Communications in Applied and Industrial Mathematics,
ISSN 2038-0909, e-530

DOI: 10.1685/journal.caim.530
Research Paper

Fractional calculus in statistical physics: the case of time fractional diffusion equation

Paolo Paradisi

Signal and Images Laboratory (SI-Lab)
Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo"
Via G. Moruzzi 1, 56124 Pisa

BCAM Basque Center for Applied Mathematics
Alameda de Mazarredo 14, 48009 Bilbao, Basque Country, Spain

E-mail: paolo.paradisi@cnr.it

Dedicated to Professor Francesco Mainardi
on the occasion of his retirement

Communicated by Gianni Pagnini and Enrico Scalas

Abstract

The relationships among intermittency with fractal Waiting Time distribution, Continuous Time Random Walk (CTRW) and the emergence of Fractional Calculus (FC) are reviewed. The derivation, in the long-time limit, of Time Fractional Diffusion Equation (TFDE) is shown and compared with the case of normal diffusion equation. Emphasis is given to the underlying connections of CTRW with concepts and results from probability theory and stochastic processes: conditional probabilities, the law of total probability, Central and (Lévy) Generalized limit theorems, renewal theory. It is shown how the emergence of a well-defined scaling rigorously emerges by imposing the invariance of the probability distribution under a group of self-similarity transformations involving space and time. The physical interpretation of some crucial mathematical passages is explained. In particular, the physical meaning of self-similarity coupled with the long-time limit is explained, having in mind a experimental point of view. Finally, the emergence of FC in complexity is discussed and associated with the ubiquitous generation of short-time transition events in the dynamics of complex systems. These renewal events are associated with the dynamical emergence (birth) and decay (death) of cooperative long-lived structures, thus giving rise to a intermittent birth-death process of cooperation.

Keywords: `fractional calculus, renewal theory, Continuous Time Random Walk, anomalous diffusion, scaling, self-similarity`

AMS subject classification: `35R11, 60G18, 60G22, 60G50, 60K05`

Received on 03 04, 2014. Accepted on 12 20, 2014. Published on 10 08, 2015.

1. Introduction

Even if Fractional Calculus (FC) has been known since the 19th century (see, e.g., [1]), its role in several applied fields is gaining momentum only in the last decades and is still a matter of debate. Some authors have been claiming for many years the potential impact of FC in fields like biology and human physiology (see recent reviews [2–4] and references therein). In more recent years, applications have been found in brain imaging [5], modeling of biological tissues [6][a], neuron modeling [8], signal processing [9], wave dynamics and viscoelasticity [10–12].

The interest here is more on applications of FC that can be traced back to complex systems, a particular topic involving different classical research fields that is becoming a subject of interest in medicine and biology Here we will adopt the point of view of statistical physics and we will discuss FC and complexity in this framework. In particular, the application of FC in statistical physics has been the subject of hundreds of papers since the $'90$, with two main directions: the first one is referred to Fractional Brownian Motion [13], while in the second one the basic model is given by the Continuous Time Random Walk (CTRW) [14–17] (see also [18] for a review). We will not consider here the first one, where long-range memory is modeled directly in the single-trajectory dynamic equations. On the contrary, we will focus on the second one, where the emergence of long-range memory is more subtle.

As far as we know, Hilfer and co-workers [19,20] were the first to find a rigorous derivation of a link between CTRW and FC, deriving the Time Fractional Diffusion Equation (TFDE) from an uncoupled CTRW with a Mittag-Leffler distribution of Waiting Times (WTs), i.e., the time duration between two successive events occurring randomly in time. In this case, the derivation of the fractional derivative is exact. However, the robustness of FC in statistical physics derive from a fundamental theorem of probability theory, regarding the sum of independent random variables, i.e., the Lévy Generalized Limit Theorem (GLT) [21,22].[b]. Compte [23] firstly

[a] It is interesting to note that in this paper the author discuss, among others, an application to the vestibulo-oculomotor neural system, citing the results of a pioneering work by Anastasio [7], which is one of the first applications (and probably the first one) of FC in biology.

[b] To understand the crucial role of this theorem, we note that the emergence of normal diffusion processes in standard statistical mechanics is an important consequence of the Central Limit Theorem (CLT) that, as well known, is based on (i) independent random variables, thus implying short memory and Markov property, and (ii) finite size fluctuations (finite variance) whose addition determines the emergence of Gaussian probability density. On the contrary, non-standard statistical mechanics is associated with (i) long-range time correlations and/or (ii) fluctuations with infinite variance. This last

DOI: 10.1685/journal.caim.530

proved the emergence of Fractional Diffusion Equations (FDEs) by means of semi-heuristic arguments applied to the Montroll-Weiss Master equation in the Laplace-Fourier space [18], thus recognizing that FDEs emerge as a general limiting description of all scale invariant diffusion processes, i.e, from CTRWs with WT and/or jump power-law distributions. Compte's arguments were made rigorous and generalized by Mainardi and Gorenflo through the concept of *well-scaled transition* (see, e.g., [24–26]).
Here the arguments of Compte-Mainardi-Gorenflo are reviewed and the relationships with concepts from probability theory are underlined and made clear. Similarly, some care is devoted to the physical interpretation of some crucial mathematical passages and constrains and, in particular, the well-scaled transition as *long-time limit* and the role of self-similarity. In Section 2 the case of normal diffusion is illustrated in order to make a comparison with the anomalous case. In Section 3 a similar route is followed to derive the TFDE from the uncoupled CTRW. Here the *long-time limit*, *continuum limit* and *self-similarity condition* are discussed in some detail. Finally, in Section 4 we give a brief discussion on the role of FC in complex systems.

2. From Markovian Random Walk to Normal Diffusion Equation

2.1. *The Markovian Random Walk*

The Markovian Random Walk (MRW) is a mesoscopic model for transport and diffusion where the walker can move only at discrete time instants: $t_n = n\Delta t$, being Δt the fixed time step of the Random Walk (RW).[c] For simplicity only the one-dimensional case is considered here, but the results are easily generalized to the multi-dimensional case.
Denoting the random jump variable at time t_n with ξ_n and with $X_n = X(t_n)$ the position of the (diffusing) random walker, we have:

$$X_n = X_{n-1} + \xi_n \quad ; \quad i = 1, ..., n \tag{1}$$

so that the MRW problem can be reduced to that of a sum of random variables: $X_n = X_0 + \xi_1 + ... + \xi_n = X_0 + \sum_{i=1}^{n} \xi_i$. In particular, we assume that ξ_i are *independent identically distributed* (i.i.d.) random variables, with Probability Density Function (PDF) p(s), defined by the relationship: $Pr\{s < \xi_i < s + ds\} = p(s)ds$[d]. The assumption of independent

condition allows a random walker to make extremely long jumps in a very short time, thus determining long-range spatial correlations. In both cases, the "long-range" behavior is mathematically described by an asymptotic slow power-law decay in memory kernels, space and time correlations, jump or WT probability distributions.

[c] Note that the statistical features of the diffusion process, such as variance growth and scaling, depend only on jump distribution and correlations among jumps.

[d] The notation $Pr\{A\}$ means *the probability of the event A*

ξ_i is equivalent to the Markovianity of X_n, while the assumption of identical distributions can be associated with homogeneous conditions in time and space. In fact, having identical distributions for the jump ξ_i means no dependence on the index i and, consequently, no dependence on the time at which the jump occurs neither on the position of the walker. Under the homogeneity assumption, the CLT (Gaussian PDF) or the GLT (Lévy stable PDF) can be applied in a straightforward way, which one of the two depending on having a finite or infinite variance [21][e].
In panel (a) of Fig. 1 a sketch of a MRW with dichotomous jumps $\xi_i = \pm 1$ is reported. In this case the walker can move on a lattice whose grid size is 1. In the dichotomous case, the general rules of conditional probabilities

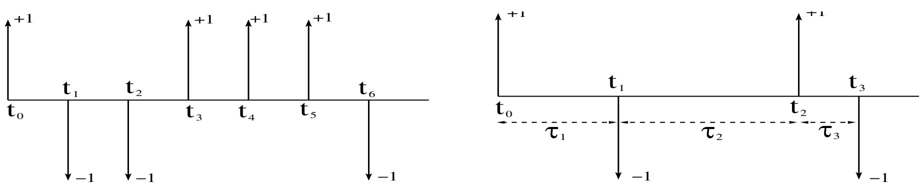

Figure 1. (a) MRW with fixed time steps. (b) CTRW with random times τ_1, τ_2,...

