

# Exact asymptotic solutions to nonlinear Hawkes processes: a systematic classification of the steady-state solutions

Kiyoshi Kanazawa<sup>1,2</sup> and Didier Sornette<sup>3,4</sup>

<sup>1</sup> Faculty of Engineering, Information and Systems,

The University of Tsukuba, Tennodai, Tsukuba, Ibaraki 305-8573, Japan

<sup>2</sup> JST, PRESTO, 4-1-8 Honcho, Kawaguchi, Saitama 332-0012, Japan

<sup>3</sup> ETH Zurich, Department of Management, Technology and Economics, Zurich, Switzerland

<sup>4</sup> Institute of Risk Analysis, Prediction and Management, Academy for Advanced Interdisciplinary Studies, Southern University of Science and Technology, Shenzhen, China

(Dated: May 9, 2022)

Hawkes point processes are first-order non-Markovian stochastic models of intermittent bursty dynamics with applications to physical, seismic, epidemic, biological, financial, and social systems. While accounting for positive feedback loops that may lead to critical phenomena in complex systems, the standard linear Hawkes process only describes excitatory phenomena. To describe the co-existence of excitatory and inhibitory effects (or negative feedbacks) as occurs for instance in seismic (so-called stress shadows) and neural systems (via glycine and gamma-aminobutyric acid (GABA) transmitters), extensions involving nonlinear dependences of the intensity as a function of past activity are needed. However, such nonlinear Hawkes processes have been found hitherto to be analytically intractable due to the interplay between their non-Markovian and nonlinear characteristics, with no analytical solutions available. Here we present various exact and robust asymptotic solutions to nonlinear Hawkes processes using the field Master equation (ME) approach introduced previously by the authors. We report explicit power law formulas for the steady-state intensity distributions  $P_{ss}(\lambda) \propto \lambda^{-1-a}$ , where the tail exponent  $a$  is expressed analytically as a function of parameters of the nonlinear Hawkes models. We present three robust interesting characteristics of the nonlinear Hawkes process: (i) for one-sided positive marks (i.e., in the absence of inhibitory effects), the nonlinear Hawkes process can exhibit any power law relation either as intermediate asymptotics ( $a \leq 0$ ) or as true asymptotics ( $a > 0$ ) by appropriate model selection; (ii) for distribution of marks with zero mean (i.e., for balanced excitatory and inhibitory effects), the Zipf law ( $a \approx 1$ ) is universally observed for a wide class of nonlinear Hawkes processes with fast-accelerating intensity map; (iii) for marks with a negative mean, the asymptotic power law tail becomes lighter as the mean mark becomes more negative. We introduce the basic analytical tools for advanced Hawkes modeling, particularly for model calibration to real time-series data in various complex systems.

## I. INTRODUCTION

Intermittent bursts are ubiquitously observed with temporal and spatial clustering characters in physical [1, 2], seismic [3–6], epidemic [7], financial [8–10], and social systems [11, 12]. Such bursty dynamics can be well described by the Hawkes process [13–15], a non-Markovian self-excited point process capturing both long memory effects and critical bursts, such that past events keep their potential influence to trigger future bursty events for a long time, potentially leading to critical bursts. However, the essential non-Markovian nature of this model has been an obstacle preventing the development of a unified analytical theory because the established framework of Markovian stochastic processes is not applicable.

Recently, however, a new theoretical scheme was developed to address such non-Markovian stochastic processes directly, in particular for the Hawkes process [16, 17]. This scheme is based on a mapping from the non-Markovian Hawkes model to an equivalent stochastic partial differential equation (SPDE). The SPDE is then mapped to an equivalent field master equation (field ME), i.e., a functional-differential equation for the probability density functional (PDF) of the intensity. The solutions of this equation can be obtained analytically in their asymptotic form, in particular near criticality. This theoretical framework predicted a novel non-universal power law relation for the intensity as an intermediate asymptotics [18]. It has the potential for further explorations of the theoretical properties of more general Hawkes processes.

Since the basic linear Hawkes (LHawkes) process is analytically solved in this framework, it is natural to seek further generalisation of the framework, such as for nonlinear Hawkes (NLHawkes) processes [19, 20]. NLHawkes processes are particularly important to account for the presence of inhibitory effects: in addition to positive feedbacks, many systems are characterised by co-existing negative feedbacks. In the context of point processes, while the standard Hawkes process describes only excitatory processes, many systems are kept in balance by the additional occurrence of inhibitory processes. For instance, inhibitory effects naturally appear in seismicity [21, 22] as any earthquake creates a tensorial stress perturbation within the visco-plasto-elastic Earth crust with the presence of “stress shadows” in

certain regions around the ruptured fault where future earthquakes are less likely [23], while other regions are brought closer to rupture by an increase in the local relevant stress component. Similarly, neurobiological brains are kept in balance by the interplay between excitatory and inhibitory neurotransmitters, with the resulting cascades of excitations exhibiting power law statistics [24–26].

The long-standing problem of combining inhibitory and excitatory effects in point processes requires considering nonlinear extensions of the Hawkes processes in order to fulfill the condition that the intensity (a probability per unit time) remains nonnegative. Our recent Letter [27] has presented a step toward a general theory of NLHawkes processes by applying the framework of the field ME [16, 17]. In this Letter, we discovered the existence of an asymptotic ubiquitous power law distribution of the intensity for NLHawkes processes in the case of mark distributions with non-positive mean. Since NLHawkes processes may have a huge variety of forms, and thus of control parameters, for instance in the tension-intensity map defined below and in the mark distribution, it would be useful to further study various NLHawkes processes by systematically classifying their solutions according to the asymptotic analyses of the field MEs.

The present article supplements our short Letter [27] by providing a systematic classification of various NLHawkes processes, together with various explicit exact and asymptotic solutions. In this paper, we present a general formulation for the NLHawkes processes and provide their explicit solutions for various cases. In particular, we report three interesting asymptotic features which are valid for a wide class of memory kernels. (i) In the absence of inhibitory effects (i.e., when events all have positive marks), we find a non-universal power law relation for the intensity distribution at criticality, with an exponent  $a$  that can take any value, i.e., corresponding to a genuine power law ( $a > 0$ ) or to an intermediate power law asymptotics ( $a \leq 0$ ). This is in contrast to the LHawkes model, where only a negative exponent  $a < 0$  exists [16, 17]. (ii) In the presence of inhibitory effects (i.e., both positive and negative marks coexist), in the case where the mark distribution has zero mean corresponding to a balance between inhibitory and excitatory effects, a wide class of NLHawkes processes exhibit Zipf's law ( $a \approx 1$ ) for their intensity distributions. (iii) For negative mean marks, we derive the asymptotic formula for the intensity PDF, whose tail becomes thinner than in the zero mean mark case. This provides a new mechanism for the ubiquity of power laws, including Zipf's law, in the form of a universal property of the NLHawkes family composed of intensity maps growing sufficiently fast as a function of the tension (to be defined below) and with balanced marks.

This article is organised as follows. We present the detailed mathematical formulation of the NLHawkes processes in Sec. II. In Sec. III, the NLHawkes processes are mapped onto Markovian SPDEs, whose time evolution are described by MEs. We also develop a mathematical scheme to analyse the MEs, such as the functional Kramers-Moyal (KM) expansion and system size expansion (SSE) for the diffusive limit. In Sec. IV, we study the exact solutions to NLHawkes processes with an exponential memory kernel without inhibitory effect (i.e., only the positive feedback effects are taken into account). In Sec. V and VI, we study the exact solutions of NLHawkes processes with an exponential memory kernel and in the presence of inhibitory effects (i.e., when both positive and negative feedback are considered). In Sec. VII and VIII, we present the asymptotic solutions of the NLHawkes models with an arbitrary memory in the absence and presence of inhibitory effects, respectively. Sec. IX discusses future possible extensions and progress that can derive from our present work. Sec. X concludes and is followed by nine appendices presenting detailed derivations omitted from the main text for the sake of conciseness.

For readers interested only in the overview of our results, go to Sec. II and Table I. Indeed, all our results are summarised in Table I, which maps the inputs of the model (i.e., setups) to the outputs (i.e., the resultant asymptotic PDFs).

## II. SETUP

We first introduce the mathematical notations used to define the NLHawkes model. We then review the NLHawkes processes and their applications for real data analysis of complex systems, to highlight their utility and importance in various contexts.

### A. Mathematical notation

We denote any stochastic variable  $\hat{A}$  with a hat to distinguish it from a non-stochastic real number  $A$ . The ensemble average of any stochastic variable  $\hat{A}$  is written as  $\langle \hat{A} \rangle$ . The probability density function (PDF) is denoted by  $P_t(A) := \langle \delta(A - \hat{A}(t)) \rangle$ , which characterizes the probability that  $\hat{A}(t) \in [A, A + dA]$  as  $P_t(A)dA$ . Using the PDF, the ensemble average can be rewritten as

$$\langle \hat{A}(t) \rangle := \int A P_t(A) dA. \quad (1)$$

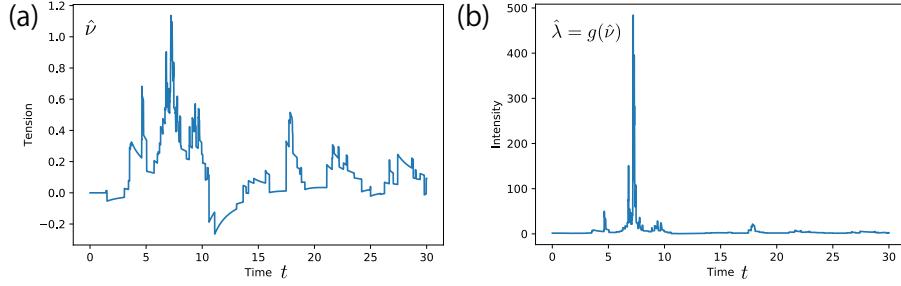


FIG. 1: A sample trajectory of the tension  $\hat{\nu}$  (left panel) and the intensity  $\hat{\lambda} = g(\hat{\nu})$  (right panel) in the NLHawkes process (3). The functions and model parameters are:  $\rho(y) = e^{-x^2/2}/\sqrt{2\pi}$ ,  $g(\nu) = e^{\beta\nu+\nu_0}$ ,  $\beta = 5$ ,  $\nu_0 = 0.5$ . The memory kernel  $h(t)$  is defined by expression (14) with  $\tilde{h}(x) = \{\eta_{\text{ini}} + c_{\text{ini}}(x - x_{\text{ini}})\}/x$  for  $x \in [x_{\text{ini}}, x_{\text{fin}}]$  and  $\tilde{h}(x) = 0$  for  $x \notin [x_{\text{ini}}, x_{\text{fin}}]$  with  $x_{\text{ini}} = 0.5$ ,  $x_{\text{fin}} = 10$ ,  $\eta_{\text{ini}} = 0.021$ ,  $c_{\text{ini}} = 0.0044$ , and we use the discrete step size  $dx = 0.0475$ .

We define the real number space by  $\mathcal{R}$ . Its nonnegative part is denoted by  $\mathcal{R}^+ := \{x \mid x \geq 0, x \in \mathcal{R}\}$ . The  $K$ -dimensional real number space is denoted as  $\mathcal{R}_K := \{(x_1, \dots, x_n) \mid x_k \in \mathcal{R} \text{ for } k = 1, 2, \dots, n\}$  and its nonnegative part is written as  $\mathcal{R}_K^+ := \{(x_1, \dots, x_K) \mid x_k \in \mathcal{R}, x_k \geq 0 \text{ for } k = 1, 2, \dots, K\}$ . We also define the functional space by  $\mathcal{S}_F$ . For example, a function  $f$  defined on  $\mathcal{R}^+$  is in the function space  $\mathcal{S}_F$ , such that  $\{f(x)\}_{x \in \mathcal{R}^+} \in \mathcal{S}_F$ .

In this paper, functionals (i.e., maps from a function space  $\mathcal{S}_F$  to a real number space  $\mathcal{R}$ ) appear to characterise the “path” probability density. For any  $\{z(x)\}_{x \in \mathcal{R}^+} \in \mathcal{S}_F$ , a functional  $f$  is denoted as  $f[z] := f[\{z(x)\}_{x \in \mathcal{R}^+}]$ . Here, the square bracket emphasizes that  $f$  is a functional (i.e., its argument is a function), but not an ordinary function.

For a stochastic variable  $\{\hat{z}(t, x)\}_{x \in \mathcal{R}^+}$  defined on a field  $x \in \mathcal{R}^+$ , the probability density functional (PDF) is written as  $P_t[z] := \langle \delta[z - \hat{z}] \rangle = P_t[\{z(x)\}_{x \in \mathcal{R}^+}]$  with the  $\delta$  functional  $\delta[z - \hat{z}] := \prod_{x \in \mathcal{R}^+} \delta(z(x) - \hat{z}(t, x))$ . Here, the PDF is defined over paths so that probability weighted quantities involve path integrals. For instance, the ensemble average is defined by

$$\langle \hat{A}(t) \rangle = \int A(t) P_t[z] \mathcal{D}z, \quad \mathcal{D}z := \prod_{x \in \mathcal{R}^+} dz(x) \quad (2)$$

where  $\mathcal{D}z$  is the path-integral volume element.

## B. Model

Let us now formulate the marked NLHawkes process studied in this paper. Let us consider an internal variable  $\hat{\nu}(t)$  that represents the total “tension” of the system, and which obeys a non-Markovian stochastic differential equation (SDE),

$$\hat{\nu}(t) = \sum_{i=1}^{\hat{N}(t)} \hat{y}_i h(t - \hat{t}_i), \quad (3a)$$

where  $\{\hat{y}_i\}_i$  is an independent and identically distributed (IID) random sequence of random numbers (“jumps”) obeying a distribution  $\rho(y)$ ,  $h(t)$  is a nonnegative memory kernel,  $\{\hat{t}_i\}_i$  is a Poisson process conditional on a time-dependent intensity  $\hat{\lambda}(t)$ , and  $\hat{N}(t)$  is the total number of events during  $[0, t)$  (called “counting process”). The jump size  $\hat{y}_i$  is called a “mark” in the point process literature. Here the intensity  $\hat{\lambda}(t)$  is assumed to be stochastic and is a nonnegative nonlinear function of the total tension  $\hat{\nu}(t)$ , defined as

$$\hat{\lambda}(t) = g(\hat{\nu}(t)) > 0. \quad (3b)$$

In this paper, we call  $g(\hat{\nu})$  the tension-intensity map or intensity function. The intensity is the probability per unit time for an event to be triggered: assuming  $\hat{N}(t) = k$ ,  $\hat{\lambda}(t)dt$  gives the probability that  $\hat{t}_{k+1} \in [t, t+dt)$  for an infinitely small time interval  $dt \rightarrow 0$ . We can rewrite Eqs. (3a) and (3b) as

$$\hat{\lambda}(t) = g \left( \sum_{i=1}^{\hat{N}(t)} \hat{y}_i h(t - \hat{t}_i) \right). \quad (3c)$$

This is the fundamental dynamical equation governing the NLHawkes processes. See Fig. 1 for a schematic trajectory.

In this article, we particularly focus on power law forms of the steady PDF of the intensity for large  $\lambda$  as

$$P_{ss}(\lambda) = \lim_{t \rightarrow \infty} \langle \delta(\lambda - \hat{\lambda}(t)) \rangle \propto \lambda^{-1-a}. \quad (4)$$

where  $a$  is the exponent of the complementary cumulative distribution function (CCDF)<sup>1</sup>.

*Remark.* Model (3) is a natural nonlinear generalisation of the conventional (linear) Hawkes process. Indeed, the LHawkes process is recovered by choosing a linear intensity function,

$$\lambda = g(\nu) = \nu + \nu_0, \quad (5)$$

assuming both  $h(t)$  and  $\nu$  are nonnegative. In contrast to the conventional Hawkes process, we do not assume nonnegativity of  $\hat{\nu}$  and  $\hat{y}_i$  for the case of general nonnegative nonlinear intensity function  $g(\hat{\nu})$ .

For the LHawkes process, the integral of the memory kernel

$$\eta := \int_0^\infty h(t) dt \quad (6)$$

is an important parameter (called the *branching ratio*) since it controls the fertility of events to trigger descendants (triggered events). Indeed, the LHawkes process is subcritical for  $\eta < 1$ , critical at  $\eta = 1$ , and supercritical for  $\eta > 1$ .

### C. Motivation and literature review

We now present a brief self-contained review of the existing literature on NLHawkes processes for statistical physics readers who may be unfamiliar with this topic. Readers interested only in our main results may skip this subsection.

NLHawkes processes were first introduced by Brémaud and Massoulié [19] in 1996, who were concerned with general conditions for the existence of the processes. Since then, there have been a few applications to seismic, financial, and neural modelling, in particular for empirical comparisons. However, beyond the derivation of general conditions for existence, obtaining analytical solutions of these models is very difficult due to the complex interplay between their nonlinear and non-Markovian structures. Only a few studies exist, such as the analysis of the stability of these processes (conditions for non-explosiveness) [19], a special solution for the ZHawkes (Zumbach Hawkes) processes with an exponential memory in the diffusive limit [28], and an asymptotic analysis for high-baseline intensity using the functional central limit theorem [29].

There are several motivations for introducing NLHawkes processes. Here we focus on two following interesting properties: (i) inhibitory effects and (ii) physical underpinning of the nonlinear tension-intensity maps. Indeed, one of the motivations for introducing NLHawkes processes is to describe inhibitory effects [20], such that previous events can produce negative feedback effects on the total tension  $\hat{\nu}$ . For simplicity, let us consider the case where the tension-intensity map  $\lambda = g(\nu)$  is an increasing function. For this setup, an event with positive mark  $\hat{y}_i > 0$  is likely to induce future events and, inversely, an event with negative mark  $\hat{y}_i < 0$  is likely to inhibit future events. This means that negative marks  $\hat{y}_i < 0$  represent inhibitory effects, while positive marks  $\hat{y}_i > 0$  represent excitatory effects.

To implement such inhibitory effects, nonlinearity in the tension-intensity map is essential because the LHawkes process cannot accommodate inhibitory effects. Indeed, if we assume an affine tension-intensity map  $g(\nu) = \nu_0 + \nu$  with nonnegative constant  $\nu_0$ ,  $\nu$  must take value larger than  $-\nu_0$  in order for the tension-intensity map to remain nonnegative. This condition requires that the mark distribution must be one-sided toward the positive direction (i.e.,  $\rho(y) = 0$  for  $y < 0$ ); otherwise,  $\nu$  takes value smaller than  $-\nu_0$  with non-zero probability and the model assumption is violated. In fact, the model cannot be defined as a negative intensity or probability density cannot be given mathematical sense.

The second nice property of NLHawkes processes is that the nonlinearity of the tension-intensity maps captures in a natural way the real mechanisms occurring in the modelled systems. Let us illustrate this point by reviewing several versions of the NLHawkes processes studied in the literature.

<sup>1</sup> We used the PDF exponent  $a_{\text{PDF}} = 1 + a_{\text{CCDF}}$  for the description of the power law relations in Refs. [16, 17], where CCDF stands for the complementary cumulative distribution function.

### 1. Example 1: seismic modelling

One of the most illustrative cases is found in the modelling of statistical seismicity. Let us regard the tension  $\hat{\nu}(t)$  as the total stress component along the fault best oriented for rupture at a given point  $\vec{r}$  in the Earth crust. Let  $\hat{t}_i$  be the time of occurrence of the  $i$ th earthquake. This earthquake creates a tensorial stress field that adds to the pre-existing stress field. Again, for our discussion, we simplify the picture by taking this stress perturbation as being a scalar, for instance the Mohr-Coulomb stress amplitude along the fault best oriented for rupture at point  $\vec{r}$ . Furthermore, we take into account the visco-elastic property of the crust, which means that a stress perturbation is progressively relaxed via a memory kernel  $h(t - \hat{t}_i)$  that tends to 0 at long times. Then, the total stress at  $\vec{r}$  is obtained as the sum of the stress perturbations created by all past earthquakes

$$\hat{\nu}(t) = \sum_{i=1}^{N(t)} \hat{y}_i h(t - \hat{t}_i). \quad (7)$$

Note that the marks  $\hat{y}_i$  can be positive (resp. negative), corresponding to the  $i$ th earthquake bringing the point  $\vec{r}$  closer to (resp. further away from) failure. The former case is the most intuitive and represents the stress load on  $\vec{r}$  due to the redistribution of forces by the earthquake fault slip in its neighborhood, especially close to its fault tips and in its stress lobes of positive influence. The later case is known as “stress shadow” [23] and is associated with the tensorial nature of the stress disturbances induced by an earthquake. Given the stochasticity in the distribution of earthquake sizes, in their positions and orientations, the marks  $\hat{y}_i$  are stochastic variables. Given the total stress (tension) (7), the next ingredient is to recognise that mechanical rupture and earthquakes are thermally activated with an effective inverse temperature  $\beta$  that is renormalised via the quenched heterogeneity of the medium [30–32]. Then, the probability for the next earthquake to occur is given by the Arrhenius formula, thus formulating the intensity  $\lambda$  as a decreasing exponential function  $e^{-\beta \Delta E(t)}$  of the energy barrier  $\Delta E(t)$  for nucleation. The key point is to approximate the energy barrier as a decreasing affine function of the stress field:  $\Delta E(t) = (E - \hat{\nu}(t))$ , where  $E$  is a constant. Putting all together, this yields

$$\lambda(t) = \lambda_0 e^{\beta \hat{\nu}(t)}. \quad (8)$$

We finally obtain the NLHawkes with an exponential intensity

$$\hat{\lambda}(t) = g \left( \sum_{i=1}^{N(t)} \hat{y}_i h(t - \hat{t}_i) \right), \quad g(\hat{\nu}) := \lambda_0 e^{\beta \hat{\nu}}. \quad (9)$$

In addition, given that the prediction of earthquake magnitudes is empirically very difficult (while the short-term prediction of their rates is rather possible [33]), it is a plausible assumption that the marks are drawn independently of the current tension  $\hat{\nu}(t)$ .

In this simplified presentation, we have restricted our attention to the temporal version of the general formulation, which is known as the multifractal stress activation (MSA) model [21, 22] and involves space in addition to time in the formulation of the tension and intensity. It is remarkable that both inhibitory effects and nonlinear intensity function appear naturally for this system, as the result of the random stress perturbations induced by earthquakes and from the Arrhenius law (renormalised by quenched disorder), respectively.

It should be noted that Refs. [21, 22] offered only an approximate scaling theory to derive magnitude dependent Omori law exponents and that no analytical results exist for the MSA model or for its temporal-only version (9).

### 2. Example 2: financial modelling

Ref. [34] is one of the very first uses in finance of the LHawkes process (in its bivariate form) in order to model the joint dynamics of trades and mid-price changes of the NYSE. Ref. [35] provided the first quantitative framework using the LHawkes process to study and quantify the level of endogeneity (or “reflexivity”) of market fluctuations. The basic idea is that trades and price changes are analysed by investors (humans or machines) as one of the useful information channels to improve trading decisions, on the basis (or belief) that past actions reveal intentions and that there is a persistence in price trends, volume, volatility and more generally of trading activity. In this sense, the self-exciting Hawkes process is a natural candidate to model the point processes of discrete trades and mid-price changes [41].

As an improved model, a nonlinear version of the Hawkes process was introduced by Blanc, Donier, and Bouchaud [28], where the intensity dynamics is given by a quadratic extension to the standard Hawkes process,

$$\hat{\lambda}(t) = \lambda_0 + \int_{-\infty}^t L(t-s) \hat{\xi}_{\rho(y); \hat{\lambda}(s)}^P(s) ds + \int_{-\infty}^t ds \int_{-\infty}^t du K(t-s, t-u) \hat{\xi}_{\rho(y); \hat{\lambda}(s)}^P(s) \hat{\xi}_{\rho(y); \hat{\lambda}(u)}^P(u) \quad (10)$$

with  $\rho(y) = (1/2)[\delta(y-1) + \delta(y+1)]$  and the term  $\hat{\xi}_{\rho(y); \hat{\lambda}(s)}^P(s)$  is the compound Poisson process with intensity  $\hat{\lambda}(t)$  and jump size distribution  $\rho(y)$  as defined below by expression (14c). This model is called the quadratic Hawkes (QHawkes) processes and has been theoretically analyzed in Ref. [28].