allow to derive the following expression:

$$P_{n+1}(j) = p_+ P_n(j-1) + p_- P_n(j+1) = \sum_{l=\pm 1} p_l \, P_n(j-l) \quad , \tag{2}$$

being $P_n(j) = P(j, t_n) = Pr\{X(t_n) = j\}$ the probability of finding the walker at position j and $p_\pm$ the probability of making a jump $\pm$. The generalization to jumps of every (discrete) size is given by the following expression:

$$P_{n+1}(j) = \sum_{l=-\infty}^{l=+\infty} p_l \, P_n(j-l) \; ; \tag{3}$$

$$P_{n+1}(j) - P_n(j) = \sum_{l=-\infty}^{l=+\infty} \left[p(j-l \to j) \, P_n(j-l) - p(j \to j+l) \, P_n(j) \right] \quad , \tag{4}$$

where the second, more familiar,form is derived form the first one by subtracting $P_{n-1}(j)$ to both sides of Equation (3) and using the equalities: $p_l = p(j-l \to j) = p(j \to j+l)$ (space homogeneity); $\sum_l p_l = 1$ (normalization). Equations (3-4) are the general forms of the (Markovian) Master

[e] Both CLT and GLT can be generalized to the general case of non-homogeneous conditions [21,22].

equation describing, in the homogeneous case, a random walker moving, at fixed time steps, on a lattice with unitary (or discrete) grid size.
The Master Equation (3-4) is essentially an application of the law of total probability. In this case, the complete set of (incompatible) events is given by the set Z of positive and negative integer numbers: $\Omega(n+1) = \Omega(n) = Z$. Then, let us recall the law of total probability:

$$Pr\{A\} = \sum_l Pr\{A|B_l\} Pr\{B_l\} \ , \qquad (5)$$

where $Pr\{A|B_l\}$ is the conditional probability of the event A given the event B_l and $\{B_l, l \text{ integer}\}$ a complete set of incompatible (disjoint) events, i.e., in formulas: $\cup_{l=1}^{\infty} B_l = \Omega$ (completeness); $B_j \cap B_k = \varnothing$ when $j \neq k$ (disjoint events). In Equation (3) the conditional probabilities of Equation (5) are given by:

$$A = \{X_{n+1} = j\} \quad ; \quad Pr\{A\} = P_{n+1}(j) \ ,$$
$$B_l = \{X_n = j - l\} \quad ; \quad Pr\{B_l\} = P_n(j-l) \ ,$$

and it results:

$$Pr\{A|B_l\} = Pr\{X_{n+1} = j | X_n = j - l\} = p(j - l \to j) = p_l \ . \qquad (8)$$

Note that the general Master Equation (3) reduces to the dichotomous Master Equation (2) by simply substituting the following conditions on the conditional probabilities of Equation (8): (i) $Pr\{A|B_{\pm 1}) = p(j \mp 1 \to j\} = p_\pm$ (ii) $Pr\{A|B_l\} = 0;\ l \neq \pm 1 \ \Rightarrow \ p_+ + p_- = 1$. When $p_+ = p_- = 1/2$ the walker's diffusion is symmetric with respect to the initial position and there is no drift. On the contrary, when $p_+ \neq p_-$, a mean drift, proportional to the difference $|p_+ - p_-|$, determine an average motion that is superposed to the walker's diffusion. However, the diffusion is still symmetric, but with respect to the time-dependent average motion.

2.2. ***Long-time limit of the MRW: the normal diffusion equation***

Let us limit ourselves to the dichotomous case, i.e., Equation (2) with $p_\pm = 1/2$, $p_l =$ for $l \neq \pm$, without any mean drift. We recall that, in Equation (2), the meaning of n is that of a time-step in units of the time scale Δt: $P_n(j) = Pr\{X(n\ \Delta t) = j\}$. Then, a microscopic time scale Δt is already included in the RW. An analogous microscopic space scale can be introduced by rescaling the grid size of the lattice from 1 to Δx and changing the interpretation of $P_n(j)$ from the probability of finding the walker, at time $n\Delta t$, in a grid point j to the probability of finding the walker in an interval of width Δx centered around $x = j\Delta x$:

$$P_n(j) = Pr\left\{ \left(j - \frac{1}{2}\right) \Delta x \leq X(n\Delta t) < \left(j + \frac{1}{2}\right) \Delta x \right\} \ . \qquad (9)$$

Now, let us re-write this equation in a *continuous form*:

$$P_n(j) = Pr\left\{x - \frac{\Delta x}{2} < X(t) < x + \frac{\Delta x}{2}\right\} = = F\left(x + \frac{\Delta x}{2}, t\right) - F\left(x - \frac{\Delta x}{2}, t\right) \ , \tag{10}$$

where $x = j\Delta x,\ t = n\Delta t$ and $F(x,t)$ is the cumulative function: $F(x,t) = Pr\{X(t) < x\}$. The long-time limit is operatively defined by the need of a laboratory time scale much greater than the microscopic one: $t \simeq t_n = n\Delta t \gg \Delta t$, thus involving $n \gg 1$. Then, the microscopic scale Δt can be treated as an infinitesimal with respect to t. Let us now make a heuristic reasoning proving that, as a consequence of the long time limit, also Δx can be treated as an infinitesimal with respect to x. In fact, the condition $n \gg 1$ means that we observe the walker after it has done many jumps and, consequently, a limit theorem for the sum of i.i.d. random variables can be applied. Being dichotomous, the jump distribution has finite variance (in particular, it is equal to 1), and the CLT can be applied[f]. In formulas: $x = j\Delta x = X(n+m) - X(m) \gg \Delta x$ for $n \gg 1$, thus also implying $j \gg 1$. Then, in analogy with Δt, also Δx can be treated as an infinitesimal with respect to x. Note that this heuristic proof could be made more rigorous by using probabilistic arguments. As a consequence of the above reasoning and assuming $F(x,t)$ to be smooth, we can apply the Taylor's expansion formula to $F(x,\cdot)$ up to the first order:

$$P_n(j) = \frac{F(x + \frac{\Delta x}{2}, t) - F(x + \frac{\Delta x}{2}, t)}{\Delta x} \Delta x \simeq \rho(x,t)\Delta x \ , \tag{12}$$

being $\rho(x,t) = \partial F/\partial x$ the PDF of the RW position at time t. By substituting Equation (12) in Equation (2) it is easy to see that ρ satisfies an equation similar to that of $P_n(j)$, but with the continuous variables $x,\ x+\Delta x,\ t,\ t+\Delta t$ instead of the discrete indices $j,\ j+1,\ n,\ n+1$:

$$\rho(x, t+\Delta t) = p_+ \rho(x - \Delta x, t) + p_- \rho(x + \Delta x, t) \ . \tag{13}$$

[f] We recall a simplified version of the Central Limit Theorem:

Theorem 2.1 (Central Limit Theorem). *Given n random variables $\xi_i,\ i = 1, n$ with the following properties: (i) ξ_i are i.i.d. random variables; (ii) $\langle \xi_i \rangle = 0$; $\langle \xi_i^2 \rangle = \sigma^2 < \infty$, then, defined the normalized sum as: $S_n = \frac{\xi_1 + \xi_2 + \ldots + \xi_n}{\sigma\ n^{1/2}}$, we have the following result:*

$$n \to \infty \ ; \quad Pr\{S_n \leq x\} \to G(x) \ ,$$

being $G(x)$ a standard Gaussian probability distribution, with zero mean and unitary variance.

From the above limit, it is clear that in the CLT the convergence is in distribution (weak convergence). A consequence of this theorem is the relation: $(X_n)^2 \sim \sigma^2 n$, being $X_n = S_n \sigma n^{1/2}$, which tells us that the distance increases with the discrete "time" n, in both average and probability.