Since this model is nonlinear and non-Markovian, its systematic analysis is difficult and only limited results are available. However, by assuming  $K(t, s) = h(s)h(t)$  and  $L(t) = 0$ , this model reduces to a simpler NLHawkes process,

$$\hat{\lambda}(t) = g \left( \int_{-\infty}^t h(t-s) \hat{\xi}_{\rho(y); \hat{\lambda}(s)}^P(s) ds \right) = g \left( \sum_{i=1}^{\hat{N}(t)} \hat{y}_i h(t - \hat{t}_i) \right), \quad g(\nu) := \lambda_0 + \nu^2, \quad (11)$$

where we have used  $\hat{\xi}_{\rho(y); \hat{\lambda}(s)}^P(s) = \sum_{i=1}^{\hat{N}(s)} \hat{y}_i \delta(s - \hat{t}_i)$ . This NLHawkes process is a special case of the Zumbach Hawkes (ZHawkes) process, without the Hawkes feedback. While the ZHawkes process is simpler than the QHawkes, it is still difficult to solve analytically. Therefore, the analysis in Ref. [28] focused on the special case of an exponential memory  $h(t) = (\eta/\tau)e^{-t/\tau}$  and considered the diffusive limit<sup>2</sup>. For this special case, the steady-state PDF of the intensity obeys a power law with non-universal exponent

$$P_{ss}(\lambda) \propto \lambda^{-1-a}, \quad a = \frac{1}{2} + \frac{1}{2n_Z} \quad (12)$$

with a constant  $n_Z$  called Zumbach norm (see Ref. [28] for details). It is remarkable that a power law relation (12) appears even for short memory kernels without introducing any power law distributions. To the best of our knowledge, this special solution was the only available analytical solution for a NLHawkes process before our work [27].

One of the main claims in Ref. [28] is that the power law relation (12) provides a validation step supporting the relevance of the QHawkes process for financial data analyses, because it matches the empirical power law price-change distribution, which is well-known stylised fact in market microstructure. From this viewpoint, the authors of Ref. [28] claim that the QHawkes process is a minimal generalisation beyond the LHawkes process that is essentially needed to account for empirical facts.

### 3. Example 3: the self-excited multifractal model

It is also useful to mention the self-excited multifractal model [36], which is not per se a point process but got its inspiration from self-excited point processes, the concept of reflexivity [37], the multifractal random walk model [38] and its generalisations [39, 40]. Reminiscent of a NLHawkes model with a much stronger nonlinearity than quadratic, the self-excited multifractal model is defined such that the amplitudes of the increments of the process are expressed as exponentials of a long memory of past increments:

$$dX(t) = \sigma \exp \left\{ -\frac{1}{\sigma} \int_{-\infty}^t h(t-t') dX(t') \right\} dW(t), \quad (13)$$

where  $dW(t)$  is the increment of the regular Wiener noise process,  $h(t)$  is a memory kernel function and  $\sigma$  controls the amplitude of the noise as well as the dimension and scale of  $X(t)$ . Interpreting  $dX(t)$  as a log-return of a financial price, the self-excited multifractal process recovers all the standard stylised facts documented in empirical financial time series. The exploration of the links between the self-excited multifractal model and the exponential NLHawkes process is left for the future.

---

<sup>2</sup> They call their analysis the low-frequency asymptotics, taking the long-time limit and a constant endogeneity rescaling. This asymptotic method is essentially equivalent to the diffusive limit in the framework of the system size expansion (SSE), a traditional asymptotic analyses developed for statistical physics, which is formulated in Sec. III E.

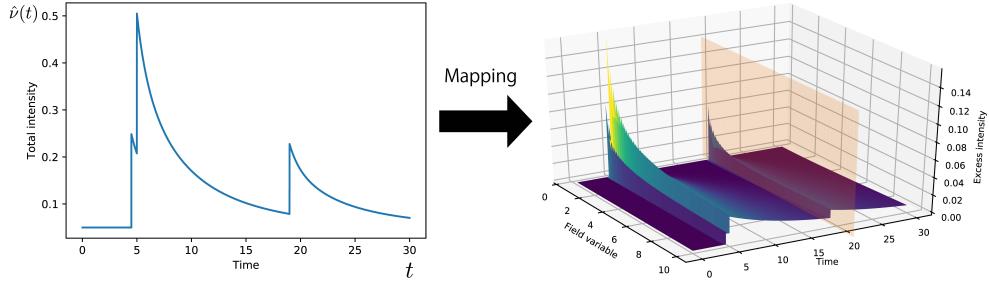


FIG. 2: Schematics of the Markovian embedding: the original non-Markovian one-dimensional dynamics (described by the SDE (3); left panel) is mapped onto the Markovian field dynamics (described by the SPDE (14); right panel).

#### 4. Goal of this study: solutions for various nonlinear Hawkes processes

The above summaries highlight the fact that analytical solutions for NLHawkes processes have not been obtained yet, except for special cases (such as the ZHawkes case with exponential memory in the diffusive limit). In this context, our goal is to systematically classify NLHawkes processes and then provide analytical (both exact and asymptotic) solutions for various NLHawkes processes, in particular for the steady-state intensity PDF  $P_{ss}(\lambda)$ . All our results are summarised as Table I, with the mapping between the inputs of the model (i.e., setups) to the outputs (i.e., the resultant asymptotic PDFs).

### III. MASTER EQUATION FORMULATION

In this section, we introduce an analytical framework for the general NLHawkes process based on the field MEs. We first provide a Markovian mapping from the original non-Markovian NLHawkes process to a Markovian SPDE. We then derive the corresponding field ME for any memory kernel, which is shown to simplify for the special case of an exponential memory kernel. We next develop two useful tools that have a long tradition in the history of physical stochastic processes: the Kramers-Moyal (KM) expansion and the system size expansion (SSE) for the diffusive limit. The field ME is then shown to reduce to the functional Fokker-Planck equations (FPEs) for a special case.

#### A. Mapping to Markovian SPDEs

Following Ref. [16], let us present the mapping from the original non-Markovian stochastic process (3) to Markovian SPDEs. Let us decompose the total tension  $\hat{\nu}(t)$  and the memory kernel  $h(t)$  as continuous sums

$$h(t) = \int_0^\infty dx \tilde{h}(x) e^{-t/x}, \quad \hat{\nu}(t) = \int_0^\infty dx \hat{z}(t, x). \quad (14a)$$

The intuition behind this decomposition is that the memory kernel is decomposed into a continuous sum of exponential terms with amplitude  $\tilde{h}(x)$ . This then suggests to use  $x$  as an auxiliary field  $x \in (0, \infty)$ , and then to decompose the tension as a continuous sum over the ‘‘excess tensions’’  $\hat{z}(t, x)$ . The excess tensions are assumed to satisfy the following SPDEs:

$$\frac{\partial \hat{z}(t, x)}{\partial t} = -\frac{\hat{z}(t, x)}{x} + \tilde{h}(x) \hat{\xi}_{\rho(y); \hat{\lambda}(t)}^{\text{CP}}, \quad \hat{\lambda}(t) = G[\hat{z}] := g \left( \int_0^\infty dx \hat{z}(t, x) \right) \quad (14b)$$

under the initial condition  $\hat{z}(t = 0, x) = 0$ . The term  $\hat{\xi}_{\rho(y); \hat{\lambda}(t)}^{\text{CP}}$  is the compound Poisson process with intensity  $\hat{\lambda}(t)$  and jump size distribution  $\rho(y)$ :

$$\hat{\xi}_{\rho(y); \hat{\lambda}(t)}^{\text{CP}} = \sum_{i=1}^{\hat{N}(t)} \hat{y}_i \delta(t - \hat{t}_i), \quad (14c)$$

which means that the random marks  $y_i$  obeys the distribution  $\rho(y)$ .

This mapping can be schematically illustrated as shown in Fig. 2: the original dynamics is one-dimensional, governed by the SDE (3). In this low-dimensional representation, the dynamics is non-Markovian. However, by applying the Markovian embedding, we can construct an infinite-dimensional Markovian dynamics governed by the SPDE (14) by adding sufficiently many auxiliary variables  $\hat{z}(t, x)$ .

#### *Proof of equivalence.*

The SPDE (14b) together with the decomposition formula (14a) is equivalent to the original marked NLHawkes process (3). Indeed, the formal solution of Eq. (14b) is given by

$$\hat{z}(t, x) = \tilde{h}(x) \int_0^t dt' e^{-(t-t')/x} \hat{\xi}_{\rho(y); \hat{\lambda}(t)}^{\text{CP}} = \sum_{i=1}^{\hat{N}(t)} \hat{y}_i \tilde{h}(x) \int_0^t dt' e^{-(t-t')/x} \delta(t - \hat{t}_i) = \sum_{i=1}^{\hat{N}(t)} \hat{y}_i \tilde{h}(x) e^{-(t-\hat{t}_i)/x}, \quad (15)$$

leading to

$$\hat{\nu}(t) = \sum_{i=1}^{\hat{N}(t)} \hat{y}_i \int_0^\infty dx \tilde{h}(x) e^{-(t-\hat{t}_i)/x} = \sum_{i=1}^{\hat{N}(t)} \hat{y}_i h(t - \hat{t}_i). \quad (16)$$

It is noteworthy that this derivation does not make explicit reference to the definition of  $\hat{\lambda}(t) = g(\hat{\nu}(t))$  and is independent of the specific function  $g(\nu)$ .

## B. Field master equation

In this subsection, we study the functional ME corresponding to the SPDE (14b). The field ME of the PDF  $P_t[z]$  is given by

$$\frac{\partial P_t[z]}{\partial t} = (\mathcal{L}_{\text{adv}} + \mathcal{L}_{\text{jump}}) P_t[z] \quad (17a)$$

with advective and jump Liouville operators

$$\mathcal{L}_{\text{adv}} P_t[z] := \int_0^\infty dx \frac{\delta}{\delta z(x)} \left( \frac{z(x)}{x} P_t[z] \right), \quad (17b)$$

$$\mathcal{L}_{\text{jump}} P_t[z] := \int_{-\infty}^\infty dy \rho(y) G[z - y \tilde{h}] P_t[z - y \tilde{h}] - G[z] P_t[z]. \quad (17c)$$

In this paper, we provide various analytical exact or asymptotic solutions of (17).

### 1. Derivation

It is useful to provide a derivation of the field ME (17) via a discrete approach, which gives a sound mathematical interpretation and control of the functional derivatives [42]. Let us consider the case of the memory kernel composed of a discrete sum of  $K$  exponentials (which we refer to as  $K$ -exponentials),

$$h(t) = \sum_{k=1}^K \tilde{h}_k e^{-t/\tau_k}. \quad (18a)$$

The NLHawkes process (3) together with the  $K$ -exponentials (18a) can be mapped onto a Markovian equation by introducing  $\hat{z} := (\hat{z}_1, \dots, \hat{z}_K)$ ,

$$\frac{d\hat{z}_k(t)}{dt} = -\frac{\hat{z}_k(t)}{\tau_k} + \tilde{h}_k \hat{\xi}_{\rho(y); \hat{\lambda}(t)}^{\text{CP}}, \quad \hat{\lambda}(t) = g(\hat{\nu}(t)), \quad \hat{\nu}(t) := \sum_{k=1}^K \hat{z}_k(t), \quad (18b)$$

which is parallel to the Markovian embedding procedure for Eqs. (14). We introduce the following function  $G$ , which will be convenient for future developments,

$$G(\hat{\mathbf{z}}) := g(\hat{\nu}) = g\left(\sum_{k=1}^K \hat{z}_k\right). \quad (18c)$$

The ME for the SDE (18) is derived as follows. Let us consider an arbitrary function  $f(\hat{\mathbf{z}})$  and its time-evolution  $df(\hat{\mathbf{z}}(t)) := f(\hat{\mathbf{z}}(t+dt)) - f(\hat{\mathbf{z}}(t))$  during  $[t, t+dt]$ :

$$df(\hat{\mathbf{z}}) = \begin{cases} -\sum_{k=1}^K \frac{\hat{z}_k}{\tau_k} \frac{\partial f(\hat{\mathbf{z}})}{\partial \hat{z}_k} dt & (\text{No jump during } [t, t+dt]: \text{ probability} = 1 - \hat{\lambda}(t)dt) \\ f(\hat{\mathbf{z}} + \hat{y}\tilde{\mathbf{h}}) - f(\hat{\mathbf{z}}) & (\text{Jump in } [t, t+dt] \text{ with } \hat{y} \in [y, y+dy]: \text{ probability} = \hat{\lambda}(t)\rho(y)dtdy) \end{cases} \quad (19)$$

with  $\tilde{\mathbf{h}} := (\tilde{h}_1, \dots, \tilde{h}_K)$ . By taking the ensemble average of both sides over realisations of the excess tensions  $\hat{\mathbf{z}} := (\hat{z}_1, \dots, \hat{z}_K)$ , we obtain

$$\langle df(\hat{\mathbf{z}}) \rangle = \left\langle -\sum_{k=1}^K \frac{\hat{z}_k}{\tau_k} \frac{\partial f(\hat{\mathbf{z}})}{\partial \hat{z}_k} dt \right\rangle + \int_{-\infty}^{\infty} dy \rho(y) \left\langle (f(\hat{\mathbf{z}} + \hat{y}\tilde{\mathbf{h}}) - f(\hat{\mathbf{z}})) dt \right\rangle, \quad (20)$$

which is equivalent to

$$\begin{aligned} \int_{-\infty}^{\infty} dz f(z) \frac{\partial P_t(z)}{\partial t} &= \int_{-\infty}^{\infty} dz P_t(z) \left[ -\sum_{k=1}^K \frac{z_k}{\tau_k} \frac{\partial f(z)}{\partial z_k} + \int_{-\infty}^{\infty} dy \rho(y) G(z) (f(z + y\tilde{\mathbf{h}}) - f(z)) \right] \\ &= \int_{-\infty}^{\infty} dz f(z) \left[ \sum_{k=1}^K \frac{\partial}{\partial z_k} \frac{z_k}{\tau_k} P_t(z) + \int_{-\infty}^{\infty} dy \rho(y) [G(z - y\tilde{\mathbf{h}}) P(z - y\tilde{\mathbf{h}}) - G(z) P(z)] \right], \end{aligned} \quad (21)$$

by using the following relation

$$\langle df(\hat{\mathbf{z}}) \rangle = \langle f(\hat{\mathbf{z}}(t+dt)) - f(\hat{\mathbf{z}}(t)) \rangle = \int_{-\infty}^{\infty} dz f(z) P_{t+dt}(z) - \int_{-\infty}^{\infty} dz f(z) P_t(z) = dt \int_{-\infty}^{\infty} dz f(z) \frac{\partial P_t(z)}{\partial t} + O(dt^2). \quad (22)$$

Since Eq. (21) is an identity holding for any  $f(z)$ , we obtain the corresponding ME

$$\frac{\partial P_t(z)}{\partial t} = \sum_{k=1}^K \frac{\partial}{\partial z_k} \frac{z_k}{\tau_k} P_t(z) + \int_{-\infty}^{\infty} dy \rho(y) [G(z - y\tilde{\mathbf{h}}) P(z - y\tilde{\mathbf{h}}) - G(z) P(z)], \quad (23)$$

where we have performed an integration par part and have used the variable transformation  $z + y\tilde{\mathbf{h}} \rightarrow z$ .

We then proceed with the continuous limit for the memory kernel. We first rewrite

$$h(t) = \sum_{k=1}^K \tilde{h}_k e^{-t/\tau_k} \rightarrow \sum_{k=1}^K dx \tilde{h}(x_k) e^{-t/x_k}, \quad \hat{\nu}(t) = g\left(\sum_{k=1}^K \hat{z}_k(t)\right) \rightarrow g\left(\sum_{k=1}^K dx \hat{z}(t, x_k)\right) \quad (24)$$

for the formal replacement

$$\tau_k \rightarrow x_k, \quad \tilde{h}_k \rightarrow \tilde{h}(x_k)dx, \quad \hat{z}_k \rightarrow \hat{z}(t, x_k)dx \quad (25)$$

obtained by introducing the lattice interval  $dx$  and  $x_k := kdx$ . By introducing the formal functional derivative and integration for the limit  $K \rightarrow \infty$  and  $dx \rightarrow 0$

$$\frac{\delta}{\delta z(x_k)}[\dots] := \lim_{dx \rightarrow 0} \frac{1}{dx} \frac{\partial}{\partial z(x_k)}[\dots], \quad \int_0^\infty dx [\dots] := \lim_{dx \rightarrow 0} \sum_{k=1}^K dx [\dots], \quad (26)$$

we obtain

$$\frac{\partial P_t[z]}{\partial t} = \int_0^\infty dx \frac{\delta}{\delta z(x)} \frac{z(x)}{x} P_t[z] + \int_{-\infty}^{\infty} dy \rho(y) G[z - y\tilde{\mathbf{h}}] P_t[z - y\tilde{\mathbf{h}}] - G[z] P_t[z] \quad (27)$$

and

$$h(t) = \int_0^\infty dx \tilde{h}(x) e^{-t/x}, \quad \hat{\nu}(t) = g\left(\int_0^\infty dx \hat{z}(t, x)\right), \quad (28)$$

which is equivalent to Eq. (17) (see Appendix A for the definition of the Dirac delta function and the functional derivative). See Appendix B for another derivation based on direct manipulation of functional derivatives.

## 2. Mathematical remark

Master (or FP) equations based on functional derivatives often appear in the description of SPDEs, such as for stochastic chemical reactions [42]. While this continuous description is a useful tool for formal calculations, unfortunately, its mathematical foundation has not been established yet. Indeed, one can easily observe that there is the potential problem of encountering a divergence, such as  $[\delta/\delta z(x)]z(x)P[z] = \delta(0)P[z] + z(x)\delta P[z]/\delta z(x)$ . This problem might be serious for nonlinear SPDEs even for physical observables (see the divergence problem of nonlinear stochastic chemical reaction; Chapter 13.3.3 in [42]), while it might not be for linear SPDE. One can find that this divergence is not serious for the LHawkes processs and the generalised Langevin equation [16] at least in understanding physical observables. Remarkably, for the generalised Langevin equation, this divergence problem is essentially the same as the one encountered in quantum field theory and can be renormalised in the same manner with which the divergence problem of the zero-point energy is solved in quantum electrodynamics. We note that, in the case of the NLHawkes process, the SPDE (14) itself is fortunately linear, while the intensity function  $g(\dot{\nu})$  is nonlinear.

To avoid such mathematically delicate issues, our strategy is to follow a safer interpretation that follows Ref. [42]: we regard the field ME (17) (or the FPE (48)) as a formal limit of the discrete ME (23). If we encounter a potential problem of divergence in Eq. (17), we return to the discrete ME (23) to proceed with the calculations, and then come back to its formal limit (17). We confirm that our main results hold for the general discrete cases (18) and we then generalise them to the continuous limit.

## C. Special case: exponential memory kernel

Let us here focus on the simplest case of the single exponential memory kernel:

$$h(t) = \frac{\eta}{\tau} e^{-t/\tau}, \quad (29)$$

or equivalently

$$\tilde{h}(x) = \frac{\eta}{\tau} \delta(x - \tau) \quad (30)$$

with positive real numbers  $\eta$  and  $\tau$ . Consistent with the definition (6), parameter  $\eta$  is the branching ratio. This special case is easier to analyse analytically, since the functional ME (17) reduces to the ME for a PDF of the total tension  $\nu$ ,

$$\frac{\partial P_t(\nu)}{\partial t} = \frac{1}{\tau} \frac{\partial}{\partial \nu} [\nu P_t(\nu)] + \int dy \rho(y) g(\nu - \eta y/\tau) P_t(\nu - \eta y/\tau) - g(\nu) P_t(\nu). \quad (31)$$

## D. Functional Kramers-Moyal expansion

One of the standard analytical prescriptions to analyse MEs is the KM expansion. The KM expansion was historically introduced for a formal validation of the FP description from MEs. This formal expansion was criticised by van Kampen due to its ambiguous validity as an asymptotic series. Later, van Kampen developed a mathematically sophisticated formulation in the form of the SSE [43]. Let us present a sound formulation of the KM functional expansion for the field ME, which will be utilised for a further generalisation of the SSE in Sec III E.

### 1. Exponential memory case

To first present the key idea, let us focus on the exponential-memory case (29). By considering the expansion

$$\int dy \rho(y) g(\nu - \eta y/\tau) P_t(\nu - \eta y/\tau) = \sum_{k=0}^{\infty} \frac{\alpha_k}{k!} \left(-\frac{\eta}{\tau}\right)^k \frac{\partial^k}{\partial \nu^k} g(\nu) P_t(\nu), \quad \alpha_k := \int_{-\infty}^{\infty} dy \rho(y) y^k, \quad (32)$$

the ME (31) can be rewritten as

$$\frac{\partial P_t(\nu)}{\partial t} = \frac{1}{\tau} \frac{\partial}{\partial \nu} [\nu P_t(\nu)] + \sum_{k=1}^{\infty} \frac{\alpha_k}{k!} \left(-\frac{\eta}{\tau}\right)^k \frac{\partial^k}{\partial \nu^k} g(\nu) P_t(\nu). \quad (33)$$

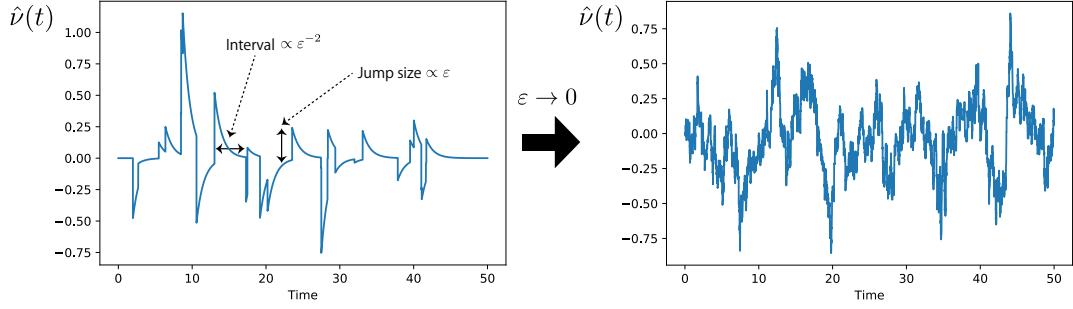


FIG. 3: Schematic trajectory of the NLHawkes process in the diffusive limit. While the trajectory is composed of sparse jumps for large  $\varepsilon$  ( $\varepsilon = 1.0$ ; left panel), the trajectory is composed of many small jumps for small  $\varepsilon$  and becomes approximately continuous ( $\varepsilon = 0.1$ ; right panel). The trajectories were generated by assuming  $h(t) = (\eta/\tau)e^{-t/\tau}$ ,  $\rho_\varepsilon(y) = e^{-y^2/(2\varepsilon^2)}/\sqrt{2\pi\varepsilon^2}$ , and  $g(\nu) = \lambda_0 + k\nu^2$  with  $\eta = 0.5$ ,  $\tau = 1$ ,  $k = 1$ , and  $\lambda_0 = 0.5$ . The discrete time step is  $\Delta t = 10^{-4}$ .

This is the Kramers-Moyal (KM) expansion for the ME (31) for this special case. We have assumed that all the KM coefficients  $\{\alpha_k\}_{k \geq 1}$  are finite, which excludes some singular classes of mark distributions (e.g., power law mark distributions).

## 2. General cases

The above formulation can be generalised by considering the functional Taylor expansion (see Appendix A)

$$\int_{-\infty}^{\infty} dy \rho(y) G[z - y \tilde{h}] P_t[z - y \tilde{h}] = \sum_{k=0}^{\infty} \frac{\alpha_k}{k!} \left( - \int_0^{\infty} dx y \tilde{h}(x) \frac{\delta}{\delta z(x)} \right)^k G[z] P_t[z] \quad (34)$$

with KM coefficients defined by

$$\alpha_k := \int_{-\infty}^{\infty} dy \rho(y) y^k, \quad (35)$$

and assuming that all the KM coefficients  $\{\alpha_k\}_{k \geq 1}$  are finite. Using this relation, the field ME can be rewritten as

$$\frac{\partial P_t[z]}{\partial t} = \int_0^{\infty} dx \frac{\delta}{\delta z(x)} \left( \frac{z(x)}{x} P_t[z] \right) + \sum_{k=1}^{\infty} \frac{\alpha_k}{k!} \left( - \int_0^{\infty} dx \tilde{h}(x) \frac{\delta}{\delta z(x)} \right)^k G[z] P_t[z]. \quad (36)$$

## E. Diffusive limit: system size expansion

We next consider the diffusive limit for the mark distribution according to the SSE, by assuming (i) that the mark distribution is symmetric,

$$\rho(y) = \rho(-y), \quad (37)$$

i.e., this is the case where inhibitory effects are as prevalent as excitatory effects. This situation will be further studied in detail in Sec. V. This model is essentially different from the positive mark cases (i.e.,  $\rho(y) = 0$  for  $y \leq 0$ ) because both positive and negative feedback effects occur with the same probability. For instance, such assumption is natural for seismic models as the stress perturbations induced by earthquakes present indeed this symmetry (which has a complex tensorial spatial rendering, see for instance [44]). With this symmetry condition, all the odd-order KM coefficients are zero:  $\alpha_{2k+1} = 0$  for nonnegative integer  $k$ .