DOI: 10.1685/journal.caim.530

As before, we can apply the Taylor's formula and we can expand $\rho(x,t)$ up to the first order in Δt and up to the third order in Δx and $-\Delta x$. Substituting into Equation (13) and making some algebraic passages and dividing both terms by Δt, we get:

$$\frac{\partial \rho(x,t)}{\partial t} = \frac{1}{2}\frac{\partial^2 \rho(x,t)}{\partial^2 x} \cdot \frac{\Delta x^2}{\Delta t} + O(\Delta t) + O\left(\frac{\Delta x^4}{\Delta t}\right) , \tag{14}$$

where $O(...)$ denotes the higher order error associated with the neglected terms in the Taylor's formula. The rigorous mathematical condition associated with the asymptotic long-time limit is given, in this case, by the *continuum limit*: $\Delta t \to 0$, $\Delta x \to 0$. As said above, from the physical point of view, this condition is equivalent to investigate time and space scales much greater than the microscopic scales Δt and Δx. This corresponds to the *long-time limit*, defined mathematically as $t \to \infty$, but actually given by the regime $t \gg \Delta t$. Note that, from the above heuristic proof exploiting the CLT, the long-time limit implies also a limit in the space variable: $x \gg \Delta x$. In this limit, three different cases can occur, but only one case generates a (normal) diffusion process:

(i) $\frac{\Delta x^2}{\Delta t} \to 0$: we get a trivial equation without any diffusion : $\frac{\partial \rho}{\partial t} = 0$

(ii) $\frac{\Delta t}{\Delta x^2} \to 0$: another trivial case without diffusion: $\frac{\partial^2 \rho}{\partial^2 x} = 0$

(iii) Normal Scaling:

$$\frac{\Delta x^2}{\Delta t} \to 2D = \text{constant} . \tag{15}$$

In the last case, we get the Normal Diffusion Equation (NDE):

$$\frac{\partial}{\partial t}\rho(x,t) = D\frac{\partial^2}{\partial^2 x}\rho(x,t) \tag{16}$$

In summary, the *normal scaling* relationship, given by Equation (15), is strictly associated with the CLT and is a necessary and sufficient condition for the emergence of a (Gaussian) diffusive phenomena and the derivation of the associated NDE. As we will explain later, the scaling constraint that must be applied in the case of CTRW is different from the normal scaling, but an anomalous scaling relationship is still necessary to get a well-defined (anomalous) diffusion process in the long-time limit.

3. From the Continuous Time Random Walk to the Time Fractional Diffusion Equation

The idea of CTRW originates in the first publications by Montroll and co-workers during the ′50 and its consequences were intensively investigated by Montroll and other collaborators in the ′60[g]. Apparently, the idea is very simple: instead of having fixed time steps, these are allowed to vary in a random way according to some statistical distribution. Despite this apparent simplicity, this idea is now recognized to have deep implications in the statistical physics of complex systems. The physical interpretation lies in the assumption that the macroscopic manifestation of the microscopic dynamics is assumed to be described not only through the statistics of spatial jumps, but also by means of the statistics of random *temporal jumps* (i.e., the WTs). Some authors associate these temporal jumps with a kind of *internal time* of the system under study. Moreover, the introduction of randomness in the temporal dimension allows for a much richer stochastic dynamics at the macroscopic scale. More recently, deep consequences of having random WTs with fat tail distributions have been found also in the ergodic properties of the system [30,31].

3.1. *The uncoupled Continuous Time Random Walk*

There are several, mathematically rigorous, definitions of the CTRW. Limiting to the one dimensional (spatial) case, a typical approach is to consider the CTRW as a two-dimensional stochastic process, with the particular feature that one of the two variables describes the temporal jumps. Let us consider the stochastic vector (X_n, T_n) evolving in the discrete time *internal time* n, where T_n has values in the positive real axis and X_n has positive or negative discrete values. Then, we have:

$$\begin{cases} X_{n+1} = X_n + \xi_{n+1} \quad ; \quad n = 0, \infty \\ T_{n+1} = T_n + \tau_{n+1} \quad ; \quad n = 0, \infty \ , \end{cases} \tag{17}$$

being $\{\tau_n\}_{n=1}^{\infty}$ and $\{\xi_n\}_{n=1}^{\infty}$ the sequences of WTs and space jumps, respectively. The CTRW is then defined by the relationship:

$$X(t) = X_n \quad ; \quad T_{n-1} < t < T_n \quad ; \quad n = 1, \infty \ , \tag{18}$$

[g] See, *e.g.*, the four fundamental papers *Random Walk on lattices I-IV* published between 1964 and 1973, [14–17], where Montroll investigated the anomalous transport of charged particles in a electrical conductor under the action of an electrical field, a condition where anomalous (non-linear) time evolution of the variance growth was observed. The dynamical system driving the transport is modelled as a a sequence of potential wells that can *trap* the charged particle for a time interval whose duration is random and depends on a complex microscopic dynamics.
It is in these papers that Montroll introduces for the first time the assumption of renewal WTs [27] with a slow power-law decay in the distribution (*i.e.*, $\langle\tau\rangle = \infty$), which is now referred to as *fractal intermittency*, and, using the formalism of the Montroll-Weiss equation in the Laplace-Fourier space, is able to give analytical solutions for the CTRW probability distribution and, consequently, for the mobility of the charged particles in the conductor. Fundamental results on the modeling of conductivity in disordered media are also given in Refs. [28,29].

DOI: 10.1685/journal.caim.530

where $T_0 = 0$. This equation defines a particular kind of subordination process [32]. In general, the increments ξ_n and τ_n are assumed to be independent[h]. Under this assumption, the stochastic process in the two dimensions (space,time), defined by Equation (17), is a Markovian process. Nevertheless, it is interesting to note that the subordinated process $X(t)$ is typically non-Markovian, with the important exception of Continuous Time Markov Chains (CTMCs), also denoted as Compound Poisson Processes, where the WT distribution and the auto-correlation are exponential functions, thus denoting a short-memory process.

Now, we want to derive the fundamental equation of CTRW under the following assumptions:

> **Assumption 1**: the initial laboratory time $t = 0$ matches the occurrence of a critical event: $T_0 = 0$ (No aging);
> **Assumption 2**: i.i.d spatial jumps (spatial homogeneity and independent jumps)
> **Assumption 3**: i.i.d. WTs (time homogeneity and renewal WTs)
> **Assumption 4**: Uncoupled WTs and spatial jumps (i.e., they are independent from each other)[i].

The spatial component X_n of Equation (17) is again described by the Markovian Master equation given in Equation (3). The time component, given by the sequence of event occurrence times $\{T_n\}$, is described in the framework of renewal theory [27], whose main ingredient is given by the WT distribution

$$Pr\{\tau < \tau_n < \tau + d\tau\} = \psi(\tau) d\tau \quad ; \quad \Psi(\tau) = Pr\{\tau_n > \tau\} = 1 - \int_0^t \psi(\tau') d\tau' \,, \tag{19}$$

being $\psi(\tau)$ the PDF of the WTs (WT-PDF) and $\Psi(\tau)$ the Survival Probability Function (SPF) of the WTs (WT-SPF)[j].

The n-th event occurrence time T_n is given by the sum of the first n random WTs: $T_n = \tau_1 + ... + \tau_n$. It is easy to see that the PDF of T_n is given by

[h] Note that the assumption of i.i.d. WTs is the condition defining the sequence $\{T_n\}_{n=1}^{\infty}$ to be a (homogeneous) *renewal* process [27], being T_n the occurrence times of some critical events (e.g., neuron spiking, turbulence bursting).

[i] The application of CTRW is particularly powerful in the case of (i) uncoupled WTs and jumps and (ii) renewal WTs. These conditions are satisfied in the majority of papers devoted to CTRW applications. However, recent investigations are being devoted also to the case of non-renewal WTs. Further, a classical example of coupled CTRW is the *Lévy Walk* [33], where the spatial jumps are exactly proportional to the WTs.

[j] Note that, as a consequence of the renewal assumption, both ψ and Ψ do not depend on the previous WT history and, as a consequence of time homogeneity, they do not depend on the event index n, neither on the laboratory time t.

the n-fold convolution of $\psi(\tau)$:

$$Pr\{\tau < T_n < \tau + d\tau\} = (\psi * ... * \psi)(\tau)d\tau = \psi^{n*}(\tau)d\tau \ , \tag{20}$$

where $*$ denotes the convolution operator and $n*$ the n-fold convolution[k]. Note that the Laplace transform[l] of ψ^{n*} is given by the n-th power of the Laplace transform of ψ: $\mathcal{L}[\psi^{n*}(\tau)](u) = \overline{\psi^{n*}}(u) = [\overline{\psi}(u)]^n$.

Now, let us introduce the CTRW probability distribution, $P(j,t) = Pr\{X(t) = j\}$, which is the analogous of $P_n(j)$ for the MRW[m]. In order to derive the evolution equation for the CTRW, we reduce the event $B = \{X(t) = j\} = \{$ the system is found in j at time $t\}$ to a union of disjoint events $B_n(t) = \{$the system reachs j exactly at time T_n of the $n-th$ event, then it remains in j up to the next jump occurring at time $T_{n+1} > t\}$. In formulas:

$$B(t) = \{X(t) = j\} = \bigcup_{n=0}^{\infty} B_n(t) \ , \tag{22}$$

$$B_n(t) = \{X(T_n) = j,\ T_{n+1} \geq t\} \ , \tag{23}$$

$$B_n(t) \cap B_m(t) = \varnothing \ ; \quad n \neq m \ . \tag{24}$$

The events B_n are disjoint as, if the system jumps in j at the $n-th$ event and time t, it cannot jump in j at some different $m-th$ event (at the same time t). Further, each event B_n is the intersection (or product) of three independent events:

(i) the $n-th$ event occurs at a time $0 < t' < t$ ($T_n = t'$);

(ii) the system jumps in j at the $n-th$ event ($X_n = j$);

(iii) the system *waits* in j for a time $t - t'$ (we know that $t - t' < \tau_{n+1}$ as $T_{n+1} > t$).