As the second assumption (ii), let us introduce a small parameter  $\varepsilon > 0$  scaling the jump size in the original Hawkes process,

$$\hat{y}_i := \varepsilon \hat{Y}_i \quad (38)$$

or equivalently,

$$\hat{\lambda}(t) = g \left( \varepsilon \sum_{i=1}^{\hat{N}(t)} \hat{Y}_i h(t - \hat{t}_i) \right). \quad (39)$$

In other words, each jump size  $\hat{y}_i$  is assumed proportional to a small parameter  $\varepsilon$  and thus the rescaled jump size  $\hat{Y}_i$  appears as the renormalised jump size independent of  $\varepsilon$  (see Fig. 3, left panel). For explicit clarification of the  $\varepsilon$  dependence, we denote below the original mark distribution  $\rho(y)$  by  $\rho_\varepsilon(y)$ . This assumption can be interpreted as a weak coupling limit between the system and the noise term. Considering the Jacobian relation (i.e., preservation of probability)

$$\rho_\varepsilon(y) dy = \tilde{\rho}(Y) dY \quad (40)$$

with the scaled jump-size distribution  $\tilde{\rho}(Y)$ , the above scaling assumption on the trajectory level is equivalent to that for the mark distribution

$$\rho_\varepsilon(y) := \frac{1}{\varepsilon} \tilde{\rho}\left(\frac{y}{\varepsilon}\right). \quad (41)$$

We note that this scaling assumption is equivalent to the system size expansion (SSE, or often called the  $\Omega$  expansion), which was originally introduced by Van Kampen for a systematic derivation of the Langevin equation within this kinetic theory (see the textbook by Van Kampen [43] and a review [45] including recent extended SSEs [46, 47]). With this assumption, the KM coefficients  $\alpha_k$  have the following scaling

$$\alpha_k = \begin{cases} \varepsilon^k \tilde{\alpha}_k & (\text{even } k) \\ 0 & (\text{odd } k) \end{cases} \quad (42)$$

with  $\varepsilon$ -independent KM coefficient  $\tilde{\alpha}_k := \int_{-\infty}^{\infty} Y^k \tilde{\rho}(Y) dY$ .

In the weak coupling limit  $\varepsilon \rightarrow 0$ , each jump size is very small and thus the noise term becomes irrelevant if the intensity is constant. To keep the effect of the noise minimally relevant, let us take the diffusive limit by increasing the intensity as a function of  $\varepsilon$  (i.e.,  $g(\nu)$  is a function of  $\varepsilon$ ). As the third assumption (iii), therefore, we assume that the intensity function satisfies the diffusive scaling

$$g(\nu) = \frac{1}{\varepsilon^2} \tilde{g}(\nu) \quad (43)$$

with  $\varepsilon$ -independent intensity function  $\tilde{g}(\nu)$  (see Fig. 3, right panel). In other words, the model is explicitly written in the following form:

$$\hat{\lambda}(t) = \frac{1}{\varepsilon^2} \tilde{g} \left( \varepsilon \sum_{i=1}^{\hat{N}(t)} \hat{Y}_i h(t - \hat{t}_i) \right). \quad (44)$$

These three assumptions enable us to rewrite the field ME exactly in terms of the functional FPE in the diffusive limit  $\varepsilon \rightarrow 0$  (see Fig. 3, right panel) as we will elaborate in the following.

It is interesting to mention a report by Gao and Zhu [29], where a similar but still different form of asymptotics is studied by assuming a one-sided mark distribution  $\rho(y) = \delta(y - 1)$  and a scaling for the tension-intensity map  $g(\nu) = (1/\varepsilon)\tilde{g}(\nu)$  for a nonlinear version of the large baseline intensity regime for the LHawkes processes [48]. For this setup, the trajectory fluctuates around a deterministic trajectory and thus shows quite different phenomenology.

### 1. Exponential memory case

To understand the main ingredients of our calculations, let us first focus on the exponential-memory case (29). The KM expansion can be rewritten as

$$\frac{\partial P_t(\nu)}{\partial t} = \frac{1}{\tau} \frac{\partial}{\partial \nu} [\nu P_t(\nu)] + \sum_{k=1}^{\infty} \varepsilon^{2k-2} \frac{\tilde{\alpha}_{2k}}{(2k)!} \left(-\frac{\eta}{\tau}\right)^{2k} \frac{\partial^{2k}}{\partial \nu^{2k}} \tilde{g}(\nu) P_t(\nu). \quad (45)$$

By taking the diffusive limit  $\varepsilon \rightarrow 0$  (Fig. 3, right panel), we obtain the exact FPE

$$\frac{\partial P_t(\nu)}{\partial t} = \frac{1}{\tau} \frac{\partial}{\partial \nu} [\nu P_t(\nu)] + D \frac{\partial^2}{\partial \nu^2} \tilde{g}(\nu) P_t(\nu), \quad D := \frac{\tilde{\alpha}_2 \eta^2}{2\tau^2}. \quad (46)$$

We note that this FPE is equivalent to an Itô process described by

$$\frac{d\hat{\nu}}{dt} = -\frac{\hat{\nu}}{\tau} + \sqrt{2D\tilde{g}(\hat{\nu})} \cdot \hat{\xi}^G \quad (47)$$

with the Itô product “.” and the standard white Gaussian noise  $\hat{\xi}^G$ , satisfying  $\langle \hat{\xi}^G \rangle = 0$  and  $\langle \hat{\xi}^G(t)\hat{\xi}^G(t') \rangle = \delta(t-t')$ .

## 2. General cases

The above formulation can be extended for the field ME. Indeed, we obtain the following exact functional FPE

$$\frac{\partial P_t[z]}{\partial t} = \int_0^\infty dx \frac{\delta}{\delta z(x)} \left( \frac{z(x)}{x} P_t[z] \right) + \int_0^\infty dx \int_0^\infty dx' D(x, x') \frac{\delta^2}{\delta z(x) \delta z(x')} \tilde{G}[z] P_t[z] \quad (48)$$

with coefficient

$$D(x, x') := \frac{\alpha_2}{2} \tilde{h}(x) \tilde{h}(x'). \quad (49)$$

The functional FPE (48) implies that the stochastic dynamics finally reduces to

$$\frac{\partial \hat{z}(t, x)}{\partial t} = -\frac{\hat{z}(t, x)}{x} + \sqrt{2\tilde{G}[\hat{z}]} \cdot \hat{\xi}^G(t; x) \quad (50)$$

for the diffusive limit (Fig. 3, right panel) with the white Gaussian noise satisfying

$$\langle \hat{\xi}^G(t; x) \rangle = 0, \quad \langle \hat{\xi}^G(t; x) \hat{\xi}^G(t'; x') \rangle = D(x, x') \delta(t-t'). \quad (51)$$

## F. Laplace transformation

Here, we introduce the relevant notations for the Laplace transformation. We first define the  $K$ -dimensional Laplace transformation as

$$\mathcal{L}_K[f(z); s] := \int_0^\infty d\mathbf{z} e^{-s \cdot \mathbf{z}} f(\mathbf{z}), \quad s \in \mathcal{R}_K^+. \quad (52)$$

In a parallel manner, the Laplace transformation in the function space can be defined as a straightforward generalisation as follows:

$$\mathcal{L}_{\text{path}}[f[z]; s] := \int_0^\infty \mathcal{D}z e^{-\int_0^\infty ds(x) z(x)} f[z], \quad s \in \mathcal{S}_F. \quad (53)$$

We note that this Laplace transformation is a kind of path integral.

## IV. SOLUTION 1: EXPONENTIAL MEMORY KERNEL WITH ONE-SIDED MARK DISTRIBUTION AND WITH LINEAR AND RAMP INTENSITY MAPS

In this section, we focus on exact solutions for the simplest case with the exponential memory kernel (29), whose dynamics is characterised by a simple ME (31). In particular, we here assume that all the marks are positive  $\hat{y} > 0$ , implying the absence of inhibitory effects.

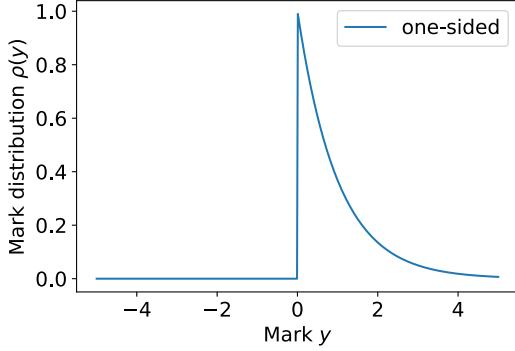


FIG. 4: Schematic of the one-sided exponential mark distribution (54) with  $y^* = 1$

#### A. Exact solutions for one-sided exponential jump

Let us consider the case with the exponential memory kernel (29) and with the one-sided exponential jump size<sup>3</sup> (see Fig. 4)

$$\rho(y) := \frac{1}{y^*} e^{-y/y^*} \Theta(y), \quad (54)$$

whose ME is known to be exactly tractable due to its special form [50]. We assume  $y^* = 1$  without losing generality because the scale can be absorbed into the branching ratio  $\eta$ . Since both memory kernel and jump size are nonnegative, the inhibitory effects are absent in this model. Interestingly, even this simple model can exhibit nontrivial steady-state distribution functions of intensities resulting from the nonlinearity of the tension-intensity map  $g(\hat{\nu})$ . This case is special because the exact steady solution to the ME (31) is available. In the steady state, the exact steady solution is given by

$$P_{ss}(\nu) = \frac{\nu^{-1}}{Z} \exp \left[ -c\nu + \tau \int \frac{g(\nu)}{\nu} d\nu \right], \quad (55)$$

with

$$c := \frac{\tau}{\eta} \quad (56)$$

and with a normalisation constant given by

$$Z := \int_0^\infty d\nu \nu^{-1} \exp \left[ -c\nu + \tau \int \frac{g(\nu)}{\nu} d\nu \right]. \quad (57)$$

*Derivation.*

By utilising the following identity (see Appendix C for the technical derivation),

$$\left( 1 + \frac{1}{c} \frac{\partial}{\partial \nu} \right) \int_0^\infty dy e^{-y} g(\nu - \eta y/\tau) P_t(\nu - \eta y/\tau) = g(\nu) P_t(\nu), \quad (58)$$

we can rewrite the ME as

$$\frac{\partial P_t(\nu)}{\partial t} = \frac{1}{\tau} \partial_\nu [\nu P_t(\nu)] - \frac{\partial_\nu/c}{1 + \partial_\nu/c} g(\nu) P_t(\nu) \quad (59)$$

---

<sup>3</sup>  $\Theta(y)$  is the Heaviside function defined by  $\Theta(y) = 1$  for  $y > 0$ ,  $\Theta(0) = 1/2$ , and  $\Theta(y) = 0$  for  $y < 0$ .

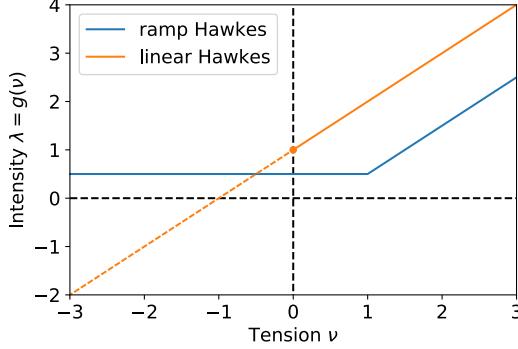


FIG. 5: Schematic of the tension-intensity maps for the linear (i.e.,  $g(\nu) = \nu_0 + \nu$  with  $\nu_0 = 1$ ) and ramp (i.e.,  $g(\nu) = \max\{\nu_0, \nu - \nu_1\}$  with  $\nu_0 = 1/2$  and  $\nu_1 = 1/2$ ) Hawkes processes. While  $\nu_0$  must be non-negative due to the non-negativity of the probability,  $\nu_1$  can be either positive or non-positive.

with the differential operator  $\partial_\nu := \partial/\partial\nu$ . We note that similar calculation technique can be found in Ref. [50]. This ME can be rewritten as

$$\frac{\partial P_t(\nu)}{\partial t} = -\frac{\partial}{\partial\nu} J_t(\nu), \quad J_t(\nu) := -\frac{1}{\tau} \nu P_t(\nu) + \frac{1/c}{1 + \partial_\nu/c} g(\nu) P_t(\nu). \quad (60)$$

Here we assume the natural boundary condition [42]:

$$\lim_{\nu \rightarrow \infty} J_t(\nu) = 0, \quad (61)$$

ensuring that the mean probability ‘‘velocity’’  $J_t(\nu)/P_t(\nu)$  is zero at infinity. We reject the possibility of periodic boundary conditions which are non-physical. In the steady state, we thus obtain the exact steady solution (55).

### B. Example 1: linear Hawkes process.

For the linear intensity function

$$g(\nu) = \nu + \nu_0 \quad (62)$$

with base intensity  $\nu_0 > 0$ , the model recovers the conventional LHawkes process. In the subcritical case  $\eta < 1$ , the exact steady-state solution is given by the gamma distribution

$$P_{ss}(\nu) = \frac{1}{Z} \nu^{-1-a} e^{-\nu/\nu_{cut}}, \quad Z = \nu_{cut}^{\tau\nu_0} \Gamma(\tau\nu_0), \quad a := -\tau\nu_0 \quad (63)$$

with the gamma function  $\Gamma(x) := \int_0^\infty dt t^{x-1} e^{-t}$ . The characteristic tension for the exponential cutoff is defined as

$$\nu_{cut} := \frac{\eta}{(1-\eta)} \frac{1}{\tau}. \quad (64)$$

The PDF  $P_{ss}(\nu)$  and thus the PDF of  $\lambda$  is a power law with a non-universal negative exponent  $a$  up to the cutoff tension  $\nu_{cut}$ . Since the cutoff tension diverges near criticality, the power law ‘‘tail’’ described by  $\nu^{-1+\tau\nu_0}$  and  $\lambda^{-1+\tau\nu_0}$  corresponds to an intermediate asymptotics [18], as reported in Ref. [16].

### C. Example 2: ramp tension-intensity map.

Let us consider the ramp tension-intensity map (also called a rectified linear unit (ReLU) in the context of recent works in machine learning),

$$g(\nu) = \max\{\nu_0, \nu - \nu_1\} \quad (65)$$

for positive  $\nu_0$  and any real number  $\nu_1$  (see Fig. 5). In this paper, the NLHawkes process with the ramp tension-intensity map (65) is called the ramp Hawkes process. While the ramp Hawkes process is quite similar to the LHawkes process, its minimal nonlinearity leads to a genuine asymptotic power law tail, thus very different from the LHawkes process. In the subcritical regime  $\eta < 1$ , the exact steady solution is given by

$$P_{\text{ss}}(\nu) = \begin{cases} \frac{1}{Z} \nu^{-1-\tau\nu_1} e^{-\nu/\nu_{\text{cut}}} & (\nu > \nu_0 + \nu_1) \\ \frac{(\nu_0 + \nu_1)^{-\tau(\nu_0+\nu_1)}}{Z} \nu^{-1+\tau\nu_0} e^{\tau(\nu_0+\nu_1)-c\nu} & (\nu \leq \nu_0 + \nu_1) \end{cases}. \quad (66)$$

with exponential cutoff  $\nu_{\text{cut}}$  given by expression (64), constant  $c$  given by (56) and normalisation constant  $Z$  given by (57).

Interestingly, for  $\nu_1 > 0$  and at criticality  $\eta = 1$ , for  $\nu > \nu_0 + \nu_1$ ,  $P_{\text{ss}}(\nu)$  becomes a pure power law

$$P_{\text{ss}}(\nu) \propto \nu^{-1-a}, \quad a := \tau\nu_1, \quad (67)$$

which is normalisable without truncation. Given the asymptotic linear relationship between  $\nu$  and  $\lambda$ , the same power law behaviour holds for the PDF of  $\lambda$ . This power law is different from the intermediate asymptotic power law distribution (63) for the LHawkes process. In this sense, the ramp Hawkes process can reproduce any power law relationship (including both true and intermediate asymptotics) at criticality, which may be useful to account for power law distributions observed empirically in various systems. It is remarkable that such a slight change from the affine structure (62) to the rectified linear (65) structure creates this large difference in the asymptotic intensity distribution. Note also that, since  $\max\{\nu_0, \nu - \nu_1\} < \nu + \nu_0, \forall \nu_1 > 0$ , the ramp tension-intensity map has a smaller intensity than that of the LHawkes process, which explains the thinner tail (67) compared with (63) (this later becoming so heavy tailed close to criticality so as to become non-normalisable). Intuitively, the base tension  $\nu_0$  in the ramp tension-intensity map (65) acts as a replenishing engine that ensures a minimum activity, which can become the source of bursts. This structure of the ramp tension-intensity map is somewhat reminiscent of the Kesten process [51–53], which is well-known to produce power law distributions with tail exponent depending on the distribution of the multiplicative factors. It is interesting that the exponent  $a = \tau\nu_1$  is independent of the “resourcing” term.

#### D. Existence of steady-state solutions

The exact solution (55) is useful in understanding the condition for the existence of a steady-state solution. For example, let us consider the case of the exponential tension-intensity map:

$$g(\nu) = \lambda_0 e^{\beta\nu}, \quad \beta > 0, \quad (68)$$

which has been used in the statistical calibration of neural spike time series in neural science [49]. The exact solution (55) predicts that this NLHawkes process has no steady-state solution. Indeed,

$$P_{\text{ss}}(\nu) \propto \nu^{-1} \exp \left[ -cv + \lambda_0 \tau \int \nu^{-1} e^{\beta\nu} d\nu \right] \propto \nu^{-1} \exp [-cv + \lambda_0 \tau \text{Ei}(\beta\nu)] \simeq \nu^{-1} \exp \left[ -cv + \lambda_0 \tau \frac{e^{\beta\nu}}{\beta\nu} \right] \quad (69)$$

for large  $\nu$  with the exponential integral  $\text{Ei}(x) := -\int_{-x}^{\infty} t^{-1} e^{-t} dt$ . This PDF is not normalisable, implying that this NLHawkes process is always unstable independently of the model parameters.

To avoid this problem, one of the easiest solutions is to introduce an upper bound in the intensity function:

$$g(\nu) = \min\{\lambda_0 e^{\beta\nu}, \lambda_{\text{max}}\} \quad (70)$$

with the finite upper boundary parameter  $\lambda_{\text{max}} > 0$ . Ref. [49] introduces a similar regularisation to guarantee the stability of their model. However, it is remarkable that this NLHawkes process is always unstable in the absence of the upper bound, and thus simulation results sensitively depend on the specific value of the cutoff  $\lambda_{\text{max}}$ .

In general, if the tension-intensity map diverges faster than the linear (or ramp) function, there is no stationary solution. Indeed, for  $g(\nu) \simeq \lambda_0 \nu^n$  with  $n > 1$ , we obtain

$$P_{\text{ss}}(\nu) \simeq \frac{\nu^{-1}}{Z} \exp \left[ -cv + \lambda_0 \tau \int \nu^{n-1} d\nu \right] \propto \nu^{-1} \exp \left[ -cv + \frac{\lambda_0 \tau}{n} \nu^n \right], \quad (71)$$

which is not normalisable. In this sense, the ramp Hawkes process is the boundary between the stationary and non-stationary Hawkes processes under the assumption of an exponential memory (29) and one-sided exponential marks (54).

Thus, an NLHawkes process with one-sided positive marks is not so flexible, if we require its stationarity. However, this situation drastically changes if we allow for the coexistence of excitatory and inhibitory effects (i.e., marks can take both positive and negative values). Indeed, as will be shown in Sec. V, NLHawkes processes with two-sided marks are flexible enough to accommodate various nonlinearities without losing their stationarity.

### E. Robust asymptotic results

The previous presentation of exact solutions for the ramp tension-intensity map (65) for the special case of (a) an exponential memory and (b) an exponential jump-size distribution, allowed us highlighting the appearance of power law tails for the distribution of tensions near and at criticality. Here, we show that such a power law behaviour is asymptotically robust for general jump-size distributions, assuming that the memory is exponential. With the following notations

$$h(t) = \frac{\eta}{\tau} e^{-t/\tau}, \quad g(\nu) \simeq \nu - \nu_1 + o(\nu^0) \quad \text{for large } \nu, \quad \alpha_k := \int_0^\infty y^k \rho(y) dy < \infty \quad \text{for any } k \geq 1, \quad \alpha_1 = 1,$$

the steady-state intensity distribution  $P_{ss}(\nu)$  is given by the following non-universal power law relation:

$$P_{ss}(\nu) \propto \nu^{-1-a}, \quad a := \frac{2\tau\nu_1}{\alpha_2}. \quad (72)$$

Given the asymptotic linear relationship between  $\nu$  and  $\lambda$ , the same power law behaviour holds for the PDF of  $\lambda$ .

We stress that  $\nu_1$  can take any real value, either positive, negative, or zero. If negative or zero, the derivation does not extend all the way to the limit  $\eta = 1$ , and the power law (72) is truncated as in (63) by an exponential cut-off. This result implies a true power law tail for positive  $\nu_1$  (i.e., normalisable without cutoff tail even at criticality) or intermediate asymptotic power law tail for non-positive  $\nu_1$  (i.e., not normalisable without cutoff tail near criticality). Notably, this recovers Eq. (67) for the one-sided exponential mark distribution (54) for which  $\alpha_2 = 2$ .

#### *Derivation*

Since we are interested only in the tail of the intensity PDF, let us focus on the asymptotic properties of the ME (31) for large  $\nu$ . The ME (31) has the asymptotic expression

$$\frac{1}{\tau} \frac{\partial}{\partial \nu} [\nu P_{ss}(\nu)] + \int dy \rho(y) (\nu - \nu_1 - \eta y / \tau) P_{ss}(\nu - \eta y / \tau) - (\nu - \nu_1) P_{ss}(\nu) \simeq 0 \quad \text{for large } \nu \quad (73)$$

in its steady state, obtained by replacing  $g(\nu)$  by  $\nu - \nu_1$  asymptotically. Applying the Laplace transform

$$\tilde{P}_{ss}(s) := \mathcal{L}_1 [P_{ss}(\nu); s] = \int_0^\infty d\nu e^{-s\nu} P_{ss}(\nu), \quad (74)$$

to this above equation yields

$$-\frac{s}{\tau} \frac{d}{ds} \tilde{P}_{ss}(s) - \Phi(s) \left[ \frac{d}{ds} \tilde{P}_{ss}(s) + \nu_1 \tilde{P}_{ss}(s) \right] \simeq 0, \quad \Phi(s) := \int_0^\infty dy (e^{-\eta y / \tau} - 1) \rho(y). \quad (75)$$

Its solution is given by

$$\log \tilde{P}_{ss}(s) \simeq -\nu_1 s + \int_0^s \frac{\nu_1 s' ds'}{s' + \tau \Phi(s')} \quad (76)$$

with the normalisation condition  $\tilde{P}_{ss}(s=0) = 1$ . Considering the expansion

$$\Phi(s) = -\frac{\eta}{\tau} s + \frac{\eta^2 \alpha_2}{2\tau^2} s^2 + \dots, \quad (77)$$

$\log \tilde{P}_{ss}(s)$  has the following asymptotic form for small  $s$  near criticality  $1 - \eta \ll 1$ ,

$$\log \tilde{P}_{ss}(s) \simeq \frac{2\tau\nu_1}{\alpha_2} \log s \quad \text{for small } s, \quad (78)$$

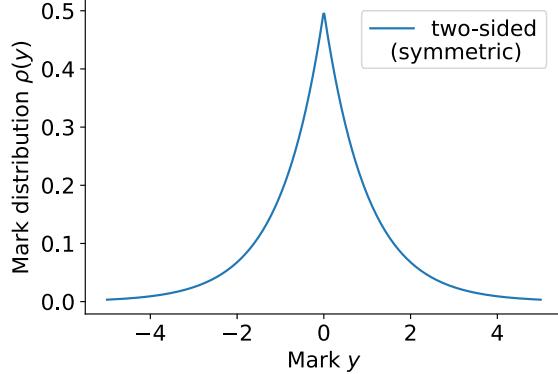


FIG. 6: Schematic of the two-sided symmetric exponential mark distribution (80) with  $y^* = 1$ .

implying, by inverse Laplace transform, a power law asymptotics for the steady intensity PDF:

$$P_{ss}(\nu) \propto \nu^{-1-2\tau\nu_1/\alpha_2}. \quad (79)$$

For non-positive  $\nu_1$ , this PDF is not normalisable and thus requires a cutoff tail, such as the exponential given by Eq. (63).

## V. SOLUTION 2: EXPONENTIAL MEMORY KERNEL WITH TWO-SIDED SYMMETRIC MARK DISTRIBUTION FOR LINEAR TO FAST-ACCELERATING INTENSITY MAPS

In this section, we study both exact and asymptotic solutions of the ME (31) for the PDF of the total tension  $\nu$  valid for an exponential memory kernel (29) and in the presence of inhibitory effects (i.e., marks  $\hat{y}$  can be both positive and negative). The inhibitory effects imply that events can sometimes suppress or decrease the amplitude of bursts, which can lead to essentially different phenomena from those in the previous section.

### A. Exact solutions to two-sided symmetric exponential mark distribution

Let us focus on the case with the two-sided symmetric exponential mark distribution (see Fig. 6):

$$\rho(y) = \frac{1}{2y^*} e^{-|y|/y^*}, \quad (80)$$

which corresponds to the existence of symmetric positive ( $y > 0$ ) and negative ( $y < 0$ ) feedback effects with zero mean. We again assume  $y^* = 1$ , without loss of generality. This negative feedback effect is called the inhibitory effect in Ref. [20] and is known to be difficult to deal with in analytical approaches. We present the exact solution of equation (31) with (80) for some specific forms of  $\lambda = g(\nu)$ .

Let us recall the identity

$$\left(1 - \frac{1}{c} \frac{\partial}{\partial \nu}\right) \int_{-\infty}^0 dy e^{-|y|} g(\nu - \eta y / \tau) P_t(\nu - \eta y / \tau) = g(\nu) P_t(\nu), \quad (81)$$

where  $c = \tau/\eta$  has been defined in (56) (see Appendix C for the derivation). This identity together with the other identity (58) implies a third identity useful to solve the ME (31):

$$\int_{-\infty}^{+\infty} dy \frac{e^{-|y|}}{2} g(\nu - \eta y / \tau) P_t(\nu - \eta y / \tau) = \frac{1}{2} \left[ \frac{1}{1 + \partial_\nu/c} + \frac{1}{1 - \partial_\nu/c} \right] g(\nu) P_t(\nu) = \frac{1}{1 - \partial_\nu^2/c^2} g(\nu) P_t(\nu). \quad (82)$$

We thus obtain a simple representation of the ME (31):

$$\frac{\partial P_t(\nu)}{\partial t} = \frac{1}{\tau} \partial_\nu [\nu P_t(\nu)] + \frac{\partial_\nu^2/c^2}{1 - \partial_\nu^2/c^2} g(\nu) P_t(\nu). \quad (83)$$

This ME can be written in the more familiar form

$$\frac{\partial P_t(\nu)}{\partial t} = -\frac{\partial J_t(\nu)}{\partial \nu}, \quad (84)$$

where the probability current is defined by

$$J_t(\nu) := -\left[\frac{\nu}{\tau} + \frac{\partial_\nu/c^2}{1-\partial_\nu^2/c^2}g(\nu)\right]P_t(\nu). \quad (85)$$

This formulation makes more transparent the meaning of the boundary condition  $\lim_{\nu \rightarrow \infty} J_t(\nu) = 0$  ensuring that the mean probability ‘‘velocity’’  $J_t(\nu)/P_t(\nu)$  is zero at infinity. We reject the possibility of periodic boundary conditions which are non-physical.

Then, the steady-state solution satisfies the following second-order differential equation

$$\frac{d^2}{d\nu^2} [\nu P_{ss}(\nu)] - \tau \frac{d}{d\nu} [g(\nu)P_{ss}(\nu)] - c^2 \nu P_{ss}(\nu) = 0, \quad (86)$$

This is obtained by putting  $\frac{\partial P_t(\nu)}{\partial t} = 0$  in (84), and using  $\lim_{\nu \rightarrow \infty} J_t(\nu) = 0$  that leads to  $J_t(\nu) = 0, \forall \nu$ , from which equation (86) derives.

#### *Example 1: ramp tension-intensity map*

For the ramp tension-intensity map

$$\lambda = g(\nu) = \max\{\nu_0, |\nu|\}, \quad (87)$$

which corresponds to setting  $\nu_1 = 0$  in Eq. (65) and adding the absolute value, the solution of (86) is a truncated-Lévy-type intensity asymptotic tail:

$$P_{ss}(\lambda) \propto \lambda^{-1} e^{-\frac{\lambda}{\lambda_{cut}}}, \quad \lambda_{cut} := \frac{2c^2}{\tau + \sqrt{4c^2 + \tau^2}}, \quad \text{for large } \lambda, \quad (88)$$

where  $c = \tau/\eta$  has been defined in (56). For  $\nu_1 \neq 0$  in Eq. (65), the exact form of the intensity distribution is also available.

Remarkably, this model has no critical point: the process is always stationary for all  $\eta < \infty$ , due to the stabilisation effect of the inhibitory component of the process, and thus  $\lambda_{cut}$  is always finite. This is in contrast to the ramp Hawkes process with one-sided exponential jumps (without inhibitory effect), and thus highlights the fact that the inhibitory effects can be crucial in understanding even the qualitative behaviour of the NLHawkes processes in general.

*Derivation of the exact solution.* The ME (86) reduces to the following set of modified Bessel differential and constant-coefficient second-order differential equations:

$$\nu^2 \frac{d^2 \phi(\nu)}{d\nu^2} + \nu \frac{d\phi(\nu)}{d\nu} - (c^2 \nu^2 + \gamma^2) \phi(\nu) = 0, \quad \gamma := \frac{1 - \tau \nu_0}{2}, \quad \phi(\nu) := \nu^\gamma P_{ss}(\nu) \quad (0 \leq \nu \leq \nu_0), \quad (89)$$

$$\frac{d^2}{d\nu^2} \psi(\nu) - \tau \frac{d}{d\nu} \psi(\nu) - c^2 \psi(\nu) = 0, \quad \psi(\nu) := \nu P_{ss}(\nu) \quad (\nu_0 \leq \nu). \quad (90)$$

The exact solution is then given by

$$P_{ss}(\nu) = \begin{cases} \nu^{-\gamma} \left( C^{[1]} I_\gamma(|\nu|/c) + C^{[2]} K_\gamma(|\nu|/c) \right) & (|\nu| \leq \nu_0) \\ C^{[3]} |\nu|^{-1} e^{-|\nu|/\nu_{cut}} & (|\nu| > \nu_0) \end{cases}, \quad \nu_{cut} := \frac{2c^2}{\tau + \sqrt{4c^2 + \tau^2}}, \quad c := \frac{\tau}{\eta} \quad (91)$$

with integral constants  $C^{[1]}, C^{[2]}, C^{[3]}$  and modified Bessel functions of the first and second kinds (denoted by  $I_\gamma(x)$  and  $K_\gamma(x)$ , respectively; see Appendix D 1). The integral constants are determined by the normalisation and continuity conditions:  $\int_{-\infty}^{\infty} d\nu P_{ss}(\nu) = 1$ ,  $\lim_{\nu \uparrow \nu_0} P_{ss}(\nu) = \lim_{\nu \downarrow \nu_0} P_{ss}(\nu)$ . We thus obtain that the intensity distribution is given by the sum of a  $\delta$  function centred on  $\nu_0$  and the truncated Lévy distribution,

$$P_{ss}(\lambda) = \left( 1 - C^{[3]} \Gamma(0, \nu_0/\nu_{cut}) \right) \delta(\nu - \nu_0) + C^{[3]} \lambda^{-1} e^{-\lambda/\nu^*} \Theta(\nu - \nu_0). \quad (92)$$

where the incomplete gamma function is  $\Gamma(a, x) := \int_x^\infty dt t^{a-1} e^{-t}$ .

*Example 2: quadratic tension-intensity map*

For the quadratic tension-intensity map corresponding to the ZHawkes process, see Sec. II C 2,

$$\lambda = g(\nu) = k\nu^2 + \lambda_0 , \quad (93)$$

the solution of (86) is a power law steady-state distribution<sup>4</sup>

$$P_{\text{ss}}(\lambda) \propto |\lambda|^{-1-a} , \quad a := \frac{1}{2} + \frac{c^2}{2k\tau} , \quad c := \frac{\tau}{\eta} \quad (94)$$

with  $\lambda_0 > 0$  and power law exponent  $a > 1/2$ . The exact form of the intensity distribution is also available. We note that this non-universal power law scaling is consistent with Eq. (12), which was reported for the diffusive limit of the ZHawkes process in Ref. [28].

*Derivation of the exact solution.* By the variable transformation  $x = k\tau\nu^2/2$ , the ME (86) for  $\nu > 0$  reduces to

$$x \frac{d^2\phi(x)}{dx^2} + \left( \frac{3}{2} - \frac{\tau\lambda_0}{2} - x \right) \frac{d\phi(x)}{dx} - \left( 1 + \frac{c^2}{2k\tau} \right) \phi(x) = 0, \quad \phi(x) := P(\nu(x)), \quad \nu(x) := \sqrt{\frac{2x}{k\tau}}. \quad (95)$$

This is the confluent hypergeometric differential equation and thus its exact solution is given by

$$\phi(x) = C^{[1]} {}_1F_1 \left( 1 + \frac{c^2}{2k\tau}, \frac{3}{2} - \frac{\tau\lambda_0}{2}; x \right) + C^{[2]} {}_1U_1 \left( 1 + \frac{c^2}{2k\tau}, \frac{3}{2} - \frac{\tau\lambda_0}{2}; x \right) \quad (96)$$

with integral constants  $C^{[1]}, C^{[2]}$  and the confluent hypergeometric functions of the first  ${}_1F_1$  and second kind  ${}_1U_1$  (see Appendix D 2). The integral constants are determined by the normalisation condition  $\int_{-\infty}^{\infty} d\nu P_{\text{ss}}(\nu) = 1$ . Interestingly, this solution has the following asymptotic form for large  $x$

$$\phi(x) \propto x^{-1-c^2/(2k\tau)}. \quad (97)$$

The steady distribution of the intensity  $\lambda$  is then given by

$$P_{\text{ss}}(\lambda) = \left| \frac{d\nu}{d\lambda} \right| P_{\text{ss}}(\nu) \propto \lambda^{-1-a} , \quad a := \frac{1}{2} + \frac{c^2}{2k\tau} , \quad c := \frac{\tau}{\eta} \quad (98)$$

for the tail  $\lambda \rightarrow \infty$ . This is a power law asymptotic distribution with a non-universal exponent  $a$  without truncation.

*Example 3: exponential intensity map*

For the exponential tension-intensity map

$$\lambda = g(\nu) = \lambda_0 |\nu| e^{\beta\nu} , \quad (99)$$

the solution of (86) is Zipf's law for the intensity distribution:

$$P_{\text{ss}}(\lambda) \propto \lambda^{-2} , \quad (100)$$

up to a logarithmic factor  $\log \lambda$ , for large  $\lambda$  with positive constants  $\lambda_0$  and  $\beta$ . This intensity map is inspired by the MSA model [21, 22], where the dominant contribution comes from the exponential factor originating from the Arrhenius law (see Sec. V B for more detail).

<sup>4</sup> If  $\lambda_0$  is zero, the steady-state distribution is singular at  $\nu = 0$  as  $P_{\text{ss}}(\lambda) \propto \nu^{-1}$  and thus is not normalisable.

*Derivation of the exact solution.* The exact steady-state solution of (86) is given by

$$P_{ss}(\nu) = \begin{cases} \frac{C^{[1]}}{\nu} e^{c\nu} {}_1U_1\left(1 + \frac{c}{\beta}, 1 + \frac{2c}{\beta}; \frac{\lambda_0\tau}{\beta} e^{\beta\nu}\right) + \frac{C^{[2]}}{\nu} e^{c\nu} L_{-1-c/\beta}^{2c/\beta}\left(\frac{\lambda_0\tau}{\beta} e^{\beta\nu}\right) & (\nu \geq 0) \\ \frac{C^{[3]}}{\nu} \exp\left[c\nu - \frac{\lambda_0\tau}{\beta} e^{\beta\nu}\right] {}_1U_1\left(\frac{c}{\beta}, 1 + \frac{2c}{\beta}; \frac{\lambda_0\tau}{\beta} e^{\beta\nu}\right) + \frac{C^{[4]}}{\nu} \exp\left[c\nu - \frac{\lambda_0\tau}{\beta} e^{\beta\nu}\right] L_{-c/\beta}^{2c/\beta}\left(\frac{\lambda_0\tau}{\beta} e^{\beta\nu}\right) & (\nu < 0) \end{cases} \quad (101)$$

with integral coefficients  $C^{[1]}$ ,  $C^{[2]}$ ,  $C^{[3]}$ , and  $C^{[4]}$ , and the generalised Laguerre function  $L_a^b(x)$  (see Appendix D 3). Considering the asymptotic formulas (D3) and (D5),  $C^{[2]}$  must be zero since  $P_{ss}(\nu) \rightarrow 0$  for  $\nu \rightarrow \infty$ . This means that the asymptotic tail is given by

$$P_{ss}(\nu) \propto \nu^{-1} e^{-\beta\nu}, \quad (102)$$

which leads to the Zipf law (100) for the steady intensity distribution, by using the Jacobian relation  $d\nu P_{ss}(\nu) = d\lambda P_{ss}(\lambda) \iff P_{ss}(\lambda) = |d\nu/d\lambda| P_{ss}(\nu)$ . As shown in Secs. V B and V C 3, this asymptotic Zipf law (100) is robust for exponential-type tension-intensity maps under general symmetric mark distribution (or more generally when the mark average is zero), on the condition that the memory kernel is exponential and the mark distribution has its moments at all orders being finite.

## B. Exact solutions in the diffusive limit

Let us now consider the diffusive limit formulated in Sec. III E, and assume an exponential memory kernel (29):

$$h(t) = \frac{\eta}{\tau} e^{-t/\tau}, \quad \rho_\varepsilon(y) = \frac{1}{\varepsilon} \tilde{\rho}\left(\frac{y}{\varepsilon}\right), \quad g(\nu) = \frac{1}{\varepsilon^2} \tilde{g}(\nu). \quad (103)$$

For this case, by solving the FPE (46) in the steady state, we obtain the explicit solution

$$P_{ss}(\nu) \propto_{\varepsilon \rightarrow 0} \frac{1}{\tilde{g}(\nu)} \exp\left[-\frac{1}{\tau D} \int \frac{\nu d\nu}{\tilde{g}(\nu)}\right], \quad D := \frac{\tilde{\alpha}_2 \eta^2}{2\tau^2} \quad (104)$$

for any  $\tilde{g}(\nu)$ , assuming that  $\tilde{\rho}(y)$  and  $\tilde{g}(\nu)$  are independent of  $\varepsilon$  and that all the integrals appropriately converge.

### Example 1: ramp tension-intensity map

Let us first consider the example of the ramp tension-intensity map

$$g(\nu) = \max\{\nu_0, |\nu - \nu_1|\} \quad (105)$$

with positive real number  $\nu_0 > 0$  and arbitrary real number<sup>5</sup>  $\nu_1$ . For  $\nu > \nu_0 + \nu_1$ , the solution of the FPE (46) is the truncated Lévy distribution for the tension (and thus for the intensity)

$$P_{ss}(\nu) = \frac{1}{Z} \frac{e^{-\nu/\nu_{cut}}}{(\nu - \nu_1)^{1+a}}, \quad \nu_{cut} := D\tau, \quad a := \frac{\nu_1}{D\tau}, \quad (106)$$

with an exponential tail tapering the intermediate power law tail. We note that this model has no critical point due to the inhibitory effects leading to the characteristic intensity for the exponential cutoff  $\nu_{cut}$  to be always finite.

### Example 2: quadratic tension-intensity map

We next study the exact solution of the quadratic intensity (93), corresponding to the ZHawkes process, see Sec. II C 2, in the presence of inhibitory effects. Using formula (104), the exact solution of the FPE (46) in the steady-state regime is given by

$$P_{ss}(\nu) = \frac{1}{Z} (k\nu^2 + \lambda_0)^{-1-1/(2kD\tau)}, \quad (107)$$

---

<sup>5</sup> Here  $\nu_1$  can be either positive, zero, or negative.

which is equivalent to

$$P_{ss}(\lambda) = \frac{1}{Z'} \frac{\lambda^{-1-1/(2kD\tau)}}{\sqrt{\lambda - \lambda_0}} \propto \lambda^{-1-a} \quad (\lambda \rightarrow \infty), \quad a := \frac{1}{2} + \frac{1}{2kD\tau} \quad (108)$$

with  $Z' := 2\sqrt{k}Z$ . This is a power law distribution without truncation and with a non-universal exponent  $a$ . We note that this non-universal power law scaling is essentially identical to Eq. (12) for the diffusive limit of the ZHawkes process reported in Ref. [28].

### Example 3: polynomial tension-intensity map

Let us consider the case of the polynomial intensity given by

$$\tilde{g}(\nu) = k|\nu|^n + \nu_0 \quad (n > 2) \quad (109)$$

with positive constant  $\nu_0 > 0$ . Using formula (104), we obtain the exact steady-state distribution, solution of the FPE (46),

$$P_{ss}(\nu) \propto \frac{1}{k|\nu|^n + \nu_0} \exp \left[ -\frac{\nu^2}{2\tau D\nu_0} {}_2F_1 \left( 1, \frac{2}{n}, 1 + \frac{2}{n}; -\frac{k|\nu|^n}{\nu_0} \right) \right] \quad (110)$$

with the hypergeometric function  ${}_2F_1$  (see Appendix D 4). By considering the following asymptotic expansion

$$\nu^2 {}_2F_1 \left( 1, \frac{2}{n}, 1 + \frac{2}{n}; -\frac{k\nu^n}{\nu_0} \right) = \begin{cases} \underbrace{\frac{\frac{2\pi}{n}}{\sin \frac{2\pi}{n}} \left( \frac{\nu_0}{k} \right)^{\frac{2}{n}}}_{\text{const.}} + o(\nu^0) & (n > 2) \\ \frac{\nu_0}{k} \log \left( \frac{k\nu^2}{\nu_0} \right) + o(\nu^0) & (n = 2) \end{cases} \quad (111)$$

for large  $\nu$ 's, we obtain the asymptotic form of the steady PDF for  $|\nu| \rightarrow \infty \implies \lambda \rightarrow \infty$  as

$$P_{ss}(\nu) \propto \frac{1}{k|\nu|^n + \nu_0} \implies P_{ss}(\lambda) \propto \lambda^{-1-a}, \quad a := 1 - \frac{1}{n}. \quad (112)$$

Note that the limit  $n \rightarrow +\infty$  recovers Zipf's law. The manner with which the exact solution (110) recovers Eq. (108) for  $n \rightarrow 2$  is now elaborated.

*Crossover between  $n = 2$  and  $n > 2$ .* Remarkably, the solution (108) for the QHawkes (i.e.,  $n = 2$ ) and the one (110) for the polynomial Hawkes with  $n > 2$  are slightly different. This qualitative difference can be seen from the analytical singularity of the hypergeometric function  ${}_2F_1$  at  $n = 2$  and suggests a crossover between two power law regimes. Here we explicitly estimate the crossover point.

Let us introduce a small positive parameter  $\epsilon$  as

$$n := \frac{2}{1-\epsilon} > 2 \quad (113)$$

and consider the limit  $\epsilon \downarrow 0$ . We focus on the discontinuous switching in Eq. (111) between  $n > 2$  and  $n = 2$ . To estimate the crossover point, it is necessary to evaluate their higher-order asymptotic behaviour for large  $\nu$  with nonzero  $\epsilon$  as given by Eq. (D10). As summarised in Appendix D 4 b, the threshold intensity is estimated to be

$$\lambda^* := \nu_0 e^{2/\epsilon}, \quad (114)$$

which characterises the crossover between the two regimes. We thus obtain the explicit crossover formula as

$$P_{ss}(\lambda) \propto \begin{cases} \lambda^{-1-a_1} & (\lambda \ll \lambda^*, a_1 := \frac{1}{2} + \frac{1}{2kD\tau}) \\ \lambda^{-1-a_2} & (\lambda \gg \lambda^*, a_2 := \frac{1}{2}) \end{cases}. \quad (115)$$

The existence of this crossover point can be intuitively understood as follows: let us go back to the SDE representation (47). Remarkably, the cases  $n = 2$  and  $n > 2$  are critically different in the sense that the relaxation term  $-\nu/\tau$  is the same order as the fluctuation  $\sqrt{2D\tilde{g}(\nu)}\xi^G$  for  $n = 2$ , whereas it is negligible for  $n > 2$

$$\left| -\frac{\nu}{\tau} \right| \ll \sqrt{2D\tilde{g}(\nu)}\xi^G \quad (116)$$

for a sufficiently large  $\nu \gg \nu^*$ . Such a crossover point can be roughly estimated by the relationship  $\nu^*/\tau = \sqrt{2D\tilde{g}(\nu^*)}$ , suggesting  $\nu^* = C^{1/\epsilon}$  with some constant  $C$ . We therefore obtain  $\log \lambda^* \propto \epsilon^{-1}$  consistently with Eq. (114).

*Example 4: Multifractal stress activation model*

An interesting example is the MSA model for earthquake triggering proposed in Refs. [21, 22] and summarised for our purpose in section II C 1, which corresponds to

$$\tilde{g}(\nu) = \lambda_0 \exp(\beta\nu), \quad (117)$$

with base intensity  $\lambda_0 > 0$  and effective inverse temperature  $\beta > 0$ . From the steady-state solution (104) of the FPE (46), we obtain the steady solution  $P_{ss}(\nu)$  for the tension  $\nu$  and  $P_{ss}(\lambda)$  of the intensity  $\lambda = \lambda_0 e^{\beta\nu}$ :

$$P_{ss}(\nu) = \frac{1}{Z} \frac{e^{-\beta\nu}}{\lambda_0} \exp\left[\frac{e^{-\beta\nu}(1 + \beta\nu)}{\lambda_0 \beta^2 \tau D}\right] \iff P_{ss}(\lambda) = \frac{1}{\beta Z} \lambda^{-2} \exp\left[\frac{\lambda^{-1}\{1 + \log(\lambda/\lambda_0)\}}{\beta^2 \tau D}\right] \quad (118)$$

The derivation of  $P_{ss}(\lambda)$  from  $P_{ss}(\nu)$  uses the Jacobian relation  $d\nu P_{ss}(\nu) = d\lambda P_{ss}(\lambda) \iff P_{ss}(\nu) = \beta\lambda P_{ss}(\lambda)$ . This steady intensity distribution exhibits Zipf's law similarly to the aforementioned result (100):

$$P_{ss}(\lambda) \propto \lambda^{-2} \quad (\text{for large } \lambda). \quad (119)$$

*Example 5: fast-accelerating intensity*

Let us focus on a large class of intensity map  $g(\nu)$  satisfying

$$\tilde{g}(\nu) \gg \nu^2 \quad (\text{for large } \nu), \quad (120)$$

which we refer to as a fast-accelerating intensity (FAI) map. For example, the polynomial intensity (109) beyond second order and the MSA intensity (117) belong to this class. FAI maps are special in the sense that the asymptotic PDF of  $\nu$ , which is solution of the FPE (46), is given by

$$P_{ss}(\nu) \propto \frac{1}{\tilde{g}(\nu)} \exp\left[-\frac{1}{\tau D} \int \frac{\nu d\nu}{\tilde{g}(\nu)}\right] = \exp[-\log \tilde{g}(\nu) - o(\nu^{-c})] \quad (121)$$

with some positive constant  $c > 0$ . This expression is derived from Eq. (104), considering that  $\nu/\tilde{g}(\nu) = o(\nu^{-1})$ . We thus obtain a general asymptotic form

$$P_{ss}(\lambda) \propto \lambda^{-1} \left| \frac{d\tilde{g}(\nu)}{d\nu} \right|_{\nu=\tilde{g}^{-1}(\lambda)}^{-1}. \quad (122)$$

### C. Robust asymptotic solutions

#### 1. Robust exponential tail for the ramp intensity

Here we show that the exponential tail (88) and (92) for the ramp intensity (87) of the steady-state solution (104) of the FPE (46) remains valid for general symmetric mark distributions, assuming appropriate convergence of the moment-generating function and with an exponential memory function:

$$h(t) = \frac{\eta}{\tau} e^{-t/\tau}, \quad \rho(y) = \rho(-y), \quad g(\nu) \simeq \nu + \nu_0 \quad (\text{for large } \nu) \implies P_{ss}(\lambda) \propto e^{-\lambda/\lambda_{\text{cut}}} \quad (\text{for large } \lambda), \quad (123)$$

up to a sub-leading contribution in the form of a truncated power law. The parameter  $\lambda_{\text{cut}}$  is given by the self-consistent relation

$$\frac{1}{\tau \lambda_{\text{cut}}} = \Phi\left(\frac{\eta}{\tau \lambda_{\text{cut}}}\right), \quad \Phi(x) := \int_{-\infty}^{\infty} dy \rho(y) (e^{xy} - 1) = \int_{-\infty}^{\infty} dy \rho(y) (\cosh yx - 1), \quad (124)$$

where  $\Phi(x)$  is the moment-generating function. The equation for  $\lambda_{\text{cut}}$  has a single positive solution (see Appendix F).