[k] The convolution operator is given by:

$$\psi^{2*}(\tau) = (\psi * \psi)(\tau) = \int_0^{\tau} \psi(s)\psi(\tau - s)ds \ .$$

The n-fold convolution is defined iteratively by: $\psi^{n*}(\tau) = (\psi * \psi^{(n-1)*})(\tau)$

[l] We recall the definition of Laplace transform of a function $f(t)$ defined on the positive real axis $[0,\infty]$:

$$\mathcal{L}[f(t)](u) = \overline{f}(u) = \int_0^{\infty} f(t)e^{-ut}dt \tag{21}$$

[m] In this definition, the variable is chosen to be integer, positive or negative. In general, it is possible to have discrete values that are not integer, i.e., $X(t) = jh$, where h is the lattice grid size.

(25) $B_n(t) = \{T_n \in [t', t' + dt'];\ 0 < t' < t\} \cap \{X_n = j\} \cap \{\tau_{n+1} > t - t'\}$,
and, passing to the probabilities of each event and using the postulates of probability theory:

(26) $Pr\{B_n(t)\} = \int_0^t Pr\{T_n \in [t', t'+dt']\}\, Pr\{X_n = j\}\, Pr\{\tau_{n+1} > t-t'\}$,

where

$$Pr\{T_n \in [t', t' + dt']\} = \psi^{n*}(t')dt' \ ; \tag{27}$$

$$Pr\{\tau_{n+1} > t - t'\} = \Psi(t - t') \ ; \tag{28}$$

$$Pr\{X_n = j\} = P_n(j) \ . \tag{29}$$

Taking the probability of both sides of Equation (22):

$$P(j, t) = Pr\{X(t) = j\} = Pr\left\{\bigcup_{n=0}^{\infty} B_n(t)\right\} = \sum_{n=0}^{\infty} Pr\{B_n(t)\} \tag{30}$$

and substituting Equations (26-29), we finally get the following result:

$$P(j, t) = \sum_{n=0}^{\infty} \int_0^t dt' \psi^{n*}(t')\Psi(t - t')P_n(j) \ . \tag{31}$$

This equation, together with Equation (3) for $P_n(j)$, represent a particular expression of the Montroll-Weiss Master equation for the CTRW under the assumptions 1-4. It is possible to eliminate $P_n(j)$ from Equation (31) substituting Equation (3):

$$P(j, t) = P_0(j)\Psi(t) + \sum_{n=1}^{\infty} \left(\psi^{n*} * \Psi\right)(t) \sum_{l=-\infty}^{+\infty} p(j - l)\ P_{n-1}(l) \ , \tag{32}$$

where $P_0(j)\Psi(t)$ is the first term of the sum, i.e., for $n = 0$, $P_0(j) = P(j, n = 0)$ is the initial condition. Note that the time integral has been rewritten as a convolution. Substituting the change of variable: $n' = n - 1$ into the previous equation, after some algebra it is easy to prove the Montroll-Weiss Master equation in the following closed form for $P(j, t)$:

$$P(j, t) = P_0(j)\Psi(t) + \int_0^t dt' \psi(t - t') \sum_{l=-\infty}^{+\infty} p(j - l)\ P(l, t') \ , \tag{33}$$

It is often convenient to study the Montroll-Weiss equation in the Laplace-Fourier space [18]. Let us define the *structure function* of the CTRW as the (discrete) Fourier transform of the jump probability $p(l)$:

$$\lambda(k) = \widehat{p}(k) = \mathcal{F}[p(l)](k) = \sum_{l=-M}^{+M} e^{ikl} p(l) \tag{34}$$

where i is the imaginary unit and k is a discrete or continuous variable depending on the lattice dimension M, i.e., the total number of sites:

- $M < +\infty$: $k = \frac{2\pi j}{M}$ where j is a (positive or negative) integer number
- $M = +\infty$: $k \in [-\pi, \pi]$.

In the following, we will consider implicitly the case of a infinite lattice dimension: $M = +\infty$. In time and space homogeneous conditions, Equation (33) is essentially a convolution in time and space. Applying the Laplace-Fourier transform to this equation[n], solving with respect to the Laplace-Fourier transform $\widehat{\overline{P}}(k,u) = \mathcal{F}\left[\mathcal{L}\left[P(x,t)\right](x,u)\right](k,u)$ and substituting $\overline{\Psi}(u) = (1-\overline{\psi}(u))/u$ (from the properties of Laplace transform), we finally get the Montroll-Weiss equation in the Laplace-Fourier space:

$$\widehat{\overline{P}}(k,u) = \frac{1-\overline{\psi}(u)}{u}\frac{\widehat{P}_0(k)}{1-\overline{\psi}(u)\lambda(k)} \ . \tag{35}$$

The Green function is defined by the initial condition $P_0(j) = \delta_{j,0}$, being δ the Kronecher symbol. This condition corresponds to all walkers starting from the position $X(t=0) = 0$. In this case, the Montroll-Weiss equation is the same as Equation (35), but with $\widehat{P}_0(k) = 1$.

3.2. *Continuum limit, long-time limit, self-similarity and scaling*

In analogy with Section (2.2), we define the rescaled probability for the CTRW. To this goal, we introduce a grid size h of the lattice and we change the interpretation of $P(j,t)$ from the probability of being in the lattice site j at time t to that of a interval of size h centered in $x_j = jh$:

$$j \to x_j = jh \Rightarrow P(j,t) = Pr\left\{X(t) \in \left[x_j - \frac{h}{2}, x_j + \frac{h}{2}\right)\right\} \ . \tag{36}$$

The main difference with respect to Equation (9), defining the rescaled distribution $P_n(j)$ for the MRW, is that the time variable is continuous, so that the long-time limit does not correspond to a continuum limit in time and space, but only in space. However, it is possible to use the formalism of self-similarity transformations [25] and, in analogy with Equation (10), we rewrite $P(j,t)$ in a *continuous form* by a formal (but heuristic) substitution $j \to x_j = x$:

$$\begin{aligned} P(j,t) &= Pr\left\{X(t) \in \left[x - \frac{\Delta x}{2}, x + \frac{\Delta x}{2}\right)\right\} = \\ &= F\left(x + \frac{\Delta x}{2}, t\right) - F\left(x - \frac{\Delta x}{2}, t\right) \simeq \rho(x,t)\Delta x \ , \end{aligned} \tag{37}$$

[n] The Fourier (Laplace) transform of the convolution of two functions is the product of the Fourier (Laplace) transforms of the two functions.

where a (smooth) cumulative function $F(x,t)$ and the corresponding PDF $\rho(x,t) = \partial F/\partial x$ have been introduced also in this case. It is easy to see that $\rho(x,t)$ satisfies the Montroll-Weiss equation (33) by substituting Equation (37), defining $P_0(t) = \rho_0(t)\Delta x$ and eliminating Δx from all the terms. In this sense, $\rho(x,t)$ is just an approximate probability density with discrete values of x, even if written in a continuous form. However, it is possible to make a more rigorous continuous form of the Montroll-Weiss equation by considering the sum over l multiplied by Δx as an approximation of the integral over x, with the additional assumption of introducing a probability density for the jumps: $p(l) \to Pr\{x_0 \to x_0 + z; z \in [x, x+\Delta x)\} = p(x)\Delta x$, where we used for simplicity the same symbol p for the discrete and continuous versions of the jump probability. Making the limit for $\Delta x \to 0$, we get the following Montroll-Weiss Master equation for continuous jump variables:

$$\rho(x,t) = \rho_0(t)\Psi(t) + \int_0^t dt'\psi(t-t') \int_{-\infty}^{+\infty} dx' p(x-x')\rho(x',t') \ , \tag{38}$$

Now, in order to explain the emergence of scaling, it is convenient to use the Montroll-Weiss equation in the Laplace-Fourier space[o].
Firstly, we consider the group of self-similarity transformations:

$$t' = at \ ; \quad x' = bx \ . \tag{39}$$

Note that the parameter a and b are adimensional as they connect two variables with the same physical dimensions.
Using these self-similarity transformations we give the following:

Definition 3.1 (Self-similarity). *A stochastic process $X(t)$ is said to be self-similar if the following relations apply:*

$$X(at) \stackrel{d}{=} bX(t) \ ; \quad b = f(a) \ , \tag{40}$$

where the equality $\stackrel{d}{=}$ is in terms of probability distribution. In this case, the cumulative probability distribution $F(x,t) = Pr\{X(t) < x\}$ is invariant with respect to the group of transformations (39) and the relationship $b = f(a)$, is denoted as scaling relationship.