*Derivation.* The solution (123) can be derived by direct substitution into the ME (31) as follows. Let us make an ansatz that the solution is given by

$$P_{\text{ss}}(\nu) \propto e^{-\nu/\lambda_{\text{cut}}} \quad \text{for large } \nu. \quad (125)$$

By considering the relations for large  $\nu$

$$\begin{aligned} & \frac{1}{\tau} \frac{d}{d\nu} \left[ \nu e^{-\frac{\nu}{\lambda_{\text{cut}}}} \right] + \int_{-\infty}^{\infty} dy \rho(y) (\nu + \nu_0 - \eta y/\tau) e^{-\frac{\nu}{\lambda_{\text{cut}}} + \frac{\eta y}{\tau \lambda_{\text{cut}}}} - (\nu + \nu_0) e^{-\frac{\nu}{\lambda_{\text{cut}}}} \\ &= \nu e^{-\frac{\nu}{\lambda_{\text{cut}}}} \left[ -\frac{1}{\tau \lambda_{\text{cut}}} + \int_{-\infty}^{\infty} dy \rho(y) e^{\frac{\eta y}{\tau \lambda_{\text{cut}}}} - 1 \right] + o\left(\nu e^{-\frac{\nu}{\lambda_{\text{cut}}}}\right) \\ &= \nu e^{-\frac{\nu}{\lambda_{\text{cut}}}} \left[ -\frac{1}{\tau \lambda_{\text{cut}}} + \Phi\left(\frac{\eta}{\tau \lambda_{\text{cut}}}\right) \right] + o\left(\nu e^{-\frac{\nu}{\lambda_{\text{cut}}}}\right), \end{aligned} \quad (126)$$

the ME (31) in the steady state reads

$$\nu e^{-\frac{\nu}{\lambda_{\text{cut}}}} \left[ -\frac{1}{\tau \lambda_{\text{cut}}} + \Phi\left(\frac{\eta}{\tau \lambda_{\text{cut}}}\right) \right] + o\left(\nu e^{-\frac{\nu}{\lambda_{\text{cut}}}}\right) = 0 \quad (127)$$

for large  $\nu$ . This relation is equivalent to the self-consistent relation (124).

## 2. Robust power law tail for quadratic intensity

We show that the power law tail for the quadratic intensity, such as Eq. (98) and Eq. (108), of the steady-state solution (104) of the FPE (46) is generally valid for general symmetric mark size distributions:

$$h(t) = \frac{\eta}{\tau} e^{-t/\tau}, \quad \rho(y) = \rho(-y), \quad g(\nu) \simeq k\nu^2 + \lambda_0 \quad \text{for large } \nu \implies P_{\text{ss}}(\lambda) \propto \lambda^{-1-a} \quad (128)$$

for large  $\lambda$  with  $a > 1/2$ , assuming appropriate convergence of the KM coefficients.

Note that the authors of Ref. [28] conjectured that the PDF of the intensity of various ZHawkes processes should be a power law with a non-universal exponent. Our results confirm this conjecture, as least for an exponential memory kernel, in the sense that the power law asymptotics with a non-universal exponent  $a > 1/2$  is a robust property of ZHawkes processes, independently of the shape of the mark distribution, as long as it is symmetric with finite moments.

*Derivation.* Let us go back to the ME in the steady state,

$$\frac{1}{\tau} \frac{\partial}{\partial \nu} [\nu P_{\text{ss}}(\nu)] + \int_{-\infty}^{\infty} dy \rho(y) (k(\nu - \eta y/\tau)^2 + \lambda_0) P_{\text{ss}}(\nu - \eta y/\tau) - (k\nu^2 + \lambda_0) P_{\text{ss}}(\nu) = 0. \quad (129)$$

We make the anzatz that the asymptotic solution is given by

$$P_{\text{ss}}(\nu) \simeq C\nu^{-\kappa} + o(\nu^{-\kappa}) \quad (\text{for large } \nu) \quad (130)$$

with a positive  $\kappa$  and a certain constant  $C$ . This implies that

$$\begin{aligned} & C \left[ \frac{1-\kappa}{\tau} \nu^{-\kappa} + \int_{-\infty}^{\infty} dy \rho(y) \left\{ k \left( \nu - \frac{\eta y}{\tau} \right)^2 + \lambda_0 \right\} \left( \nu - \frac{\eta y}{\tau} \right)^{-\kappa} - (k\nu^2 + \lambda_0)\nu^{-\kappa} + o(\nu^{-\kappa}) \right] \\ &= C \left[ \frac{1-\kappa}{\tau} \nu^{-\kappa} + \int_{-\infty}^{\infty} dy \rho(y) \left\{ k\nu^{-\kappa+2} \left( 1 - \frac{\eta y}{\tau \nu} \right)^{-\kappa+2} + \lambda_0 \nu^{-\kappa} \left( 1 - \frac{\eta y}{\tau \nu} \right)^{-\kappa} \right\} - (k\nu^2 + \lambda_0)\nu^{-\kappa} + o(\nu^{-\kappa}) \right] \\ &= C \left[ \frac{1-\kappa}{\tau} \nu^{-\kappa} + \int_{-\infty}^{\infty} dy \rho(y) \left\{ k\nu^{-\kappa+2} \left( 1 + \frac{(\kappa-2)\eta y}{\tau \nu} + \frac{\eta^2 y^2 (\kappa-1)(\kappa-2)}{2\tau^2 \nu^2} \right) + \lambda_0 \nu^{-\kappa} \right\} - (k\nu^2 + \lambda_0)\nu^{-\kappa} + o(\nu^{-\kappa}) \right] \\ &= C \left[ \frac{1-\kappa}{\tau} \nu^{-\kappa} + \frac{k\alpha_2 \eta^2}{2\tau^2} (\kappa-1)(\kappa-2)\nu^{-\kappa} + o(\nu^{-\kappa}) \right] \simeq 0, \end{aligned} \quad (131)$$

which leads to the self-consistent relation

$$\kappa = 2 + \frac{1}{kD\tau}, \quad D := \frac{\alpha_2 \eta^2}{2\tau^2}. \quad (132)$$

We thus obtain the power law tail of the intensity distribution:

$$P_{\text{ss}}(\lambda) \propto \lambda^{-1-a}, \quad a := \frac{1}{2} + \frac{1}{2kD\tau}. \quad (133)$$

### 3. Robust Zipf's law for the multifractal stress activation model

We have shown that the exact steady-state solution of the FPE (46) exhibits Zipf's law for the MSA model (117) with exponential memory kernel in the diffusive limit. Here, we show that Zipf's law universally and robustly appears for the MSA model with any general symmetric mark distribution, on the condition that the memory is exponential and the appropriate integrals converge. In other words,

$$h(t) = \frac{\eta}{\tau} e^{-t/\tau}, \quad g(\nu) = \lambda_0 e^{\beta\nu}, \quad \rho(y) = \rho(-y) \implies P_{ss}(\lambda) \propto \lambda^{-2} \quad (\text{for large } \lambda). \quad (134)$$

*Derivation.* By defining  $\phi(\nu) := g(\nu)P_{ss}(\nu)$ , the steady-state ME is given by

$$\frac{1}{\tau} \frac{\partial}{\partial \nu} [\nu e^{-\beta\nu} \phi(\nu)] + \int_{-\infty}^{\infty} dy \rho(y) \phi(\nu - \eta y/\tau) - \phi(\nu) = 0. \quad (135)$$

For large  $\nu$ , the first term in the left-hand side is negligible due to the exponential factor  $e^{-\beta\nu}$ , implying

$$\int_{-\infty}^{\infty} dy \rho(y) \phi(\nu - \eta y/\tau) - \phi(\nu) \simeq 0 \quad \text{for large } \nu. \quad (136)$$

Assuming that  $\phi(\nu)$  is nonnegative, this integral equation has a general solution

$$\phi(\nu) = C_0 + C_1 \nu \quad (137)$$

with constants  $C_0$  and  $C_1$  (see Appendix F). By imposing the natural boundary condition,  $C_1$  must be zero as shown later, and, therefore, the general solution is given by  $\phi(\nu) = C_0$ . This implies the following asymptotic form of the steady-state PDF

$$P_{ss}(\nu) = \frac{\phi(\nu)}{g(\nu)} \propto e^{-\beta\nu}. \quad (138)$$

We thus obtain Zipf's law for the intensity PDF, from the Jacobian relation  $P_{ss}(\lambda) = P_{ss}(\nu)|d\nu/d\lambda|$ .

*Natural boundary condition.* Here we impose the natural boundary condition to remove  $C_1$ . Let us use the KM expansion (33) to define the probability current as

$$\frac{\partial P_t(\nu)}{\partial t} = -\frac{\partial}{\partial \nu} J_t(\nu), \quad J_t(\nu) := -\frac{1}{\tau} [\nu P_t(\nu)] - \sum_{k=1}^{\infty} \frac{\alpha_{2k}}{(2k)!} \frac{\eta^{2k}}{\tau^{2k}} \frac{\partial^{2k-1}}{\partial \nu^{2k-1}} g(\nu) P_t(\nu). \quad (139)$$

For the steady-state distribution, let us ignore the first term in  $J_t(\nu)$  for large  $\nu$  to obtain

$$J_{ss}(\nu) \simeq - \sum_{k=1}^{\infty} \frac{\alpha_{2k}}{(2k)!} \frac{\eta^{2k}}{\tau^{2k}} \frac{\partial^{2k-1}}{\partial \nu^{2k-1}} g(\nu) P_{ss}(\nu) \quad \text{for large } \nu. \quad (140)$$

By direct substitution of the general solution  $g(\nu)P_{ss}(\nu) = \phi(\nu) = C_0 + C_1 \nu$ , we obtain

$$J_{ss}(\nu) \simeq - \sum_{k=1}^{\infty} \frac{\alpha_{2k}}{(2k)!} \frac{\eta^{2k}}{\tau^{2k}} \frac{\partial^{2k-1}}{\partial \nu^{2k-1}} (C_0 + C_1 \nu) = -\frac{\eta^2 \alpha_2}{2\tau^2} C_1. \quad (141)$$

Since the natural boundary condition implies  $\lim_{\nu \rightarrow \infty} J_t(\nu) = 0$  for any  $t$ , we obtain  $C_1 = 0$ .

### 4. Robust asymptotic form for fast-accelerating intensity maps

We now show that the asymptotic form (122) of the steady-state solution (104) of the FPE (46) is robust even for general mark distribution for any FAI map:

$$h(t) = \frac{\eta}{\tau} e^{-t/\tau}, \quad g(\nu) \gg \nu^2 \quad (\text{for large } \nu), \quad \rho(y) = \rho(-y) \implies P_{ss}(\lambda) \propto \lambda^{-1} \left| \frac{dg(\nu)}{d\nu} \right|_{\nu=g^{-1}(\lambda)}^{-1} \quad (\text{for large } \lambda). \quad (142)$$

*Derivation.* By defining  $\phi(\nu) := g(\nu)P_{\text{ss}}(\nu)$ , the steady-state ME is given by

$$\frac{1}{\tau} \frac{\partial}{\partial \nu} \left\{ \frac{\nu}{g(\nu)} \phi(\nu) \right\} + \int_{-\infty}^{\infty} dy \rho(y) \phi(\nu - \eta y/\tau) - \phi(\nu) = 0. \quad (143)$$

For large  $\nu$ , the first term in the left-hand side is negligible because  $g(\nu)$  is a FAI. The self-consistency of this assumption will be confirmed later. This implies

$$\int_{-\infty}^{\infty} dy \rho(y) \phi(\nu - \eta y/\tau) - \phi(\nu) \simeq 0 \quad \text{for large } \nu. \quad (144)$$

Assuming the nonnegativity of  $\phi(\nu)$  and the natural boundary condition, this integral equation has a single solution  $\phi(\nu) = C_0$  with a constant  $C_0$  in the same logic to that in Sec. V C 3. Finally this implies the following asymptotic form of the steady-state PDF of  $\nu$ :

$$P_{\text{ss}}(\nu) \propto \frac{1}{g(\nu)}. \quad (145)$$

Formula (142) for the intensity PDF then derives from the Jacobian relation  $P_{\text{ss}}(\lambda) = P_{\text{ss}}(\nu)|d\nu/d\lambda|$ .

*Self-consistency of the assumption.* Let us check whether this solution is consistent with the assumption that the first term in Eq. (143) is irrelevant for large  $\nu$ . For simplicity, we focus on the case of  $g(\nu) = \nu^n$  with integer  $n \geq 2$ . We first assume the expansion of the solution

$$\phi(\nu) = \phi_0(\nu) + \phi_1(\nu) + \dots, \quad \phi_0(\nu) = C^{[0]}, \quad |\phi_0(\nu)| \gg |\phi_1(\nu)| \quad (146)$$

for large  $\nu$  with a constant  $C^{[0]}$ . By assuming the the first term in the left-hand side of Eq. (143) is subleading, we obtain

$$\frac{(1-n)C^{[0]}}{\tau} \nu^{-n} + \int_{-\infty}^{\infty} dy \rho(y) \phi_1(\nu - \eta y/\tau) - \phi_1(\nu) \simeq 0. \quad (147)$$

We solve this non-homogeneous integral equation by assuming a solution anzatz:

$$\phi_1(\nu) \simeq C^{[1]} \nu^{-\kappa}, \quad \kappa > 0. \quad (148)$$

By using  $\int_{-\infty}^{\infty} y \rho(y) dy = 0$ , we obtain

$$\begin{aligned} \int_{-\infty}^{\infty} dy \rho(y) \phi_1(\nu - \eta y/\tau) &= \int_{-\infty}^{\infty} dy \rho(y) C^{[1]} (\nu - \eta y/\tau)^{-\kappa} \\ &= \int_{-\infty}^{\infty} dy \rho(y) C^{[1]} \nu^{-\kappa} \left(1 - \frac{\eta y}{\tau} \nu^{-1}\right)^{-\kappa} \\ &= \int_{-\infty}^{\infty} dy \rho(y) C^{[1]} \nu^{-\kappa} \left(1 + \frac{\kappa \eta y}{\tau} \nu^{-1} + \frac{\eta^2 \kappa(\kappa+1)y^2}{2\tau^2} \nu^{-2} + O(\nu^{-3})\right) \\ &= C^{[1]} \nu^{-\kappa} \left(1 + C^{[2]} \nu^{-2} + O(\nu^{-3})\right) \end{aligned} \quad (149)$$

with the constant  $C^{[2]}$  defined by

$$C^{[2]} := \frac{\eta^2 \kappa(\kappa+1)}{2\tau^2} \int_{-\infty}^{\infty} y^2 \rho(y) dy. \quad (150)$$

Equation (147) is thus equivalent to

$$C^{[1]} C^{[2]} \nu^{-\kappa-2} \simeq \frac{(n-1)C^{[0]}}{\tau} \nu^{-n}. \quad (151)$$

This implies that the exponent  $\kappa$  must satisfy

$$\kappa = n - 2. \quad (152)$$

This means that the subleading term  $\phi_1(\nu)$  is actually negligible when  $\kappa > 0 \iff n > 2$ . We thus confirm that the first term in Eq. (143) can be dropped for FAI maps with dependence on  $\nu$  faster than  $\nu^2$  for large  $\nu$ 's.

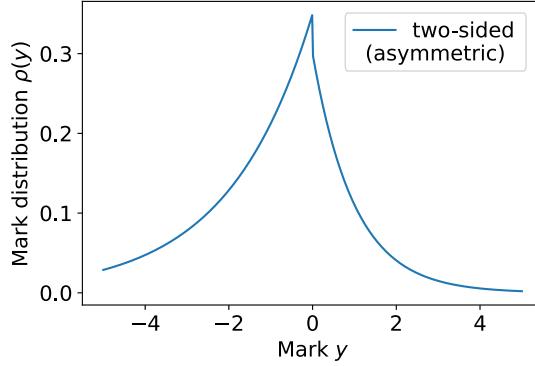


FIG. 7: Schematic of the two-sided asymmetric exponential mark distribution (154) with  $(p_+, p_-, y_+^*, y_-^*) = (0.3, 0.7, 1, 2)$ .

## VI. SOLUTION 3: EXPONENTIAL MEMORY KERNEL WITH TWO-SIDED ASYMMETRIC MARK DISTRIBUTION FOR LINEAR TO FAST-ACCELERATING INTENSITY MAPS

We here study both exact and asymptotic results for the case with the exponential memory kernel, and two-sided asymmetric mark distribution with negative mean mark

$$h(t) = \frac{\eta}{\tau} e^{-t/\tau}, \quad p_+ := \int_0^\infty dy \rho(y) > 0, \quad p_- := \int_{-\infty}^0 dy \rho(y) > 0, \quad m := \int_{-\infty}^\infty y \rho(y) dy < 0. \quad (153)$$

and consider FAI maps  $g(\nu) \gg \nu^2$ .

### A. Exact solution for two-sided asymmetric exponential mark distribution

Let us focus on the case with the two-sided asymmetric exponential mark distribution:

$$\rho(y) = \begin{cases} \frac{p_+}{y_+^*} e^{-y/y_+^*} & (y \geq 0) \\ \frac{p_-}{y_-^*} e^{y/y_-^*} & (y < 0) \end{cases}, \quad (154)$$

where  $p_+ + p_- = 1$ ,  $y_+^* > 0$ , and  $y_-^* > 0$ . The mean mark is given by

$$m := y_+^* p_+ - y_-^* p_- < 0. \quad (155)$$

By using the identities (58) and (81), the ME reads

$$\frac{\partial P_t(\nu)}{\partial t} = \frac{1}{\tau} \partial_\nu [\nu P_t(\nu)] + \partial_\nu \frac{\left( \frac{p_-}{c_-} - \frac{p_+}{c_+} \right) + \frac{\partial_\nu}{c_+ c_-}}{\left( 1 + \frac{\partial_\nu}{c_+} \right) \left( 1 - \frac{\partial_\nu}{c_-} \right)} g(\nu) P_t(\nu), \quad c_\pm := \frac{\tau}{\eta y_\pm^*}. \quad (156)$$

This means that the ME expresses the condition of probability conservation,

$$\frac{\partial P_t(\nu)}{\partial t} = -\frac{\partial J_t(\nu)}{\partial \nu}, \quad J_t(\nu) := -\frac{\nu}{\tau} P_t(\nu) - \frac{\left( \frac{p_-}{c_-} - \frac{p_+}{c_+} \right) + \frac{\partial_\nu}{c_+ c_-}}{\left( 1 + \frac{\partial_\nu}{c_+} \right) \left( 1 - \frac{\partial_\nu}{c_-} \right)} g(\nu) P_t(\nu). \quad (157)$$

By requiring the natural boundary condition  $\lim_{\nu \rightarrow \infty} J_t(\nu) = 0$ , we obtain the second-order partial differential equation that the steady-state PDF  $P_{ss}(\nu)$  satisfies:

$$\tau \left[ A + B \frac{d}{d\nu} \right] \phi(\nu) + \left[ 1 + C \frac{d}{d\nu} - B \frac{d^2}{d\nu^2} \right] \left\{ \frac{\nu}{g(\nu)} \phi(\nu) \right\} = 0, \quad \phi(\nu) := g(\nu) P_{ss}(\nu) \quad (158)$$

with

$$A := \frac{p_-}{c_-} - \frac{p_+}{c_+} := -\frac{\eta}{\tau} m > 0, \quad B := \frac{1}{c_+ c_-} > 0, \quad C := \frac{1}{c_+} - \frac{1}{c_-}. \quad (159)$$

We note that the coefficients are simplified for the symmetric mark distribution  $y_+^* = y_-^*$  and  $p_+ = p_- = 1/2$ , such that  $A = C = 0$ .

*Example 1: ramp tension-intensity map*

For the ramp tension-intensity map

$$\lambda = g(\nu) = \max\{\nu_0, |\nu|\}, \quad (160)$$

we obtain the exact solution for  $\lambda > |\nu_0|$

$$P_{ss}(\lambda) \propto \lambda^{-1} e^{-\lambda/\lambda_{cut}}, \quad \lambda_{cut} := \frac{C + \tau B + \sqrt{(C + \tau B)^2 + 4B(1 + \tau B)}}{2(1 + \tau A)} > 0. \quad (161)$$

under the natural boundary condition. This implies that an exponential tail is observed for this model.

*Example 2: exponential tension-intensity map*

For the exponential tension-intensity map

$$\lambda = g(\nu) = \lambda_0 |\nu| e^{\beta\nu}, \quad (162)$$

we obtain the exact solution for  $\nu > 0$  as

$$\phi(\nu) = C^{[1]} y^{\gamma_1} {}_1F_1(q_1, r_1; y) + C^{[2]} y^{\gamma_2} {}_1F_1(q_2, r_2; y), \quad y(\nu) = \frac{\tau e^{\beta\nu}}{\beta} \quad (163)$$

where

$$\gamma_1 := \frac{C + 2\beta B - \sqrt{4B + C^2}}{2\beta B}, \quad q_1 := 1 + \frac{C + 2A - \sqrt{4B + C^2}}{2\beta B}, \quad r_1 := 1 - \frac{\sqrt{4B + C^2}}{\beta B}, \quad (164a)$$

$$\gamma_2 := \frac{C + 2\beta B + \sqrt{4B + C^2}}{2\beta B}, \quad q_2 := 1 + \frac{C + 2A + \sqrt{4B + C^2}}{2\beta B}, \quad r_2 := 1 + \frac{\sqrt{4B + C^2}}{\beta B}, \quad (164b)$$

and  $C^{[1]}, C^{[2]}$  are two integral constants. For large  $\nu \rightarrow \infty$ , assuming the boundary condition  $\lim_{\nu \rightarrow \infty} P_{ss}(\nu) = 0$ , we obtain the asymptotic formula

$$\phi(\nu) \propto y^{-A/(\beta B)}, \quad (165)$$

leading to the PDF tail of the intensity:

$$P_{ss}(\lambda) \propto \lambda^{-2-\beta^{-1}u}, \quad u = \frac{A}{B} = -\frac{\tau m}{\eta y_+^* y_-^*}. \quad (166)$$

This result implies that the power-law scaling deviates from Zipf's law in proportion to the amplitude  $m$  of the asymmetry of the mark distribution.

## B. Robust asymptotic solutions

In this subsection, we generalise the above exact results in the form of robust asymptotic results under a wide range of two-sided mark distributions with negative mean, in the presence of an exponential memory kernel.

### 1. Robust exponential tail for the ramp intensity

Let us assume that the ramp intensity is asymptotically

$$g(\nu) \simeq \nu - \nu_1 \quad \text{for large } \nu \quad (167a)$$

with an exponential memory with two-sided mark distribution of negative mean:

$$h(t) = \frac{\eta}{\tau} e^{-t/\tau}, \quad p_+ := \int_0^\infty \rho(y) dy > 0, \quad p_- := \int_{-\infty}^0 \rho(y) dy > 0, \quad m := \int_{-\infty}^\infty y \rho(y) dy < 0. \quad (167b)$$

Here  $\nu_1$  is an arbitrary real number, either positive or nonpositive in contrast to the LHawkes process. Under this assumption, we obtain

$$P_{ss}(\lambda) \propto e^{-\lambda/\lambda_{cut}} \quad \text{for large } \lambda. \quad (167c)$$

The parameter  $\lambda_{cut}$  is given by the self-consistent relation

$$\frac{1}{\tau \lambda_{cut}} = \Phi \left( \frac{\eta}{\tau \lambda_{cut}} \right), \quad \Phi(x) := \int_{-\infty}^\infty dy \rho(y) (e^{xy} - 1), \quad (168)$$

where  $\Phi(x)$  is the moment-generating function. The equation for  $\lambda_{cut}$  has a single positive solution (see Appendix F). This relation can be derived by a straightforward generalisation of the derivation in Sec. V C 1.

## 2. Robust power law tail for fast-accelerating intensity maps

We show that, under the following general assumptions

$$h(t) = \frac{\eta}{\tau} e^{-t/\tau}, \quad g(\nu) \gg \nu^2 \quad (\text{for large } \nu), \quad p_+ := \int_0^\infty \rho(y) dy > 0, \quad p_- := \int_{-\infty}^0 \rho(y) dy > 0, \quad m := \int_{-\infty}^\infty y \rho(y) dy < 0, \quad (169)$$

we obtain the robust asymptotic relationship

$$P_{ss}(\lambda) \propto \lambda^{-1} \left[ \left| \frac{dg(\nu)}{d\nu} \right|^{-1} e^{-u\nu} \right]_{\nu=g^{-1}(\lambda)}, \quad u := \frac{\tau c^*}{\eta}, \quad (170)$$

where  $c^*$  is the unique positive root of  $\Phi(c^*) = 0$ , where the moment-generating function is defined by  $\Phi(x) := \int_{-\infty}^\infty dy \rho(y) (e^{xy} - 1)$ .

*Examples.* From this formula, we readily deduce the power law PDF for the exponential intensity

$$g(\nu) \simeq \lambda_0 e^\beta \implies P_{ss}(\lambda) \propto \lambda^{-2-\beta^{-1}u}. \quad (171)$$

We note that this result is consistent with the aforementioned exact result (166) by considering Eq. (F19) in Appendix F for the case with the exponential intensity and the two-sided asymmetric exponential mark distribution.

In addition, we obtain the truncated power law PDF for the polynomial intensity

$$g(\nu) \simeq \lambda_0 \nu^n, \quad n > 2 \implies P_{ss}(\lambda) \propto \lambda^{-2+\frac{1}{n}} e^{-u \left( \frac{\lambda}{\lambda_0} \right)^{\frac{1}{n}}}, \quad (172)$$

where the cutoff length appears due to the asymmetry of the mark distribution. For the zero mean mark limit  $m \uparrow 0$ , the cutoff disappears as  $u \downarrow 0$ .

*Derivation.* By defining  $\phi(\nu) = g(\nu) P_{ss}(\nu)$ , the ME is given by

$$\frac{1}{\tau} \frac{\partial}{\partial \nu} \left\{ \frac{\nu}{g(\nu)} \phi(\nu) \right\} + \int_{-\infty}^\infty dy \rho(y) \phi(\nu - \eta y / \tau) - \phi(y) = 0. \quad (173)$$

As an asymptotic assumption for the solution, let us first neglect the first term of Eq. (173) to obtain

$$\int_{-\infty}^\infty dy \rho(y) \phi(\nu - \eta y / \tau) - \phi(y) \simeq 0 \quad \text{for large } \nu \quad (174)$$

for case (169). The self-consistency of this assumption will be confirmed later. According to Appendix E, the general solution is given by the superposition of exponentials,

$$\phi(\nu) \simeq \sum_i C_i e^{-(\tau c_i / \eta) \nu}, \quad (175)$$

where the  $c_i$ 's are the roots of the moment-generating function  $\Phi(x) = 0$ . The moment-generating function is defined by

$$\Phi(c) = 0, \quad \Phi(x) := \int_{-\infty}^{\infty} dy \rho(y) (e^{xy} - 1), \quad (176)$$

whose analytical characters are summarised in Appendix F. According to Appendix F,  $\Phi(x) = 0$  has only two roots at  $x = 0$  and  $x = c^* > 0$ . This means that the general asymptotic solution is given by

$$\phi(\nu) \simeq C_0 e^{-(\tau c^*/\eta)\nu} + C_1 \quad (177)$$

with integral constants  $C_0$  and  $C_1$ . By imposing the natural boundary condition,  $C_1$  must be zero (see below for the natural boundary condition). We thus have the solution

$$\phi(\nu) \simeq C_0 e^{-(\tau c^*/\eta)\nu}. \quad (178)$$

This implies that the steady-state intensity PDF has the following asymptotic form

$$P_{ss}(\nu) = \frac{\phi(\nu)}{g(\nu)} \propto \frac{1}{g(\nu)} e^{-(\tau c^*/\eta)\nu}, \quad (179)$$

which implies Eq. (170) from the Jacobian relation  $P_{ss}(\lambda) = P_{ss} |d\nu/d\lambda|$ .

*Self-consistency of the assumption.* Finally, we here confirm the self-consistency of the ansatz for the solution under the assumption of FAI maps. Let us assume that the solution is given by the following expansion

$$\phi(\nu) = \phi_0(\nu) + \phi_1(\nu) + \dots, \quad \phi_0(\nu) := C^{[0]} e^{-(\tau c^*/\eta)\nu}, \quad |\phi_0(\nu)| \gg |\phi_1(\nu)| \quad \text{for large } \nu \quad (180)$$

with an integral constant  $C^{[0]}$ . For simplicity, let us focus on the case  $g(\nu) = \nu^n$  with integer  $n > 2$ . By assuming that the first term in Eq. (173) is subleading, we substitute this expansion into Eq. (173) to obtain

$$\begin{aligned} \int_{-\infty}^{\infty} dy \rho(y) \phi_1(\nu - \eta y/\tau) - \phi_1(y) &\simeq -\frac{1}{\tau} \frac{\partial}{\partial \nu} \{ \nu^{1-n} \phi_0(\nu) \} \\ \implies \int_{-\infty}^{\infty} dy \rho(y) \phi_1(\nu - \eta y/\tau) - \phi_1(y) &\simeq C^{[0]} \frac{c^*}{\eta} \nu^{1-n} e^{-(\tau c^*/\eta)\nu}. \end{aligned} \quad (181)$$

We make the anzatz for the solution in the form

$$\phi_1(\nu) \simeq C^{[1]} \nu^{-\kappa} e^{-(\tau c^*/\eta)\nu}, \quad \kappa > 0 \quad (182)$$

to obtain the special solution with a constant  $C^{[1]}$ . Here the condition  $\kappa > 0$  is essential; otherwise the consistency relationship  $|\phi_0(\nu)| \gg |\phi_1(\nu)|$  does not hold. By direct substitution, Eq. (181) is equivalent to

$$\begin{aligned} \int_{-\infty}^{\infty} dy \rho(y) \phi_1(\nu - \eta y/\tau) &\simeq C^{[1]} \int_{-\infty}^{\infty} dy \rho(y) (\nu - \eta y/\tau)^{-\kappa} e^{c^* y - (\tau c^*/\eta)\nu} \\ &= C^{[1]} \int_{-\infty}^{\infty} dy \rho(y) \nu^{-\kappa} \left( 1 - \frac{\eta y}{\tau} \nu^{-1} \right)^{-\kappa} e^{c^* y} e^{-(\tau c^*/\eta)\nu} \\ &= C^{[1]} \int_{-\infty}^{\infty} dy \rho(y) \nu^{-\kappa} \left( 1 + \frac{\kappa \eta y}{\tau} \nu^{-1} + O(\nu^{-2}) \right) e^{c^* y} e^{-(\tau c^*/\eta)\nu} \\ &= C^{[1]} \nu^{-\kappa} e^{-(\tau c^*/\eta)\nu} \left( \Phi(c^*) + 1 + C^{[2]} \nu^{-1} + O(\nu^{-2}) \right) \end{aligned} \quad (183)$$

with

$$C^{[2]} := \frac{\kappa \eta}{\tau} \int_{-\infty}^{\infty} y e^{c^* y} \rho(y) dy. \quad (184)$$

By using  $\Phi(c^*) = 0$ , we thus obtain

$$C^{[1]} C^{[2]} \nu^{-\kappa-1} e^{-(\tau c^*/\eta)\nu} \simeq C^{[0]} \frac{c^*}{\eta} \nu^{1-n} e^{-(\tau c^*/\eta)\nu}. \quad (185)$$

This implies that the power law exponent  $\kappa$  must satisfy the relationship

$$\kappa = n - 2. \quad (186)$$

Because of the assumption  $\kappa > 0$ , we obtain the self-consistency condition

$$n > 2, \quad (187)$$

which is equivalent to the assumption that the tension-intensity maps must be FAI ( $g(\nu) \gg \nu^2$ ).

*Natural boundary condition.* Here we impose the natural boundary condition to remove  $C_1$ . Let us use the KM expansion (33) to define the probability current as

$$\frac{\partial P_t(\nu)}{\partial t} = -\frac{\partial}{\partial \nu} J_t(\nu), \quad J_t(\nu) := -\frac{1}{\tau} [\nu P_t(\nu)] - \sum_{k=1}^{\infty} \frac{(-1)^k \alpha_k}{k!} \frac{\eta^k}{\tau^k} \frac{\partial^{k-1}}{\partial \nu^{k-1}} g(\nu) P_t(\nu). \quad (188)$$

For the steady-state distribution, let us ignore the first term in  $J_t(\nu)$  for large  $\nu$  to obtain

$$J_{ss}(\nu) \simeq - \sum_{k=1}^{\infty} \frac{(-1)^k \alpha_k}{k!} \frac{\eta^k}{\tau^k} \frac{\partial^{k-1}}{\partial \nu^{k-1}} g(\nu) P_{ss}(\nu) \quad \text{for large } \nu. \quad (189)$$

By direct substitution of the general solution  $g(\nu)P_{ss}(\nu) = \phi(\nu) = C_0 + C_1 e^{-(\tau c^*/\eta)\nu}$ , we obtain

$$\begin{aligned} J_{ss}(\nu) &\simeq - \sum_{k=1}^{\infty} \frac{(-1)^k \alpha_k}{k!} \frac{\eta^k}{\tau^k} \frac{\partial^{k-1}}{\partial \nu^{k-1}} \left( C_1 + C_0 e^{-(\tau c^*/\eta)\nu} \right) \\ &= \frac{\eta m}{\tau} C_1 + \frac{\eta C_0}{c^* \tau} e^{-(\tau c^*/\eta)\nu} \sum_{k=1}^{\infty} \frac{\alpha_k}{k!} c^{*k} \\ &= \frac{\eta m}{\tau} C_1 + \frac{\eta C_0}{c^* \tau} e^{-(\tau c^*/\eta)\nu} \Phi(c^*) \\ &= \frac{\eta m}{\tau} C_1, \end{aligned} \quad (190)$$

where we have used  $\Phi(x) = \sum_{k=1}^{\infty} (\alpha_k/k!)x^k$  and  $\Phi(c^*) = 0$ . Since the natural boundary condition implies  $\lim_{\nu \rightarrow \infty} J_t(\nu) = 0$  for any  $t$ , we obtain  $C_1 = 0$ .

## VII. SOLUTION 4: GENERAL MEMORY KERNEL FOR ONE-SIDED MARK DISTRIBUTION FOR THE RAMP HAWKES PROCESS

We have studied the exact solution for the NLHawkes process assuming that (i) the memory kernel is exponential and (ii) the jump size obeys the one-sided exponential distributions. In particular, we derived the power law tail (67) without truncation, at the critical point for the ramp Hawkes process (65). As shown in the following in this section, this exact power law relation is robust for general ramp Hawkes processes with any memory kernel and jump-size distribution, only assuming the finiteness of

$$\langle \tau \rangle := \int_0^\infty t h(t) dt = \int_0^\infty x^2 \tilde{h}(x) dx < \infty, \quad (191)$$

where  $\tilde{h}(x)$  has been defined in (14a). Note that the critical condition is characterised by

$$\eta := \int_0^\infty h(t) dt = \int_0^\infty x \tilde{h}(x) dx = 1. \quad (192)$$

### A. Discrete sum of exponentials

Let us first consider the case of a discrete sum of exponentials. In this case, we find a power law asymptotics at the critical point  $\eta = 1$

$$\begin{aligned} h(t) &= \sum_{k=1}^K \tilde{h}_k e^{-t/\tau_k}, \quad \lambda = g(\nu) \simeq \nu - \nu_1 + o(\nu^0) \quad \text{for large } \nu, \quad \rho(y) = 0 \quad \text{for negative } y, \quad \int_0^\infty y \rho(y) dy = 1 \\ \implies P_{ss}(\lambda) &\propto \lambda^{-1-a}, \quad a := \frac{2\nu_1 \langle \tau \rangle}{\alpha_2}, \quad \langle \tau \rangle := \sum_{k=1}^K \tau_k^2 \tilde{h}_k \end{aligned} \quad (193)$$

for either negative or non-negative  $\nu_1$ . This relation is a true power law for positive  $\nu_1$  (i.e. normalizable even without cutoff), while it is an intermediate asymptotics for non-positive  $\nu_1$  (i.e., not normalizable without cutoff). Note that the critical condition is given by

$$\eta := \sum_{k=1}^K \tau_k \tilde{h}_k = 1. \quad (194)$$

*Derivation.*

Let us first write the asymptotic form of  $P_{ss}(\mathbf{z})$  as  $S(\mathbf{z})$ :

$$P_{ss}(\mathbf{z}) = S(\mathbf{z}) + R(\mathbf{z}), \quad S(\mathbf{z}) \gg R(\mathbf{z}) \text{ for large } \mathbf{z} \quad (195)$$

with a correction term  $R(\mathbf{z})$  for small  $\mathbf{z}$ . Here  $S(\mathbf{z})$  is assumed to have a fat tail represented by a power law, while  $R(\mathbf{z})$  is assumed to have a thinner tail. The ME (23) in the steady state reduces asymptotically to

$$\sum_{k=1}^K \frac{\partial}{\partial z_k} \frac{z_k}{\tau_k} S(\mathbf{z}) + \int_0^\infty dy \rho(y) \left[ \sum_{k=1}^K (z_k - y \tilde{h}_k) - \nu_1 \right] S(\mathbf{z} - y \tilde{\mathbf{h}}) - \left[ \sum_{k=1}^K z_k - \nu_1 \right] S(\mathbf{z}) \simeq 0 \quad \text{for large } \nu. \quad (196)$$

Since the asymptotic form of this ME is the same as that for the LHawkes process presented in Ref. [16] except that  $\nu_1$  can be either negative or non-negative, its asymptotic solution for large  $\nu$  can be obtained from a similar calculation to that presented in Ref. [16]. While we refer the reader to Ref. [16] for an elementary introduction to the calculations, let us sketch the main steps of the derivation. We first define the Laplace transformations,

$$\tilde{P}_{ss}(\mathbf{s}) := \mathcal{L}_K[P_{ss}(\mathbf{z}); \mathbf{s}], \quad \tilde{S}(\mathbf{s}) := \mathcal{L}_K[S(\mathbf{z}); \mathbf{s}], \quad \tilde{R}(\mathbf{s}) := \mathcal{L}_K[R(\mathbf{z}); \mathbf{s}]. \quad (197)$$

Since  $P_{ss}(\mathbf{z})$  is a PDF, the normalisation implies  $\int d\mathbf{z} P_{ss}(\mathbf{z}) = \tilde{P}_{ss}(\mathbf{s} = \mathbf{0}) = 1$ . However,  $S(\mathbf{z})$  is just an asymptotic form of the PDF, and there is no guarantee that  $\tilde{S}(\mathbf{s} = \mathbf{0}) = 1$ . For example, assuming that  $\tilde{S}(s\mathbf{1}) = As^a + o(s^a)$  with any non-integer number  $a$  and indicator vector  $\mathbf{1} := (1, 1, \dots, 1)$ , we can expand  $\tilde{P}_{ss}(s) := \tilde{P}_{ss}(s\mathbf{1})$  as

$$\tilde{P}_{ss}(s) = \tilde{S}(s\mathbf{1}) + \tilde{R}(s\mathbf{1}) \simeq As^a + \sum_{k=0}^m c_k s^k + o(s^a), \quad \tilde{R}(s\mathbf{1}) = \sum_{k=0}^m c_k s^k + o(s^a) \quad (198)$$

for small  $s$  with  $m := \max(\lfloor a \rfloor, 0)$  and the floor function  $\lfloor x \rfloor = \max\{k \in \mathbf{Z} \mid k \leq x\}$  with the set of integers  $\mathbf{Z}$ . The normalisation condition requires  $\tilde{P}(s = 0) = c_0 = 1$ . By applying the Laplace transformation to the steady-state ME (196), we obtain

$$-\sum_{k=1}^K \frac{s_k}{\tau_k} \frac{\partial \tilde{S}(\mathbf{s})}{\partial s_k} - (\Phi(\mathbf{s}) - 1) \left( \nu_1 + \sum_{k=1}^K \frac{\partial}{\partial s_k} \right) \tilde{S}(\mathbf{s}) \simeq 0, \quad \Phi(\mathbf{s}) := \int_0^\infty dy \rho(y) e^{-y \tilde{\mathbf{h}} \cdot \mathbf{s}}, \quad (199)$$

which is valid for small  $\mathbf{s}$ . Considering  $(1/\tilde{S}(\mathbf{s})) \partial \tilde{S}(\mathbf{s}) / \partial s_k = (\partial/\partial s_k) \log |\tilde{S}(\mathbf{s})|$ , this equation can be rewritten as

$$\sum_{k=1}^K \left( 1 - \Phi(\mathbf{s}) - \frac{s_k}{\tau_k} \right) \frac{\partial \Psi(\mathbf{s})}{\partial s_k} \simeq \nu_1 (\Phi(\mathbf{s}) - 1), \quad \Psi(\mathbf{s}) := \log |\tilde{S}(\mathbf{s})|. \quad (200)$$

Since this equation belongs to the class of first-order partial differential equations, it can be solved by the method of characteristics. Let us thus consider the corresponding Lagrange-Charpit equations:

$$\frac{ds_k}{dl} = 1 - \frac{s_k}{\tau_k} - \Phi(\mathbf{s}), \quad \frac{d\Psi}{dl} = \nu_1 (\Phi(\mathbf{s}) - 1) \quad (201)$$

with a parameter  $l$  describing the position on characteristic curves. By regarding  $l$  as an imaginary “time” of this system, we can apply the standard bifurcation theory of dynamical systems. Since we are interested in the regime of small  $\mathbf{s}$ , let us consider the “long time” asymptotic limit  $l \rightarrow \infty$ , where  $\mathbf{s}(l)$  relaxes to the attractor at  $\mathbf{s} = \mathbf{0}$  (see the

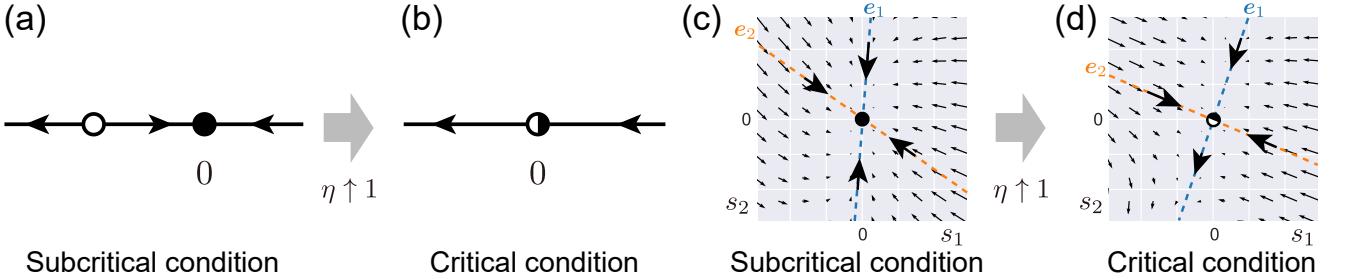


FIG. 8: Schematics of the “vector field”  $\mathbf{V}(\mathbf{s}) := d\mathbf{s}/dl$  for the special cases  $K = 1$  (Figs. a and b below and at criticality, respectively) and  $K = 2$  (Figs. c and d below and at criticality, respectively). The origin  $\mathbf{s} = \mathbf{0}$  is the stable attractor below criticality  $\eta < 1$ . An unstable attractor merges into  $\mathbf{s} = \mathbf{0}$  at criticality  $\eta \uparrow 1$ , which is consistent with the standard picture of the transcritical bifurcation.

schematic figures 8a and c for the vector field  $\mathbf{V}(\mathbf{s}) := d\mathbf{s}/dl$  for the cases  $K = 1$  and  $K = 2$  below criticality  $\eta < 1$ , respectively), such that  $\lim_{l \rightarrow \infty} \mathbf{s}(l) = \mathbf{0}$ . Let us expand the equations for small  $s$ :

$$\frac{ds_k}{dl} \simeq -\mathbf{H}\mathbf{s} - \frac{\alpha_2}{2} \left( \sum_{j=1}^K \tilde{h}_j s_j \right)^2 + o(s^2), \quad (202a)$$

$$\frac{d\Psi}{dl} \simeq -\nu_1 \mathbf{K}\mathbf{s} + o(s), \quad (202b)$$

with

$$\mathbf{H} := \begin{pmatrix} \frac{1}{\tau_1} - \tilde{h}_1, & -\tilde{h}_2, & \dots & -\tilde{h}_K \\ -\tilde{h}_1, & \frac{1}{\tau_2} - \tilde{h}_2, & \dots & -\tilde{h}_K \\ \vdots & \vdots & \ddots & \vdots \\ -\tilde{h}_1, & -\tilde{h}_2, & \dots & \frac{1}{\tau_K} - \tilde{h}_K \end{pmatrix}, \quad \mathbf{K} := (\tilde{h}_1, \dots, \tilde{h}_K). \quad (203)$$

Note that the matrix  $\mathbf{H}$  is the same as that in Ref. [16, 17]. Defining its eigenvalues  $\{\lambda_k\}_{k=1,\dots,K}$  and corresponding eigenvectors  $\{\mathbf{e}_k\}_{k=1,\dots,K}$  by

$$\mathbf{H}\mathbf{e}_k = \lambda_k \mathbf{e}_k, \quad (204)$$

$\mathbf{H}$  has the following mathematical properties (see Ref. [17] and Appendix G for details):

1. All the eigenvalues are real:  $\lambda_k \in \mathbf{R}^1$ . Accordingly, we assume that  $\lambda_i \leq \lambda_j$  for  $i < j$ .
2. The determinant of  $\mathbf{H}$  is given by

$$\det \mathbf{H} = \frac{1 - \sum_{k=1}^K \tau_k \tilde{h}_k}{\prod_{k=1}^K \tau_k}. \quad (205)$$

This means that the zero eigenvalue appears at criticality  $\eta := \sum_{k=1}^K \tau_k \tilde{h}_k = 1$ .