[o] The passage to the continuum in the Laplace-Fourier variables can be done similarly to that made in (x,t). By substituting the (discrete) Laplace transform (or Fourier series) of Equation (34) with its continuous version, and limiting to the case of a infinite lattice ($M \to \infty$), it is easy to see that the Montroll-Weiss equation identical to the same equation for $P(j,t)$, i.e., Equation (35):

$$\widehat{\rho}(k,t) = \mathcal{F}[\rho](k) = \int_{-\infty}^{+\infty} dx e^{ikx} p(x) \quad \Rightarrow \quad \widehat{\overline{\rho}}(k,u) = \frac{1-\overline{\psi}(u)}{u} \frac{\widehat{\rho}_0(k)}{1-\overline{\psi}(u)\lambda(k)} \ .$$

Figure 2 gives a simple sketch of the idea underlying self-similarity applied to a diffusion process: when the time scale is changed ($t \to at$), then it is possible to find the same values of the probability distribution $F(x,t)$ if suitably rescaled space intervals are considered: $x \to bx$, where b depends on a, typically by means of a power-law relationship. Then, it is easy to see that

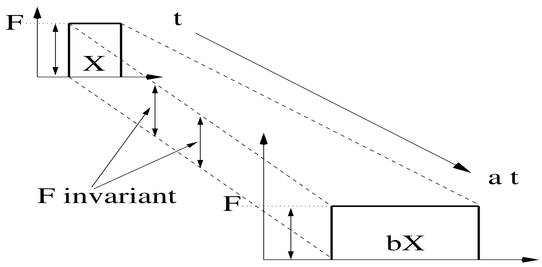

Figure 2. Sketch of self-similarity in a diffusion process. The probability F (in the y-axis) is mantained unchanged under the affine transformation: $t \to at,\ x \to bx$.

the following relationship applies to a self-similar cumulative probability distribution:

(41) $$F(x,t) = F(bx, at) \ ,$$

which is the mathematical formulation of the invariance of the cumulative distribution $F(x,t)$[p]. Notice that, in Equation (39), it is always possible to put $t' = 1$, so that we can substitute $a = 1/t$ in the right-hand side of Equation (41) and, thus, also in the scaling relationship that becomes: $b = f(1/t)$. Consequently, we find that the diffusion is actually described by a single *similarity variable* z:

(42) $$z = xf\left(\frac{1}{t}\right) \Rightarrow F(x,t) = F(z,1) = g\left(xf\left(\frac{1}{t}\right)\right)$$

In the case of a unique time scale and a unique space scale, the dimensional analysis imposes a power-law relationship of the kind:

(43) $$b = f(a) = a^{\beta} \ ,$$

which is also denoted as *scaling relationship*, so that z is also called *scaling variable*. Notice that β is a fundamental parameter emerging as a meso-(macro-)scopic manifestation of the microscopic dynamics.
As $F(x,t)$ is assumed to be smooth, a similar relationship can be derived

[p] In fact, we have: $Pr\left\{X(at) < x'\right\} = Pr\left\{bX(t) < x'\right\} = Pr\left\{X(t) < \frac{x'}{b}\right\}$ and, substituting $x' = bx$ we have: $Pr\left\{X(at) < bx\right\} = Pr\left\{X(t) < x\right\}$ and, thus, Equation (41).

DOI: 10.1685/journal.caim.530

for the PDF $\rho(x,t)$[q] and, by applying the joint Laplace-Fourier transform, for the transformed function $\widehat{\widetilde{\rho}}(k,u)$, thus finding the following results:

$$\rho(x,t) = b\rho\,(bx, at)\ \ ; \quad \widehat{\widetilde{\rho}}\,(k,u) = \frac{1}{a}\,\widehat{\widetilde{\rho}}\left(\frac{k}{b},\frac{u}{a}\right)\ , \tag{44}$$

being the Fourier transform given by the continuous version, Equation (o). We note that the limit $a \to \infty$ and $b \to \infty$ has two complementary interpretations, even if apparently opposite to each other. In both cases, we can consider x and t in Equation (39) as the microscopic space and time scales, and x' and t' as the experimental space and time scales that are observed in the laboratory. In the first interpretation, we take the variables (x', t') fixed, so that the above limit corresponds to the limit $x \to 0$ and $t \to 0$, which is is a mathematical condition imposing to the *noise* terms of both jump and WT dynamics (i.e., the terms ξ_n and τ_n in Equation (17)) to be asymptotically set to zero. In the second interpretation, actually corresponding to the real physical interpretation, we consider the variables (x,t) as fixed and the above limit corresponds to the *long-time limit*: $x' \to \infty$ and $t' \to \infty$. This interpretation is much closer to experimental reality, as the microscopic dynamical scales of a real system are fixed and typically cannot be changed, while the long-time experimental scales, at which the system is observed, define the operational laboratory conditions and the instrumental setup and tuning. In particular, this determines the constraint on the experimental sampling time, that must be much greater that the microscopic time in order to observe a diffusive behavior with a well-defined scaling[r].

[q] This can be done by differentiating both terms of Equation (41). Alternatively, it is possible to derive ρ directly from its probabilistic meaning:
$b\rho(bx,at)\Delta x = Pr\{X(at) \in [bx, bx + b\Delta x)\} = Pr\{bX(t) \in [bx, bx + b\Delta x)\} =$
$= Pr\{X(t) \in [x, x + \Delta x)\} = \rho(x,t)\Delta x$ and, eliminating Δx from both sides, we finally get Equation (44).

[r] If the device performance would allow to obtain a high time resolution so as to reach the microscopic time scales (or slightly above), it is clear that it will be necessary to perform long-time averages in order to get a diffusive behavior. The recent advance in the technical features of instrumentation is actually allowing to investigate the CTRW model at the microscopic scales that are of some interest in biology (e.g., the cell level), where some authors are finding very interesting results about the applicability of CTRW model and, even more important, on its consequences on the ergodicity breaking found, for example, at the level of cell internal dynamics (see, e.g., [30,31])
Finally, we note that a diffusive behavior can be observed if the size of the system is large enough to allow for an intermediate range of temporal scales that are much greater than the microscopic time scales and much smaller that the time the system takes to reach the borders.

3.3. *The Time-Fractional Diffusion Equation as the long-time limit of the uncoupled CTRW*

Let us first consider the case of a finite first moment for the WT distribution: $\langle\tau\rangle < \infty$. We have the following [21,22,34]:

Theorem 3.1 (Expansion for distributions with finite moments).
(i)
Given the Fourier trasform defined by Equation (34), if the first N moments of the jump distribution $p(l)$ exist and have finite values, then it is possible to write the following expansion:

$$\widehat{p}(k) = \sum_{l=0}^{N} \frac{(ik)^l}{l!} \langle \xi^l \rangle + O\left(k^{N+1}\right) \tag{45}$$

(ii)
Given the Laplace trasform defined by Equation (21), if the first N moments of the WT distribution $\psi(\tau)$ exist and have finite values, then it is possible to write the following expansion:

$$\overline{\psi}(u) = \sum_{l=0}^{N} \frac{(-u)^l}{l!} \langle \tau^l \rangle + O\left(u^{N+1}\right) \tag{46}$$

Note that, in both cases, the expansion is valid up to $N = \infty$ if all moments exist and have finite values.

Then, we can consider the following approximations for jump and WT statistical distributions:

$$k \to 0^+ : \lambda(k) = \widehat{p}(k) \simeq 1 - \frac{\langle \xi^2 \rangle}{2} k^2 \ , \tag{47}$$

$$u \to 0^+ : \overline{\psi}(u) \simeq 1 - \langle \tau \rangle u \ . \tag{48}$$

Being the second moment finite, the first assumption is equivalent to assume jump statistics in the Gaussian basin of attraction, while the effects of the second assumption are clarified in the following[s]. To do this, we use the Montroll-Weiss equation for $\widehat{\overline{\rho}}$, Equation (o), but rescaled with a and b as in Equation (44). In the limit $a, b \to \infty$ we can substitute the previous approximate expressions for $\lambda(k/b)$ and $\overline{\psi}(u/a)$ and, after some algebra, we get:

[s] However, we can notice that the exponential distribution satisfies this assumption. Being the exponential distribution a fundamental feature of Continuous Time Markov Chains, then, to our goal, this can be seen as a particular case of CTRW with exponential WT distribution. Without going into details, it is possible to prove that, when the WT-PDF is an exponential function, the Montroll-Weiss Master equation, Equation (33), reduces exactly to a Markovian Master equation similar to Equation 4 but with a continuous time derivative dP/dt instead of a discrete time difference $P_{n+1} - P_n$.

DOI: 10.1685/journal.caim.530

$$\frac{1}{a}\,\widehat{\widetilde{\rho}}\left(\frac{k}{b},\frac{u}{a}\right) = \frac{1}{u + \frac{\langle\xi^2\rangle}{2\langle\tau\rangle}\frac{a}{b^2}k^2 - \frac{\langle\xi^2\rangle k^2 u}{2b^2}} \quad , \tag{49}$$

where, for simplicity, we considered the initial condition $\rho_0(x) = \delta(x)$, defining the Green function of the problem and giving $\widehat{\rho}_0(k) = 1$. In the self-similarity assumption this expression must be equal to $\widehat{\widetilde{\rho}}(k,u)$ (without a and b). It is easy to see that this condition can be imposed in the asymptotic limit $a, b \to \infty$. In fact, the last term in the denominator becomes negligible and, in analogy with the observations made at the end of Subsection 2.2, a non-trivial and diffusive solution exists if and only if we assume the *normal scaling*:

$$b^2 = a \Rightarrow b = f(a) = a^{1/2} \ . \tag{50}$$

Under this assumption we get:

$$\widehat{\widetilde{\rho}}(k,u) = \frac{1}{u + Dk^2} \quad ; \quad D = \frac{\langle\xi^2\rangle}{2\langle\tau\rangle} \quad , \tag{51}$$

which is exactly the Laplace-Fourier form of the NDE, Equation (16), for the Green function, i.e., $\rho(x,t=0) = \delta(x)$.
This result proves that, in order to get anomalous diffusion from the uncoupled CTRW, it is necessary to admit a WT distribution with an infinite average time, thus implying a very slow power-law decay in the range of long WTs: $\psi(\tau) \sim 1/\tau^{\mu},\ 1 < \mu \leq 2$[t]. In this case, instead of Theorem 3.1, the following theorem applies [25,32]:

[t] This kind of WT distributions is expected to be ubiquitous as it belongs to the basin of attraction of a one-sided Lévy stable density, as established by the following theorem [21,22,34]:

Theorem 3.2 (Lévy Generalized Limit Theorem for one-sided distributions). *Let us consider n i.i.d. random variables: $\tau_i,\ i = 1, n$, with values in the positive real axis. Given the PDF $\psi(\tau)$ and the survival probability $\Psi(\tau)$ (see Equation (19)), let us assume the following long-time behavior:*

$$\tau \to \infty : \quad \Psi(\tau) \sim \frac{A}{\tau^{\mu-1}} \ ; \quad 1 < \mu \leq 2 \ . \tag{52}$$

Then, defined the normalized sum as: $S_n = \frac{\xi_1+\xi_2+\ldots+\xi_n}{n^{1/\alpha}}$, where

$$\alpha = \mu - 1 \tag{53}$$

is the Lévy index restricted to the range $0 < \alpha \leq 1$, we have the following result:

$$n \to \infty \ ; \quad Pr\{S_n < x\} \to G_\alpha(x) \ ,$$

being $G_\alpha(x)$ a one-sided Lévy density of index α.

This is a particular case, limited to one-sided distributions, of the GLT, which includes distributions on the entire real axis, which can be symmetric or not, and with Lévy index in the range $0 < \alpha < 2$ [21,22,34]. [35,36] Finally, we note that μ, denoted as *intermittency exponent* or *complexity index* [37–39] is, similarly to the exponent β of the scaling relationship (43), a emerging (meso-)macro-scopic property of the unknown microscopic dynamics.

Theorem 3.3 (Tauberian Lemma). *For a WT-PDF with an asymptotic power-law decay $\psi(\tau) \sim 1/\tau^{\mu}$, $1 < \mu \leq 2$, given the associated Survival Probability $\Psi(\tau)$, the following statements are equivalent:*

(i) $\tau \to \infty : \quad \Psi(\tau) \sim \frac{A}{\tau^{\mu-1}} \ ; \quad 1 < \mu \leq 2 \quad .$

(ii) $u \to 0 : \quad \overline{\psi}(u) \sim 1 - A u^{\mu-1} \ .$

This is a particular case of a more general Tauberian Theorem relating the limiting behavior of Laplace and Fourier transforms to the asymptotic power-law behavior of statistical distributions[u]*.*

Substituting this theorem, together with Equation (47) for the jump distribution, into the Montroll-Weiss Equation (o) in the rescaled version, Equation (44), we get the following expression:

$$\frac{1}{a}\,\widehat{\overline{\rho}}\left(\frac{k}{b},\frac{u}{a}\right) = \frac{u^{\mu-2}}{u^{\mu-1} + \frac{\langle\xi^2\rangle}{2A}\frac{a^{\mu-1}}{b^2}k^2 - \frac{\langle\xi^2\rangle k^2 u^{\mu-1}}{2b^2}} \quad , \tag{55}$$

Similarly to the previous case, self-similarity, in the form of invariance of Laplace-Fourier transform, can be applied in the long-time limit $a, b \to \infty$, as the last term in the denominator becomes negligible also in this case. It is then easy to see the emergence of the anomalous scaling relationship:

$$b^2 = a^{\alpha} \Rightarrow b = f(a) = a^{\alpha/2} \ ; \quad z = x f\left(\frac{1}{t}\right) = \frac{x}{t^{\alpha/2}} \ ; \alpha = \mu - 1 \ , \tag{56}$$

where also the similarity variable z has been reported. Then, the long-time limit non-trivial solution for the Montroll-Weiss equation in the Laplace-Fourier space is given by:

$$\widehat{\overline{\rho}}(k,u) = \frac{u^{\alpha-1}}{u^{\alpha} + D_{\alpha}k^2} \quad ; \quad D_{\alpha} = \frac{\langle\xi^2\rangle}{2A} = \frac{\langle\xi^2\rangle}{2T^{\alpha}} \quad , \tag{57}$$

where the Lévy index α has been used instead of the intermittency exponent μ. In general, we can rewrite the previous expression:

$$u^{\alpha}\widehat{\overline{\rho}}(k,u) - u^{\alpha-1}\widehat{\rho}_0(k) = -D_{\alpha}k^2\widehat{\overline{\rho}}(k,u) \ , \tag{58}$$

where $\widehat{\rho}_0(k) = \widehat{\rho}(k, t=0) = 1$ and, using the properties of Laplace transform of Caputo fractional derivative operators[v], we finally derive the fol-

[u] Notice that the constant A is related to a time scale through the dimensional relationship: $A = T^{\mu-1}$.

[v] The fractional derivative in the Caputo sense is given by (see, e.g., [40]):

$$\frac{\partial^{\alpha}}{\partial t^{\alpha}} f(t) = \frac{1}{\Gamma(n-\alpha)}\int_0^t dt' \frac{f^{(n)}(t')}{(t-t')^{\alpha+1-n}} \ , \quad n-1 \leq \alpha < n \ , \tag{59}$$

being $f^{(n)}$ the n-th order time derivative of the function $f(t)$. The Laplace transform is given by:

$$\mathcal{L}[D_t^{\alpha} f(t)](u) = u^{\alpha}\mathcal{L}[f(t)](u) - \sum_{k=0}^{n-1} f^{(k)}(0^+)u^{\alpha-1-k} \tag{60}$$

.

lowing TFDE [26][w]:

$$\frac{\partial^\alpha}{\partial t^\alpha}\rho(x,t) = D_\alpha \frac{\partial^2}{\partial x^2}\rho(x,t) \tag{61}$$

Mainardi, Gorenflo and co-workers gave important contributions in finding both analytical solutions and numerical algorithms for the above TFDE (see, e.g., [40,41]), and also in clarifying some mathematical aspects related to a rigorous derivation of fractional diffusion equations from both MRW (Space Fractional DE) and CTRW (Space-Time Fractional DE), adopting a technique that they called *well-scaled transition*. These aspects have been discussed here in order to explain the relationships with the concepts of self-similarity and scale invariance. In particular, we underlined how the scaling relationship emerges, for given jump and WT statistics, as a necessary condition for the existence of a non-trivial, diffusive, solution. Among other contributions, we cite Refs. [42,43], where the fundamental solutions of the TFDE are given in terms of the Mainardi's function:

$$M_\nu(z) = \sum_{n=0}^{\infty} \frac{(-z)^n}{n!\Gamma(-\nu n + (1-\nu))} = \frac{1}{\pi}\sum_{n=1}^{\infty} \frac{(-z)^{n-1}}{(n-1)!}\Gamma(\nu n)\sin(\pi\nu n) \ . \tag{62}$$

In particular, the solution to the Cauchy problem for the Green function, i.e., with initial condition: $\rho(x, t = 0) = \delta(x)$, is given by:

$$\rho_G(x,t) = \frac{1}{2\sqrt{D_\alpha} t^{\alpha/2}} M_{\alpha/2}(|z|) \tag{63}$$

where the subscript G refers to Green and z is the similarity variable given in Equation (56).