3.  $\mathbf{H}$  can be diagonalised by  $\mathbf{P}$  as follows:

$$\mathbf{P} := (\mathbf{e}_1, \dots, \mathbf{e}_K), \quad \mathbf{P}^{-1} \mathbf{H} \mathbf{P} = \begin{pmatrix} \lambda_1, & 0, & \dots & 0 \\ 0, & \lambda_2, & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0, & 0, & \dots & \lambda_K \end{pmatrix}. \quad (206)$$

4. Let us introduce a new representation based on the eigenvectors:

$$\mathbf{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_K \end{pmatrix} := \mathbf{P}^{-1} \mathbf{s}, \quad \mathbf{s} = \begin{pmatrix} s_1 \\ s_2 \\ \vdots \\ s_K \end{pmatrix}, \quad \mathbf{P}^{-1} = \begin{pmatrix} \mathbf{g}_1 \\ \mathbf{g}_2 \\ \vdots \\ \mathbf{g}_K \end{pmatrix}. \quad (207)$$

At criticality  $\eta = 1$ , the smallest eigenvalue is zero,  $\lambda_1 = 0$ , and its eigenvector is given by

$$\mathbf{e}_1 = \begin{pmatrix} \tau_1 \\ \tau_2 \\ \vdots \\ \tau_K \end{pmatrix}. \quad (208)$$

In addition,  $X_1$  is represented by

$$X_1 = \mathbf{g}_1 \cdot \mathbf{s} = \frac{1}{\langle \tau \rangle} \sum_{k=1}^K \tau_k \tilde{h}_k s_k, \quad \mathbf{g}_1 = \left( \frac{\tau_1 \tilde{h}_1}{\langle \tau \rangle}, \dots, \frac{\tau_K \tilde{h}_K}{\langle \tau \rangle} \right), \quad \langle \tau \rangle := \sum_{k=1}^K \tau_k^2 \tilde{h}_k. \quad (209)$$

Given these properties, let us consider the Lagrange-Charpit equation (202a) in the representation  $\mathbf{X} := (X_1, \dots, X_K)^T$ . At criticality  $\eta = 1$ , the leading order contribution in the Lagrange-Charpit Eq. (202a) is given by

$$\frac{dX_1}{dl} \simeq 0 - \frac{\alpha_2}{2\langle \tau \rangle} \left( \sum_{k=1}^K \tilde{h}_k s_k \right)^2 + o(\mathbf{X}^2), \quad \frac{dX_j}{dl} = -\lambda_j X_j + o(\mathbf{X}) \text{ for } j \geq 2. \quad (210)$$

Since the leading order contribution will come from the  $X_1$  direction, we can assume that  $|X_1| \gg |X_j|$  for  $j \geq 2$  for large  $l$ . We thus ignore contribution other than  $X_1$  by assuming  $X_j \simeq 0$  for  $j \geq 2$ :

$$\mathbf{s} = \mathbf{P} \mathbf{X} \simeq (\mathbf{e}_1, \dots, \mathbf{e}_K) \begin{pmatrix} X_1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = X_1 \mathbf{e}_1. \quad (211)$$

We thus obtain

$$\frac{dX_1}{dl} \simeq 0 - \frac{\alpha_2}{2\langle \tau \rangle} X_1^2 + o(\mathbf{X}^2), \quad \frac{dX_j}{dl} = -\lambda_j X_j + o(\mathbf{X}) \text{ for } j \geq 2. \quad (212)$$

This is the standard normal form of the transcritical bifurcation when regarding  $l$  as a physical time (see the schematic figures 8b and d for  $K = 1$  and  $K = 2$  at criticality  $\eta \uparrow 1$ , respectively). The solution is given by

$$X_1(l) \simeq \frac{2\langle\tau\rangle}{\alpha_2} \frac{1}{l - l_0}, \quad X_j(l) \simeq C_j e^{-\lambda_j(l - l_0)} \quad (213)$$

with integral constants  $l_0$  and  $C_j$  for  $j \geq 2$ . We can assume  $l_0 = 0$  as the initial point of the characteristic curve without losing generality. From expanding  $\Phi(s)$ , we obtain the solution

$$\Psi(l) \simeq -\nu_1 \int dl \sum_{k=1}^K \tilde{h}_k s_k(l) \simeq \frac{2\nu_1\langle\tau\rangle}{\alpha_2} \log |X|_1 + O(\mathbf{X}) + C_0 \quad (214)$$

with an integral constant  $C_0$ . According to the method of characteristics, the general solution is given by

$$\mathcal{H}(C_2, \dots, C_K) = C_0 \quad (215)$$

with a function  $\mathcal{H}$  which needs to be determined by the initial condition. The constants  $C_j$  with  $j \geq 2$  are related to each other, such that

$$l = \frac{2\langle\tau\rangle}{\alpha_2 X_1}, \quad C_j = X_j \exp \left\{ \frac{2\langle\tau\rangle\lambda_j}{\alpha_2 X_1} \right\}. \quad (216)$$

This means that the explicit form of the general solution is given by

$$\Psi(\mathbf{X}) \simeq \frac{2\nu_1\langle\tau\rangle}{\alpha_2} \log |X_1| + O(\mathbf{X}) + \mathcal{H} \left( X_2 \exp \left\{ \frac{2\langle\tau\rangle\lambda_2}{\alpha_2 X_1} \right\}, \dots, X_K \exp \left\{ \frac{2\langle\tau\rangle\lambda_K}{\alpha_2 X_1} \right\} \right). \quad (217)$$

Note the existence of the divergent term  $\log |X_1|$  resulting from neglecting the UV cutoff. Since  $\Psi(\mathbf{X})$  must be constant for  $\mathbf{s} \rightarrow \mathbf{0}$ , except for the artificial log divergence, we obtain

$$\lim_{\mathbf{X} \rightarrow 0} \mathcal{H} \left( X_2 \exp \left\{ \frac{2\langle\tau\rangle\lambda_2}{\alpha_2 X_1} \right\}, \dots, X_K \exp \left\{ \frac{2\langle\tau\rangle\lambda_K}{\alpha_2 X_1} \right\} \right) = \text{const.} \quad (218)$$

Let us now consider the specific limit  $X_1 \rightarrow 0$ , by writing

$$X_j = Z_j \exp \left\{ -\frac{2\langle\tau\rangle\lambda_j}{\alpha_2 X_1} \right\} \quad (219)$$

with any positive number  $Z_j$  for  $j \geq 2$ . This specific limit satisfies the relation,

$$\lim_{X_1 \rightarrow 0} \mathbf{X} = \mathbf{0}. \quad (220)$$

Since Eq. (218) should hold for any path taken to reach the limit  $\mathbf{X} \rightarrow 0$ , we obtain the relation even for the specific limit

$$\lim_{X_1 \rightarrow 0} \mathcal{H}(Z_2, \dots, Z_K) = \text{const.} \quad (221)$$

for any positive  $\{Z_j\}_{j=2,\dots,K}$ , implying that  $\mathcal{H}$  is a constant function. We thus obtain

$$|\tilde{S}(\mathbf{z})| = \exp \Psi(\mathbf{X}) \simeq \tilde{A} s^a, \quad a := \frac{2\nu_1\langle\tau\rangle}{\alpha_2} \quad (222)$$

with some positive number  $\tilde{A}$ .

a. *Case with negative  $a < 0$ .* When  $\nu_1$  is negative, we have

$$\tilde{P}_{ss}(s) \simeq As^a + o(s^a) \quad (223)$$

for small  $s$  with some constant  $A$  satisfying  $|A| = \tilde{A}$  and negative value  $a < 0$ . By applying the inverse Laplace transform (see Appendix H 1), we obtain the power law asymptotic form (193). In this case, the sign of  $A$  is determined to be positive (i.e.,  $A = \tilde{A}$ ) for the consistency with the probability interpretation.

b. *Case with  $0 < a < 1$ .* This case is equivalent to  $m := \lfloor a \rfloor = 0$ . We obtain

$$\tilde{P}_{ss}(s) \simeq 1 - As^a + o(s^a) \quad (224)$$

with some constant  $A$  for small  $s$ . Assuming that  $\tilde{A}$  is a positive real number, we obtain the power law asymptotic form (193) (see Appendix H 2).

c. *Case with positive non-integer  $a$ .* Let us define  $m := \lfloor a \rfloor$  in order to classify the solutions. Since the asymptotic series of the Laplace transformation is given by

$$\tilde{P}_{ss}(s) \simeq As^a + \sum_{k=0}^m c_k s^k + o(s^a), \quad (225)$$

we obtain the power law asymptotic form (193) (see Appendix H 3), by setting  $A$  to a positive (negative) number for even (odd)  $m$ , for consistency with the probability interpretation.

d. *Case with positive integer  $a$ .* Technically, the positive integer case requires a special treatment since the gamma function in the Laplace transformation formula (H14) diverges:  $\Gamma(-a) = \infty$ . However, since the power law asymptotics (193) is valid for any non-integer  $a$ , it is straightforward to obtain the power law asymptotics (193) for positive integer  $a$ , assuming that the power law exponent  $a$  is a continuous function in terms of  $\nu_1$ :

$$a(\nu_1) = \lim_{x \rightarrow \nu_1} a(x). \quad (226)$$

While we have numerically checked the validity of this result (193) for some specific cases (see Sec. VII C for the numerical results), a rigorous proof of the continuity assumption (226) is beyond the scope of this paper, as it requires further technical investigation while the continuity assumption (226) is physically reasonable.

In summary, we obtain the power law asymptotics (193) for general  $a$ .

## B. General memory kernel

Since the power law asymptotics (193) holds for general discrete sums of exponentials, as a straightforward generalisation, we find a power law asymptotics at the critical point  $n = 1$  for general memory kernel  $h(t)$ , such that

$$\begin{aligned} h(t) &= \int_0^\infty dx \tilde{h}(x) e^{-t/x}, \quad \lambda = g(\nu) \simeq \nu - \nu_1 + o(\nu^0) \text{ for large } \nu, \quad \rho(y) = 0 \text{ for negative } y, \quad \int_0^\infty y \rho(y) dy = 1 \\ &\implies P_{ss}(\lambda) \propto \lambda^{-1-a}, \quad a := \frac{2\nu_1 \langle \tau \rangle}{\alpha_2}, \quad \langle \tau \rangle := \int_0^\infty x^2 \tilde{h}(x) dx \end{aligned} \quad (227)$$

for either negative or non-negative  $\nu_1$ . This relation is a true power law for positive  $\nu_1$  (i.e. normalisable even without cutoff), while it is an intermediate asymptotics for non-positive  $\nu_1$  (i.e., not normalisable without cutoff).

## C. Numerical confirmation

Figure 9 shows the numerical results based on the Monte Carlo simulation of the SDE (3) obtained for a memory function constructed as a discrete sum of exponentials, for the ramp intensity (65), and a mark distribution reducing to the Dirac function centred on  $y = 1$ :

$$h(t) = \sum_{k=1}^K \tilde{h}_k e^{-t/\tau_k}, \quad \lambda = g(\nu) = \max\{\nu_0, \nu - \nu_1\}, \quad \rho(y) = \delta(y - 1). \quad (228)$$

The simulations are performed by using an adaptive time discretisation scheme (see Appendix I for the detailed numerical scheme). All panels in Fig. 9 exhibit the predicted power law tail of the intensity distribution, in excellent agreement with our theoretical prediction (193). Notably, the power law exponents varies continuously as a function of  $\nu_1$  and the power law formula (193) is found to be valid even for integer exponents such as  $a = 0$  and  $a = 1$ .

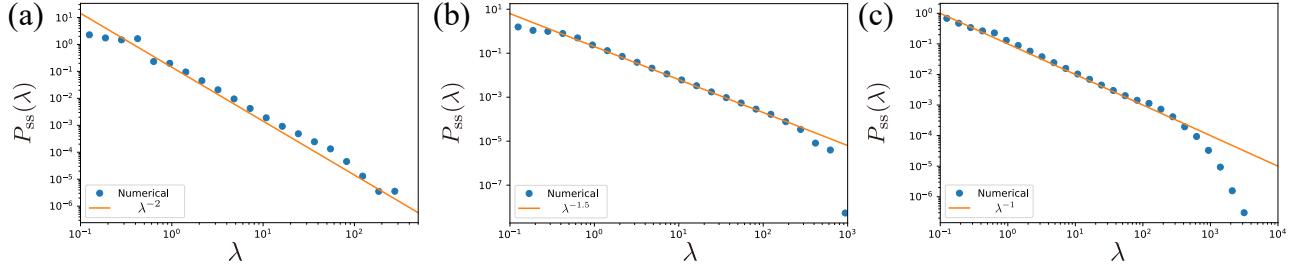


FIG. 9: Numerical confirmation of our theoretical prediction on the power law exponents (193). (a) Simulation based on  $K = 2$ ,  $(\tau_1, \tau_2) = (1, 2)$ ,  $(\tilde{h}_1, \tilde{h}_2) = (0.7, 0.14995)$ ,  $\eta = 0.9999$ ,  $\nu_0 = 0.01$ , and  $\nu_1 \simeq 0.385$ , predicting  $a \simeq 1.0$  (i.e., a true power law; Zipf's law). (b) Simulation based on  $K = 3$ ,  $(\tau_1, \tau_2, \tau_3) = (1, 2, 3)$ ,  $(\tilde{h}_1, \tilde{h}_2, \tilde{h}_3) = (0.5, 0.15, 0.1999/3)$ ,  $\eta = 0.9999$ ,  $\nu_0 = 0.01$ , and  $\nu_1 \simeq 0.147$ , predicting  $a \simeq 0.5$  (i.e., a true power law). (c) Simulation based on  $K = 3$ ,  $(\tau_1, \tau_2, \tau_3) = (1, 2, 3)$ ,  $(\tilde{h}_1, \tilde{h}_2, \tilde{h}_3) = (0.5, 0.15, 0.1999/3)$ ,  $\eta = 0.9999$ ,  $\nu_0 = 0.01$ , and  $\nu_1 = 0$ , predicting  $a = 0$  (i.e., an intermediate power law asymptotic).

### VIII. SOLUTION 5: GENERAL MEMORY KERNEL FOR FAST-ACCELERATING INTENSITY MAPS AND TWO-SIDED MARK DISTRIBUTION WITH NONPOSITIVE MEAN MARK

In Secs. V and VI, we have shown that a general asymptotic formula is available for the exponential memory and the two-sided mark distributions with non-positive mean mark. Here we show that, by solving the corresponding MEs, the asymptotic formula is valid for a wider class of memory kernels with FAI.

#### A. Discrete sum of exponentials

We first show that the power law tail of the PDF of intensities is robust for various memory kernel  $h(t)$  for the MSA intensity function in the presence of a two-sided mark distribution with non-positive mean mark. Specifically, we make the following assumptions (i.e., discrete-sum of exponentials, MSA intensity, two-sided mark distribution with nonpositive mean mark)

$$h(t) = \sum_{k=1}^K \tilde{h}_k e^{-t/\tau_k}, \quad g(\nu) = \lambda_0 e^{\beta\nu}, \quad p_+ := \int_0^\infty \rho(y) dy > 0, \quad p_- := \int_{-\infty}^0 \rho(y) dy > 0, \quad m := \int_{-\infty}^\infty y \rho(y) dy \leq 0. \quad (229)$$

Under these conditions, we obtain the power law intensity PDF

$$P_{ss}(\lambda) \propto \lambda^{-2-\beta^{-1}u} \quad \text{for large } \lambda, \quad u := \frac{c^*}{h(0)}, \quad (230)$$

where  $c^*$  is the positive root of  $\Phi(x)$  for  $m < 0$  (or  $c^* = 0$  for  $m = 0$ ). Remarkably, we recover Zipf's law exactly for the zero mean mark case  $m = 0$ .

#### Derivation

From Eq. (23), the steady-state ME is given by

$$\sum_{k=1}^K \frac{1}{\tau_k \lambda_0} \frac{\partial}{\partial z_k} \left( z_k e^{-\sum_{k'=1}^K z_{k'} \phi(z)} \right) + \int dy \rho(y) \phi(z - y \tilde{h}) - \phi(z) = 0, \quad (231)$$

where we have defined  $\phi(z) := G(z)P_{ss}(z)$ . For large  $z$ , the first term in the r.h.s. is negligible due to the exponential factor, leading to

$$\int dy \rho(y) \phi(z - y \tilde{h}) - \phi(z) \simeq 0 \quad \text{for large } z. \quad (232)$$

We then apply the transformation from  $\mathbf{z} = (z_1, \dots, z_K)$  to  $\mathbf{Z} := (W, Z_2, \dots, Z_K)$ :

$$z_1 = \tilde{h}_1 W, \quad z_2 = \tilde{h}_2 W + Z_2, \quad z_3 = \tilde{h}_3 W + Z_3, \quad \dots, \quad z_K = \tilde{h}_K W + Z_K. \quad (233)$$

Using this variable set, we can rewrite

$$\psi(W - y; Z_2, \dots, Z_K) := \phi(\mathbf{z} - y\tilde{\mathbf{h}}) = \phi\left(\tilde{h}_1(W - y), \tilde{h}_2(W - y) + Z_2, \dots, \tilde{h}_K(W - y) + Z_K\right). \quad (234)$$

The integral equation (232) is then reduced to

$$\int dy \rho(y) \psi(W - y; \mathbf{Z}') - \psi(W; \mathbf{Z}') \simeq 0 \quad \text{for large } \mathbf{Z} \quad (235)$$

with  $\mathbf{Z}' := (Z_2, \dots, Z_K)$ . This variable transformation is useful because Eq. (232) is an effectively one-dimensional integral equation. Since the variable subset  $Z'$  is irrelevant in this integral equation, its solution is given by

$$\psi(W; \mathbf{Z}') = C_0(\mathbf{Z}') e^{-c^* W} + C_1(\mathbf{Z}') \quad (236)$$

with arbitrary nonnegative functions  $C_0(\mathbf{Z}')$  and  $C_1(\mathbf{Z}')$  without the variable  $W$  (see Appendix E). In addition, by defining the moment-generating function  $\Phi(x) := \int_{-\infty}^{\infty} dy \rho(y)(e^{xy} - 1)$ , the constant  $c^*$  is the positive root of  $\Phi(c^*) = 0$  for the case of negative mean mark  $m < 0$  or  $c^* = 0$  for the case of zero mean mark  $m = 0$  (see Appendix F for the detailed properties of  $\Phi(x)$ ). Assuming the natural boundary condition,  $C_1(\mathbf{Z}')$  must be set zero as shown later. We then derive the steady distribution  $P_{ss}(\nu)$  as

$$\begin{aligned} P_{ss}(\nu) &:= \int_{-\infty}^{\infty} d\mathbf{z} P_{ss}(\mathbf{z}) \delta\left(\nu - \sum_{k=1}^K z_k\right) \\ &\simeq \frac{1}{\lambda_0} \int_{-\infty}^{\infty} d\mathbf{z} e^{-\beta \sum_{k=1}^K z_k} C_0(\mathbf{Z}') e^{-c^* W} \delta\left(\nu - \sum_{k=1}^K z_k\right) \\ &= \frac{e^{-\beta\nu}}{\lambda_0} \int_{-\infty}^{\infty} dz_1 \int_{-\infty}^{\infty} \left( \prod_{j=2}^K dz_j \right) C_0\left(z_2 - \frac{\tilde{h}_2}{\tilde{h}_1} z_1, \dots, z_K - \frac{\tilde{h}_K}{\tilde{h}_1} z_1\right) \exp\left(-\frac{c^*}{\tilde{h}_1} z_1\right) \delta\left(\nu - \sum_{k=1}^K z_k\right) \end{aligned} \quad (237)$$

Applying the transformation

$$z'_j := z_j - \frac{\tilde{h}_j}{\tilde{h}_1} z_1, \quad \text{for } j = 2, \dots, K, \quad (238)$$

we obtain

$$\begin{aligned} P_{ss}(\nu) &\simeq \frac{e^{-\beta\nu}}{\lambda_0} \int_{-\infty}^{\infty} dz_1 \int_{-\infty}^{\infty} \left( \prod_{j=2}^K dz'_j \right) C_0(z'_2, \dots, z'_K) \exp\left(-\frac{c^*}{\tilde{h}_1} z_1\right) \delta\left(\nu - r z_1 - \sum_{k=2}^K z'_k\right) \\ &= \frac{e^{-\beta\nu}}{\lambda_0} \int_{-\infty}^{\infty} \left( \prod_{j=2}^K dz'_j \right) C_0(z'_2, \dots, z'_K) \int_{-\infty}^{\infty} dz_1 \exp\left(-\frac{c^*}{\tilde{h}_1} z_1\right) \delta\left(\nu - r z_1 - \sum_{k=2}^K z'_k\right) \\ &= \frac{e^{-(\beta-u)\nu}}{r\lambda_0} \int_{-\infty}^{\infty} \left( \prod_{j=2}^K dz'_j \right) C_0(z'_2, \dots, z'_K) \exp\left(\frac{c^*}{h(0)} \sum_{k=2}^K z'_k\right), \end{aligned} \quad (239)$$

where we have used

$$\delta\left(\nu - r z_1 - \sum_{k=2}^K z'_k\right) = \frac{1}{r} \delta\left(z_1 - \frac{\nu - \sum_{k=2}^K z'_k}{r}\right) \quad (240)$$

with

$$r := \frac{1}{\tilde{h}_1} \sum_{k=1}^K \tilde{h}_k = \frac{h(0)}{\tilde{h}_1}, \quad u := \frac{c^*}{h(0)}. \quad (241)$$

Assuming that

$$\frac{1}{r} \int_{-\infty}^{\infty} \left( \prod_{j=2}^K dz'_j \right) C_0(z'_2, \dots, z'_K) \exp \left( \frac{c^*}{h(0)} \sum_{k=2}^K z'_k \right) < \infty, \quad (242)$$

we find that the asymptotic PDF for large  $\nu$  is given by

$$P_{ss}(\nu) \propto e^{-(\beta-u)\nu} \text{ for large } \nu. \quad (243)$$

This asymptotic form implies the power law (230) for the intensity  $\lambda := g(\nu)$ .

#### Natural boundary condition

By neglecting the first term in the ME (23) and by applying the variable transformation (233), we obtain an approximate ME

$$\frac{\partial P_t(W; \mathbf{Z}')}{\partial t} \simeq \int_{-\infty}^{\infty} dy \{ G(W-y; \mathbf{Z}') P_t(W-y; \mathbf{Z}') - G(W; \mathbf{Z}') P_t(W; \mathbf{Z}') \}. \quad (244)$$

By applying the KM expansion, we obtain the conservation of probability:

$$\frac{\partial P_t(W; \mathbf{Z}')}{\partial t} \simeq - \frac{\partial J_t(W; \mathbf{Z}')}{\partial W} \quad (245)$$

with the probability current

$$J_t(W; \mathbf{Z}') := \sum_{n=1}^{\infty} \frac{(-1)^{n-1} \alpha_n}{n!} \frac{\partial^{n-1}}{\partial W^{n-1}} G(W; \mathbf{Z}') P_t(W; \mathbf{Z}'). \quad (246)$$

By substituting the solution (236), we obtain

$$J_{ss}(W; \mathbf{Z}') = \sum_{n=1}^{\infty} \frac{(-1)^{n-1} \alpha_n}{n!} \frac{\partial^{n-1}}{\partial W^{n-1}} \left( C_1(\mathbf{Z}') + e^{-c^* W} C_0(\mathbf{Z}') \right) = m C_1(\mathbf{Z}') + \frac{C_0(\mathbf{Z}')}{c^*} \Phi(c^*) e^{-c^* W}, \quad (247)$$

where we have used  $\Phi(x) = \sum_{n=1}^{\infty} (\alpha_n/n!) x^n$ . Since  $\Phi(c^*) = 0$  by definition, we obtain  $J_{ss}(W; \mathbf{Z}') = m C_1(\mathbf{Z}')$ . The natural boundary condition requires

$$\lim_{W \rightarrow \infty} J_{ss}(W; \mathbf{Z}') = 0 \quad (248)$$

for any  $\mathbf{Z}'$ , implying  $C_1(\mathbf{Z}') = 0$ .

## B. General memory kernel

As done before, any memory kernel can be approximated by a sum of exponentials, such that

$$h(t) = \int_0^{\infty} dx \tilde{h}(x) e^{-t/x} \approx \sum_{k=1}^K \tilde{h}_k e^{-t/\tau_k}. \quad (249)$$

Since the power law tail for the PDF of the intensity is found for any discrete sum of exponentials, it remains valid for general superpositions of exponential memory. Under the assumption

$$\forall h(t), \quad g(\nu) = \lambda_0 e^{\beta\nu}, \quad p_+ := \int_0^{\infty} \rho(y) dy > 0, \quad p_- := \int_{-\infty}^0 \rho(y) dy > 0, \quad m := \int_{-\infty}^{\infty} y \rho(y) dy \leq 0. \quad (250)$$

we obtain

$$P_{ss}(\lambda) \propto \lambda^{-2-\beta^{-1}u} \text{ for large } \lambda, \quad u := \frac{c^*}{h(0)}, \quad (251)$$

where  $c^*$  is the positive root of  $\Phi(c^*) = 0$  for  $m < 0$  or  $c^* = 0$  for  $m = 0$ . For the zero mean mark case  $m = 0$ , the PDF obeys Zipf's law exactly.