4. Discussion: On the role of fractional calculus in complexity

In statistical physics, the transition from the MRW to the CTRW with a slow power-law decay in asymptotic WT statistics corresponds to a passage from the CLT to the Lévy Generalized Limit Theorem. In the first case, only PDFs belonging to the Gaussian basin of attraction are considered, which limits us to consider a noise term with independent increments and finite variance jumps. The failure of the *finite variance* hypothesis determines the emergence of super-diffusion driven by so-called Lévy flights and, in the long-time limit, of (Riesz-Feller) space fractional derivatives [40]. This is always in agreement with the general Markovian Master Equation (3), but with the assumption of an infinite variance for the jump distribution. To get this, a slow power-law decay must be assumed: $p_l \sim \frac{1}{l^{1+\alpha}}$; $0 < \alpha < 2$, thus implying p_l to belong to the basin of attraction of a Lévy stable

[w] The Normal Diffusion Equation is recovered for $\mu = 2$ ($\alpha = 1$).

density of order α. This can be summarized in a simplified conceptual chain: infinite variance $\rightarrow$ power-law (scale invariant) jump distribution $\rightarrow$ power-law space kernel (non-locality) $\rightarrow$ (Riesz-Feller) space fractional derivative. However, the emergence of TFDE is, in some sense, more interesting, at least from the point of view of complexity and this is related to the underlyng renewal process describing a sequence of critical events [27]. Complex systems are characterized by some specific features that are commonly accepted in the scientific community as peculiar properties of complexity[x]. Among others, the main feature, probably the unique one to be universally accepted, is the concept of *emergence* of cooperative behaviour and self-organized structures. These emergent properties cannot be explained only in terms of the single components (*non-reducibility*). Moreover, it is found that emergent cooperative structures are not equilibrium states, but metastable states. A crucial aspect for the application of CTRW and, then, for the emergence of FC, is the typical intermittent behavior of these structures, that is, an alternance between (i) long life-times in which a global structure emerges showing some level of *coherence*, in the sense that is can be clearly identified (e.g., a quasi-stable large-scale vortex in turbulence) and (ii) short-time *critical events*, erasing memory of the past (i.e., *renewal* condition [27]) and associated with the emergence (birth) and decay (death) of coherent, self-organized structures or states. In fact, in many complex systems, ranging from earthquakes to Blinking Quantum Dots, turbulence and brain dynamics [37–39,44–48], there's an experimental evidence of power-law behavior emerging in the statistics of WTs, where WTs are defined as the time intervals between two successive critical events. These are, in turn, defined as bursting events in the time series, i.e., as rapid transitions (see, e.g., brain dynamics [37–39]) marked by abrupt changes in the derivative of the signal [49]. The interesting case, given by power-law WT statistics, is denoted as *fractal intermittency.* The definition of complex systems in terms of self-organized structures that are not stable, but evolve according to some fractal intermittent birth-death process of cooperation is still a matter of discussion in the scientific community. In fact, it is largely accepted that the signatures of complexity should be some power-law behavior in the topological structure of the system (e.g., degree distribution in a complex network), in the spatial correlations (e.g., critical phenomena) and in the event intensity (e.g., avalanche size distribution in a Self-Organized System) [50],

[x] Similarly to the case of turbulence, a precise and unique definition of *complexity* is still a matter of debate and, up to our knowledge, there are no definitions that are accepted by overall scientific community. However, as in turbulence, it is possible to make a list of features that are recognized as general properties of complex systems by the most part of the scientific community.

all conditions that can be referred as *structural* or *spatial* complexity. On the contrary, the investigations on the so-called *Temporal Complexity* [51], typically emerging in complex systems in the form of fractal intermittency, are still at the beginning. This seems to be a fruitful research direction as, being anomalous diffusion a signature of complexity and being, in turn, anomalous diffusion related to fractal intermittency through CTRW, temporal complexity should play some central role.

Finally, we notice that the interpretation of critical events in terms of (fast) birth-death transition events among metastable, long-lived cooperative or self-organized structures, is something more than a simple conjecture, as it has been recently confirmed by some theoretical investigations on simple stochastic systems interconnected through simple or complex networks [52].

Acknowledgements and a historical note

In this paper I reviewed some results that can be found in papers published by Montroll and co-workers during the $'$60s [14–17] and in more recent papers by Mainardi, Gorenflo and co-workers [25,26,40–43]). In particular, I gave a detailed explanation of the mathematical derivation of CTRW (MRW) from fundamental ideas of probability theory and of TFDE (NDE) from CTRW (MRW), also discussing the corresponding physical interpretation of the most crucial mathematical passages. In particular, the role of self-similarity and the need for the emergence of a scaling relationship has been discussed.
Moreover, I took the chance of this Special Issue devoted to the 70th birthday of Prof. Francesco Mainardi to make a partial revision of my *laurea* thesis *Analysis of diffusive phenomena by Non-Markovian stochastic processes* (in italian *Analisi di fenomeni diffusivi tramite processi stocastici Non-Markoviani*), written during the first half of 1997 under the supervision of Prof. Mainardi and discussed during the 1997 summer session. The greatest part of this paper is a revised version of the second chapter, dedicated to the introduction of CTRW, and I added several discussions about some subtle points, such as self-similarity and long-time limit. Other parts are also taken from chapters 1,3 and 4.

I thank very much Prof. Mainardi that introduced me to fractional calculus and gave me the chance to study the CTRW model in the perspective of investigating the relationships with fractional diffusion equations, a topic that is still giving me new stimuli and ideas in my current research activity in applied fields such as biology, bio-chemistry, neuro-physiology and turbulence.

REFERENCES

1. A. Kilbas, H. Srivastava, and J.J.Trujillo, *Theory and applications of fractional differential equations.* Amsterdam: Elsevier, 2006.
2. B. West, Fractal Physiology and the fractional calculus: a perspective, *Frontiers in Physiology*, vol. **1**, p. 12, 2010.
3. B. West and D. West, Fractional dynamics of allometry, *Fractional Calculus and Applied Analysis*, vol. **15**, no. 1, pp. 70–96, 2012.
4. G. Werner, Fractals in the nervous system: conceptual implications for theoretical neuroscience, *Frontiers in Physiology*, vol. **1**, p. 15, 2010.
5. R. Magin, C. Ingo, L. Colon-Perez, W. Triplett, and T. Mareci, Characterization of anomalous diffusion in porous biological tissues using fractional order derivatives and entropy, *Microporous and Mesoporous Materials*, vol. **178**, pp. 39–43, 2013.
6. R. Magin, Fractional calculus models of complex dynamics in biological tissues, *Computers and Mathematics with Applications*, vol. **59**, 2010.
7. T. Anastasio, The fractional-order dynamics of brainstem vestibulo-oculomotor neurons, *Biological Cybernetics*, vol. **72**, pp. 69–79, 1994.
8. B. Lundstrom, M. Higgs, W. Spain, and A. Fairhall, Fractional differentiation by neocortical pyramidal neurons, *Nature Neuroscience*, vol. **11**, pp. 1335–1342, 2008.
9. H. Sheng, Y. Chen, and T. Qiu, *Fractional Processes and Fractional-Order Signal Processing: Techniques and Applications.* London: Springer-Verlag, 2012.
10. F. Mainardi and P. Paradisi, *A model of diffusive waves in viscoelasticity based on fractional calculus*, pp. 4961–4966. IEEE Conference on Decision and Control - Proceedings of the 36th Ieee Conference on Decision and Control, Vols **1-5**, 1997.
11. F. Mainardi and P. Paradisi, Fractional diffusive waves, *Journal of Computational Acoustics*, vol. **9**, no. 4, pp. 1417–1436, 2001.
12. F. Mainardi, *Fractional calculus and waves in linear viscoelasticity.* London: Imperial College, 2010.
13. B. Mandelbrot and J. van Ness, Fractional Brownian motions, fractional noises and applications, *SIAM Review*, vol. **10**, no. 4, pp. 422–437, 1968.
14. E. Montroll, Random Walks on Lattices, *Proc. Symp. Appl. Math., Am. Math. Soc.*, vol. **16**, pp. 193–220, 1964.
15. E. Montroll and G. Weiss, Random Walks on Lattices II, *J. Math. Phys.*, vol. **6**, no. 2, pp. 167–181, 1965.
16. E. Montroll, Random Walks on Lattices III. Calculation of First-Passage Times with Application to Exciton Trapping on Photosynthetic Units, *J. Math. Phys.*, vol. **10**, no. 4, pp. 753–765, 1969.

DOI: 10.1685/journal.caim.530

17. E. Montroll and H. Scher, Random Walks on Lattices IV. Continuous-Time Walks and Influence of Absorbing Boundaries, *J. Stat. Phys.*, vol. **9**, no. 2, pp. 101–135, 1973.
18. G. Weiss and R. Rubin, Random Walks:Theory and Selected Applications, in *Advances in Chemical Physics* (P. I. and R. S.A., eds.), vol. **52**, pp. 363–505, John Wiley and Sons, 1983.
19. R. Hilfer, Classification Theory for Anequilibrium Phase Transitions, *Physical Review E*, vol. **48**, no. 4, pp. 2466–2475, 1993.
20. R. Hilfer and L. Anton, Fractional Master Equations and Fractal Time Random Walks, *Physical Review E*, vol. **51**, no. 2, pp. 848–851, 1995.
21. P. Lèvy, *Thèorie de l'Addition des Variables Alèatoires.* Paris: Gauthier-Villars, 2 ed., 1954.
22. B. Gnedenko and A. Kolmogorov, *Limit Distributions for Sums of Independent Random Variables.* Reading, MA: Addison-Wesley, 1954. Transl. from Russian by K.L. Chung.
23. A. Compte, Stochastic foundations of fractional dynamics, *Physical Review E*, vol. **53**, pp. 4191–4193, 1996.
24. E. Scalas, R. Gorenflo, and F. Mainardi, Uncoupled continuous-time random walks: solution and limiting behavior of the master equation, *Physical Review E*, vol. **69**, p. 011107, 2004.
25. R. Gorenflo and F. Mainardi, Simply and multiply scaled diffusion limits for continuous time random walks, *Journal of Physics: Conference Series*, vol. **7**, pp. 1–16, 2005.
26. R. Gorenflo and F. Mainardi, Subordination pathways to fractional diffusion, *European Physical Journal - Special Topics*, vol. **193**, pp. 119–132, 2011.
27. D. Cox, *Renewal Theory.* London: Methuen, 1962.
28. H. Scher and M. Lax, Stochastic transport in a disordered solid. I: Theory, *Physical Review B*, vol. **7**, pp. 4491–4502, 1973.
29. H. Scher and M. Lax, Stochastic transport in a disordered solid. II: Impurity condition, *Physical Review B*, vol. **7**, pp. 4502–4519, 1973.
30. A. Lubelski, I. Sokolov, and J. Klafter, Nonergodicity Mimics Inhomogeneity in Single Particle Tracking, *Physical Review Letters*, vol. **100**, p. 250602, 2008.
31. J.-H. Jeon, V. Tejedor, S. Burov, E. Barkai, C. Selhuber-Unkel, K. Berg-Srensen, L. Oddershede, and R. Metzler, In vivo anomalous diffusion and weak ergodicity breaking of lipid granules, *Physical Review Letters*, vol. **106**, p. 048103, 2011.
32. W. Feller, *An Introduction to Probability Theory and Its Applications.* New York: John Wiley and Sons, 2 ed., 1966.

33. M. Shlesinger, J. Klafter, and B. West, Lèvy Walks with Applications to Turbulence and Chaos, *Physica*, vol. **140A**, pp. 212–218, 1986.

34. V. Zolotarev, *One-dimensional Stable Distributions.* Providence: American Mathematical Society, 1983.

35. O. C. Akin, P. Paradisi, and P. Grigolini, Periodic trend and fluctuations: The case of strong correlation, *Physica A*, vol. **371**, no. 2, pp. 157–170.

36. O. C. Akin, P. Paradisi, and P. Grigolini, Perturbation-induced emergence of poisson-like behavior in non-poisson systems, *Journal of Statistical Mechanics-Theory and Experiment.* P01013, JAN 2009, DOI: 10.1088/1742-5468/2009/01/p01013.

37. P. Allegrini, D. Menicucci, R. Bedini, L. Fronzoni, A. Gemignani, P. Grigolini, B. J. West, and P. Paradisi, Spontaneous brain activity as a source of ideal 1/f noise, *Physical Review E*, vol. **80**, no. 6, 2009.

38. P. Allegrini, D. Menicucci, R. Bedini, A. Gemignani, and P. Paradisi, Complex intermittency blurred by noise: Theory and application to neural dynamics, *Physical Review E*, vol. **82**, no. 1, 2010.

39. P. Paradisi, P. Allegrini, A. Gemignani, M. Laurino, D. Menicucci, and A. Piarulli, Scaling and intermittency of brain events as a manifestation of consciousness, *AIP Conference Proceedings*, vol. **1510**, pp. 151–161, 2013.

40. R. Gorenflo, F. Mainardi, D. Moretti, G. Pagnini, and P. Paradisi, Discrete random walk models for space-time fractional diffusion, *Chemical Physics*, vol. **284**, no. 1-2, pp. 521–541, 2002.

41. R. Gorenflo and F. Mainardi, Some recent advances in theory and simulation of fractional diffusion processes, *Journal of Computational and Applied Mathematics*, vol. **229**, pp. 400–415, 2009.

42. F. Mainardi, Fractional Relaxation-Oscillation and Fractional Diffusion-Wave Phenomena, *Chaos, Solitons & Fractals*, vol. **7**, no. 9, pp. 1461–1477, 1996.

43. F. Mainardi, The fundamental solutions for the fractional diffusion-wave equation, *Applied Mathematics Letters*, vol. **9**, pp. 23–28, 1996.

44. P. Paradisi, P. Allegrini, F. Barbi, S. Bianco, and P. Grigolini, Renewal, modulation and blinking quantum dots, *AIP Conference Proceedings*, vol. **800**, pp. 92–97, 2005.

45. P. Paradisi, R. Cesari, A. Donateo, D. Contini, and P. Allegrini, Scaling laws of diffusion and time intermittency generated by coherent structures in atmospheric turbulence, *Nonlinear Processes in Geophysics*, vol. **19**, no. 1, pp. 113–126, 2012.

46. P. Paradisi, R. Cesari, A. Donateo, D. Contini, and P. Allegrini, Corrigendum to Scaling laws of diffusion and time intermittency generated

by coherent structures in atmospheric turbulence (vol 19, pg 113, 2012), *Nonlinear Processes in Geophysics*, vol. **19**, no. 6, pp. 685–685, 2012.

47. P. Paradisi, R. Cesari, F. Mainardi, A. Maurizi, and F. Tampieri, A generalized Fick's law to describe non-local transport effects, *Physics and Chemistry of the Earth Part B - Hydrology, oceans and atmosphere*, vol. **26**, no. 4, pp. 275–279, 2001.
48. P. Paradisi, R. Cesari, A. Donateo, D. Contini, and P. Allegrini, Diffusion scaling in event-driven random walks: an application to turbulence, *Reports on mathematical physics*, vol. **70**, no. 2, pp. 205–220, 2012.
49. A. A. Fingelkurts and A. A. Fingelkurts, Brain-Mind Operational Architectonics Imaging: Technical and Methodo- logical Aspects, *The Open Neuroimaging Journal*, vol. **2**, pp. 73–93, 2008.
50. S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D. Hwang, Complex networks: Structure and dynamics, *Physics Reports*, vol. **424**, p. 175308, 2006.
51. M. Turalska, B. J. West, and P. Grigolini, Temporal complexity of the order parameter at the phase transition, *Physical Review E*, vol. **83**, p. 061142, 2011.
52. E. Lovecchio, P. Allegrini, E. Geneston, B. J. West, and P. Grigolini, From self-organized to extended criticality, *Frontiers in Physiology*, vol. **3**, p. 98, 2012.