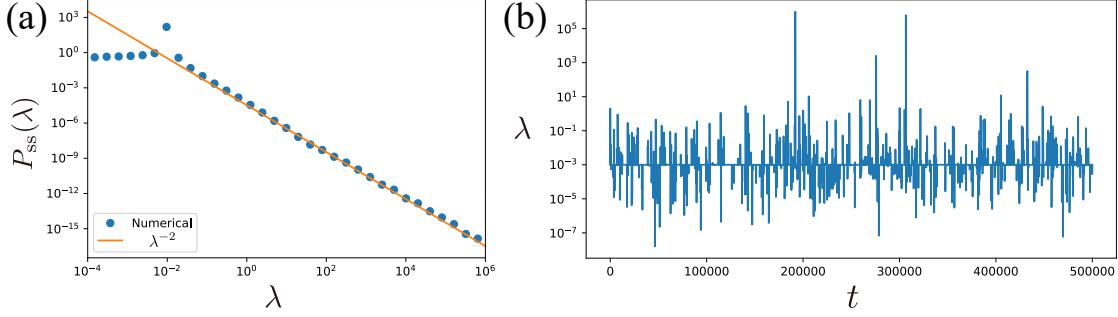


FIG. 10: (a) Numerical confirmation of our theoretical prediction (230) of the Zipf law for the intensity distribution. This simulation corresponds to the following set of parameters:  $K = 3$ ,  $(\tau_1, \tau_2, \tau_3) = (1, 0.5, 2)$ ,  $(\tilde{h}_1, \tilde{h}_2, \tilde{h}_3) = (0.5, 0.6, 0.1)$ ,  $\lambda_0 = 0.01$ ,  $\beta = 6$ ,  $\lambda_{\max} = 10^6$ , and  $\sigma = 0.3$ . The predicted power law exponent is given by  $a = 1.0$  (i.e., a true power law; the Zipf law) and is shown as the red straight line. (b) Sample trajectory of the intensity for  $K = 3$ ,  $(\tau_1, \tau_2, \tau_3) = (1, 0.5, 2)$ ,  $(\tilde{h}_1, \tilde{h}_2, \tilde{h}_3) = (0.5, 0.6, 0.1)$ ,  $\lambda_0 = 0.001$ ,  $\beta = 10$ ,  $\lambda_{\max} = 10^6$ , and  $\sigma = 0.3$ . Note that the horizontal reference line is determined by  $\lambda_0 = 0.001$ .

### C. Numerical confirmation

Figure 10 shows numerical results obtained by Monte Carlo simulations of the SDE (3) for the memory function made of a discrete sum of exponential functions, for the exponential intensity MSA with finite cutoff (to ensure convergence of the numerical scheme), and a zero-mean Gaussian mark distribution

$$h(t) = \sum_{k=1}^K \tilde{h}_k e^{-t/\tau_k}, \quad \lambda = g(\nu) = \min \{ \lambda_0 e^{\beta \nu}, \lambda_{\max} \}, \quad \rho(y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{y^2}{2\sigma^2}}. \quad (252)$$

We use an adaptive time discretisation scheme (see Appendix I for the detailed numerical scheme). Here  $\lambda_{\max}$  is a cutoff parameter to control numerical rounding errors. Figure 10a exhibits the Zipf law in the intensity distribution up to the cutoff  $\lambda_{\max}$ , showing agreement with our theoretical prediction (230).

Figure 10b shows a sample trajectory of the intensity obtained for the parameter set  $K = 3$ ,  $(\tau_1, \tau_2, \tau_3) = (1, 0.5, 2)$ ,  $(\tilde{h}_1, \tilde{h}_2, \tilde{h}_3) = (0.5, 0.6, 0.1)$ ,  $\lambda_0 = 0.001$ ,  $\beta = 10$ ,  $\lambda_{\max} = 10^6$ , and  $\sigma = 0.3$ . This semi-log plot illustrates that the NLHawkes model exhibits an intermittent behaviour in terms of its intensity, which is qualitatively consistent with observed phenomena in various complex systems, such as seismic activity.

### D. Generalisation to fast-accelerating intensity maps

The above framework can be readily generalised to FAI maps defined by  $g(\nu) \gg \nu^2$ . Our general result can be formulated as follows. Under the assumptions

$$h(t) = \sum_{k=1}^K \tilde{h}_k e^{-t/\tau_k}, \quad g(\nu) \gg \nu^2 \text{ for large } \nu, \quad p_+ := \int_0^\infty \rho(y) dy > 0, \quad p_- := \int_{-\infty}^0 \rho(y) dy > 0, \quad m := \int_{-\infty}^\infty y \rho(y) dy \leq 0, \quad (253a)$$

we obtain

$$P_{ss}(\lambda) \propto \lambda^{-1} \left[ e^{-u\nu} \left\{ \frac{dg(\nu)}{d\nu} \right\}^{-1} \right]_{\nu=g^{-1}(\lambda)} \quad \text{for large } \lambda, \quad u := \frac{c^*}{h(0)} \quad (253b)$$

with  $c^*$  being the positive root of  $\Phi(c^*) = 0$  for  $m < 0$  or zero for  $m = 0$ , where  $\Phi(x) := \int_{-\infty}^\infty dy \rho(y)(e^{xy} - 1)$ . The derivation of this result is essentially the same as that in Sec. VIII A, by replacing the MSA intensity map with the general FAI map.

Since any memory kernel can be approximated by a discrete sum of exponentials (see Eq. (249)), the continuous

version of the statement (253) also holds:

$$\forall h(t), \quad g(\nu) \gg \nu^2, \quad p_+ > 0, \quad p_- > 0, \quad m \leq 0 \implies P_{ss}(\lambda) \propto \lambda^{-1} \left[ e^{-u\nu} \left\{ \frac{dg(\nu)}{d\nu} \right\}^{-1} \right]_{\nu=g^{-1}(\lambda)}, \quad u := \frac{c^*}{h(0)} \quad (254)$$

with  $c^*$  being the positive root of  $\Phi(c^*) = 0$  for  $m < 0$  or zero for  $m = 0$ , where  $\Phi(x) := \int_{-\infty}^{\infty} dy \rho(y)(e^{xy} - 1)$ . Remarkably, this result implies that the power law tail for the steady-state PDF of intensities holds robustly for super-polynomial intensity maps, such as  $g(\nu) \propto e^{\beta\nu}$  (i.e., the MSA case) and  $g(\nu) \propto e^{\beta\nu^2}$ .

## IX. DISCUSSION

### A. Relationship to nonlinear Kesten processes

We have shown that power law asymptotics robustly appears for the quadratic and FAI cases. Here we provide another derivation based on more heuristic arguments, by removing inessential technicalities, using the viewpoint of Kesten processes [51]. Let us focus on the case with exponential memory  $h(t) = (\eta/\tau)e^{-t/\tau}$  and in the diffusive limit (i.e., for the symmetric mark distribution  $\rho(y) = \rho(-y)$ ) described by the FPE (46) and the corresponding SDE (47).

#### 1. Case with quadratic intensity map: $\tilde{g}(\nu) = k\nu^2 + \nu_0$

Inserting  $\tilde{g}(\nu) = k\nu^2 + \nu_0$  in the SDE (47), we obtain

$$\frac{d\hat{\nu}}{dt} = -\frac{\hat{\nu}}{\tau} + \sqrt{k\nu^2 + \nu_0} \sqrt{2D} \cdot \hat{\xi}_t^G. \quad (255)$$

Here,  $\hat{\xi}_t^G$  is the standard white Gaussian noise satisfying  $\langle \hat{\xi}_t^G \rangle = 0$  and  $\langle \hat{\xi}_t^G(t) \hat{\xi}_{t'}^G(t') \rangle = \delta(t - t')$ . With respect to the power law structure of the tail of the PDF of  $\nu$ , by using the discretisation  $d\nu/dt = (\nu(t + dt) - \nu(t))/dt$ , this SDE can be regarded as a continuous version of the discrete-time Kesten process [51]

$$\nu(t + dt) = a_t \nu(t) + b_t, \quad \text{with } a_t = 1 + (2Dk)^{1/2} \hat{\xi}_t^G dt - (dt/\tau) \quad \text{and} \quad b_t = \nu_0 (2k)^{1/2} \hat{\xi}_t^G dt. \quad (256)$$

The first term  $a_t \nu(t)$  controls the intermittent excursions of  $\nu(t)$  to large values, for which  $\nu_0$  can be neglected in the last term of the r.h.s. of Eq. (255). The second term  $b_t$  in the Kesten map (256) is the ‘‘reinjecting term’’ obtained when  $\nu$  becomes smaller than  $\nu_0$ . As shown in Ref. [53], the detailed shape of this ‘‘reinjecting term’’ has no impact on the existence of a power law tail and on the value of its exponent. The only important point is that the ‘‘reinjecting term’’ exists to prevent  $\nu$  from being too small. Remaining no less than a stochastic variable proportional to  $\nu_0$ , intermittent runs of exponential growth occur when there is a succession of positive realisations of  $\hat{\xi}_t^G$  for several consecutive times such that the multiplicative factor  $a_t$  is larger than 1 over this run [52, 53].

The condition for the existence of a steady-state PDF for the Kesten process (256) is that  $\langle \ln a_t \rangle < 0$  [51, 53]. For infinitesimal  $dt$ ,  $\ln a_t$  can be expanded as  $\ln a_t = (2Dk)^{1/2} \hat{\xi}_t^G dt - (dt/\tau)$  and its mean is then  $\langle \ln a_t \rangle = -(dt/\tau)$  since  $\langle \hat{\xi}_t^G \rangle = 0$  by definition. Hence, the condition for a stationarity process holds true. It is then easy to show by explicitly writing the self-consistent equation for the steady-state PDF of  $\nu$  that it is a power law with exponent  $a$  given as the solution of the equation

$$\langle |a_t|^a \rangle = 1. \quad (257)$$

Using the fact that  $\hat{\xi}_t^G dt = dW$  is a Gaussian random variable with zero mean and variance  $dt$  (i.e. it is the infinitesimal increment of the Wiener process), the average in (257) is obtained by using the saddle-node approximation, and we find that the corresponding solution recovers exactly expression (94), namely  $a := 1/2 + c^2/(2k\tau)$  with  $c := \tau/\eta$ . This confirms that our treatment in terms of the diffusive limit gives equations in the general class of Kesten processes. Other forms of the tension-intensity map  $\lambda = g(\nu)$  can thus be interpreted as continuous nonlinear extension of the Kesten process.

Therefore, from an intuitive point of view, the power law tails of the PDFs of  $\nu$  and  $\lambda$  can be qualitatively related to an underlying multiplicative structure together with additional ingredients to ensure the existence and stationarity of the process.

## 2. Case with fast-accelerating intensity map: $\tilde{g}(\nu) \gg \nu^2$

We here consider the case of FAI maps satisfying

$$\tilde{g}(\nu) \gg \nu^2 \text{ (for positive large } \nu\text{), } \tilde{g}(\nu) \simeq \text{const. (for negative large } \nu\text{).} \quad (258)$$

Let us take  $D = 1/2$  to simplify notations, so that the corresponding SDE is given by

$$\frac{d\hat{\nu}}{dt} = -\frac{\hat{\nu}}{\tau} + \sqrt{\tilde{g}(\hat{\nu})} \cdot \hat{\xi}^G. \quad (259)$$

Since  $\tilde{g}(\nu) \simeq \text{const.}$  for negative large  $\nu$ ,  $\hat{\nu}$  cannot go to  $-\infty$  due to the relaxation term  $-\hat{\nu}/\tau$ . On the other hand, for positive large  $\hat{\nu}$ , the dynamics is approximated by

$$\frac{d\hat{\nu}}{dt} \simeq \sqrt{\tilde{g}(\hat{\nu})} \cdot \hat{\xi}^G, \quad (260)$$

because  $\sqrt{\tilde{g}(\hat{\nu})} \gg \hat{\nu}/\tau$  for FAI maps.

This model is thus similar to a Brownian motion with a position dependent variance or diffusion coefficient. Interestingly, such a Brownian model has a well-defined steady-state PDF for FAI maps. Let us thus consider a Brownian motion obeying the following SDE

$$\frac{d\hat{\nu}}{dt} = \sqrt{\tilde{g}(\hat{\nu})} \cdot \hat{\xi}^G \text{ (for } \nu \geq 0\text{)}, \quad (261)$$

which is complemented by the condition of a repulsive hard wall at  $\hat{\nu} = 0$ , which prohibits the ‘‘Brownian particle’’ from going to  $-\infty$ . This condition is a simplification to and ensures a similar result as in the presence of the relaxation effect in the original model (259). The steady FPE is given by

$$\frac{d^2}{d\nu^2} \{\tilde{g}(\nu) P_{ss}(\nu)\} = 0 \text{ (for } \nu > 0\text{), } P_{ss}(\nu) = 0 \text{ (for } \nu < 0\text{).} \quad (262)$$

If  $\tilde{g}(\nu)$  were not a FAI map, this steady FPE might not have a normalisable steady solution. For example, if  $\tilde{g}(\nu) = \text{const.}$ , the general solution of the FPE is given by a non-normalisable steady solution  $\tilde{P}_{ss}(\nu) = c_0 + c_1\nu$ , satisfying  $\int_0^\infty d\nu P_{ss}(\nu) = \infty$ . This model is therefore non-stationary. In contrast, when  $\tilde{g}(\nu)$  is a FAI map, the FPE (262) has a normalisable steady solution for any FAI map, with

$$P_{ss}(\nu) \propto \begin{cases} \frac{1}{\tilde{g}(\nu)} & \text{(for } \nu > 0\text{)} \\ 0 & \text{(for } \nu < 0\text{)} \end{cases} \implies \int_{-\infty}^\infty d\nu P_{ss}(\nu) < \infty. \quad (263)$$

We then obtain the robust expression of the steady-state PDF of the intensity  $\lambda := \tilde{g}(\nu)$ :

$$P_{ss}(\lambda) \propto \lambda^{-1} \left| \frac{d\tilde{g}(\nu)}{d\nu} \right|_{\nu=\tilde{g}^{-1}(\lambda)}^{-1} \text{ (for large } \lambda\text{).} \quad (264)$$

This result readily implies that the Zipf law

$$P_{ss}(\lambda) \propto \lambda^{-2} \text{ (for large } \lambda\text{)} \quad (265)$$

is observed for a wide class of superpolynomial intensity maps, such as  $\tilde{g}(\nu) = e^{\beta\nu}$  and  $\tilde{g}(\nu) = e^{\beta\nu^2}$ .

Let us complete this discussion by mentioning that the rigorous mathematical demonstration of the existence of steady-state solutions of the SDE (261) is obtained from the theorems presented in Ref. [54]. In particular, we refer to the theorems in section 5.2 in Ref. [54].

*a. Numerical simulation.* Figure 11 presents the PDF of the intensity obtained from the numerical solution of the SDE (261) with an exponential intensity  $\tilde{g}(\nu) = \lambda_0 e^{\beta\nu}$ . The Zipf law (265) is obtained for the steady-state intensity distribution of the SDE describing a Brownian motion with  $\hat{\nu}$ -dependent diffusion constant  $\tilde{g}(\nu)/2$  in the presence of a reflecting barrier at  $\hat{\nu} = 0$ . The detailed numerical implementation is described in Appendix I.

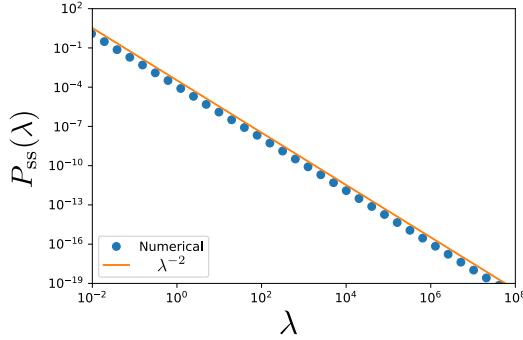


FIG. 11: Numerical simulations confirming our theoretical prediction on the Zipf law (265) for the intensity distribution. The simulated SDE describes a Brownian motion with the  $\hat{\nu}$ -dependent diffusion constant  $\tilde{g}(\hat{\nu})/2$  and a reflecting barrier at  $\hat{\nu} = 0$ .

*b. Intuitive discussion.* Why is the Brownian model with position dependent variance stationary for FAI maps? This might be understandable from the viewpoint of step-size explosion for large  $\hat{\nu}$ . As  $\hat{\nu}$  grows to very large values, the steps of the random walk explode even faster and thus it is very likely that a negative step occurs of huge size which brings back the Brownian particle to 0 or even pushes it to negative values if the repelling boundary was absent. In the presence of the repelling boundary, the huge negative steps bring the Brownian particle close to 0, for which the random step sizes become small, which implies that  $\hat{\nu}$  remains quite a long time in the vicinity of the origin. Eventually,  $\hat{\nu}$  escapes again to large values of  $\hat{\nu}$  but then the huge random walk step sizes, when negative, bring it back again to a neighbourhood of 0. This process occurs repeatedly and leads to a stationary PDF decaying rapidly as  $1/\tilde{g}(\hat{\nu})$ , due to the effect of the negative random steps that push back  $\hat{\nu}$  to the left and the boundary somehow traps the process in its neighbourhood. In summary, this picture implies that the explosive step size leads to an effective strong “trapping potential”, which might be counter-intuitive at the first glance.

It might be interesting to rephrase the above argument from the viewpoint of the recurrence time of one-dimensional Brownian motion. It is well-known that the recurrence probability of one-dimensional Brownian motion with constant variance  $\tilde{g}(\nu) = \text{const.}$  is unity, while the average recurrence time is infinity. In other words, a Brownian particle will surely come back to the origin after a long time, but this waiting time may be too large so that repeated recurrence events cannot be expected within a finite observation time. In the case of Eq. (261), the variance of the step lengths depends on the distance from the origin and becomes larger and larger for large  $\nu$ . Thus, the time evolution of this Brownian particle becomes faster and faster for large  $\hat{\nu}$ , such that the particle can come back to the origin much sooner than with a constant diffusion coefficient and repeated recurrence events can be expected in a finite time.

Last, it is useful to stress the difference between the mechanism underlying the existence of a steady-state power law distribution in the conventional linear Kesten process [51] and that of our FAI Hawkes model. In usual linear Kesten processes, the underlying mechanism is “proportional growth” or multiplicative proportional growth in order for Zipf’s law and related power laws to occur, in the presence of an average contraction, i.e., the mean growth rate is negative, together with a reinjection mechanism. In contrast, our FAI Hawkes models are characterised by explosive expansions in the presence of a reflecting or bounded boundary condition. This is in stark contrast to the conventional approaches based on the Kesten type models, proportional growth type, and preferential attachment type mechanisms.

## B. Implication to financial data analyses

From a broader perspective, our results have significant implications for financial modelling. Recall that one of the motivations for introducing the NLHawkes family is to reproduce empirical stylised facts, in particular the power law distribution of returns (see Sec. II C 2 for a brief review). Indeed, according to Ref. [28], one of the advantages of the QHawkes process lies in the fact that it can reproduce a power law intensity distribution with non universal exponents (see Eq. (12)), from which the power law distribution of price changes derives.

From this point of view, our results summarised in Table I show that various NLHawkes processes can reproduce power law intensity PDFs, and not just the QHawkes processes. Even the ramp Hawkes processes with one-sided marks (which is arguably a minor modification of the LHawkes process) can reproduce a power law distribution with arbitrary exponent, when near criticality. If one focuses only on power law intensity PDFs, various Hawkes models can be suitable candidate models. Moreover, by assuming symmetric marks, note that QHawkes is the model at the boundary between a class of models with universal power law exponents and a class of models with non-universal

exponents.

Another debatable point is whether the exponent should be universal or non-universal. The linear and ramp Hawkes processes with one-sided marks, the symmetric-mark QHawkes processes, and the two-sided FAI Hawkes process with nonpositive mean mark have non-universal exponents for the power law intensity PDF. In contrast, the symmetric-mark FAI Hawkes processes have universal exponents. The non-universal exponents are useful for flexible data calibration, while one would like to have strong justifications of why such parameters are selected in empirical data analyses. On the other hand, universal exponents are useful if the empirical exponent seems robust and universal, but this removes flexible data calibration. It might be necessary to construct a suitable framework for model selection with the goal of developing practical reverse engineering approaches. This is beyond the scope of the present paper.

### C. Mathematical relation to quantum field theories

We have studied a wide variety of generalised Hawkes processes. Here we discuss their mathematical relation with quantum field theories. Let us rewrite  $z(x) \rightarrow \phi(x)$  and introduce a “momentum” operator  $\pi(x)$

$$\pi(x) := -i \frac{\delta}{\delta \phi(x)}, \quad (266)$$

which satisfies the canonical commutative relation

$$[\phi(x), \pi(x')] = i\delta(x - x'). \quad (267)$$

By introducing the state vector

$$|P_t\rangle := \int \mathcal{D}\phi P_t[\phi] |\phi\rangle, \quad (268)$$

the ME (17) becomes a Schrödinger-like equation for the field  $\{\phi(x)\}_x$  as

$$\frac{\partial}{\partial t} |P_t\rangle = H |P_t\rangle, \quad H := i \int_0^\infty \frac{dx}{x} \pi(x) \phi(x) + \int_{-\infty}^\infty dy \rho(y) \left( T[y \tilde{h}] - 1 \right) G[\phi] \quad (269)$$

with the non-Hermitian Hamiltonian  $H$  and the translation operator  $T[y]$ , defined by

$$T[y] := \exp \left[ -i \int_0^\infty dx y(x) \pi(x) \right] \quad (270)$$

satisfying  $T[y]P_t[z] = P_t[z - y]$ . This equation is non-local, since the Hamiltonian includes infinite-order of “momentum” operators.

In the diffusive limit (48), the Hamiltonian reduces to local version as a result of the SSE:

$$\frac{\partial}{\partial t} |P_t\rangle = H |P_t\rangle, \quad H := i \int_0^\infty \frac{dx}{x} \pi(x) \phi(x) - \int_0^\infty dx \int_0^\infty dx' D(x, x') \pi(x) \pi(x') G[\phi], \quad D(x, x') := \frac{\alpha_2}{2} \tilde{h}(x) \tilde{h}(x'), \quad (271)$$

where the “momentum” operator appears via a quadratic form. In this sense, the SSE for the field ME can be regarded as a mathematical procedure to obtain local forms of non-Hermitian field quantum theories in an appropriate limit.

### D. Future application: field master equation for the general QHawkes processes

Our current formulation covers a part of the ZHawkes processes, but not yet the whole class of general QHawkes processes (10). Including the general QHawkes processes as special cases of our formalism can in principle be easily performed as follows. The linear and quadratic kernels  $L(t - s)$  and  $K(t - s, t - u)$  can be decomposed according to a Laplace representation:

$$L(t - s) = \int_0^\infty dx e^{-\frac{t-s}{x}} \tilde{L}(x), \quad K(t - s, t - u) = \int_0^\infty dx \int_0^\infty dx' e^{-\frac{t-s}{x} - \frac{t-u}{x'}} \tilde{K}(x, x'). \quad (272)$$

The QHawkes process then reads

$$\begin{aligned}
\hat{\lambda}(t) &= \lambda_0 + \int_{-\infty}^t ds \hat{\xi}_{\rho(y); \hat{\lambda}(s)}^P(s) \left( \int_0^\infty dx e^{-\frac{t-s}{x}} \tilde{L}(x) \right) \\
&\quad + \int_{-\infty}^t ds \hat{\xi}_{\rho(y); \hat{\lambda}(s)}^P(s) \int_{-\infty}^t du \hat{\xi}_{\rho(y); \hat{\lambda}(u)}^P(u) \left( \int_0^\infty dx \int_0^\infty dx' e^{-\frac{t-u}{x'}} e^{-\frac{t-s}{x}} \tilde{K}(x, x') \right) \\
&= \lambda_0 + \int_0^\infty dx \tilde{L}(x) \left( \int_{-\infty}^t ds e^{-\frac{t-s}{x}} \hat{\xi}_{\rho(y); \hat{\lambda}(s)}^P(s) \right) \\
&\quad + \int_0^\infty dx \int_0^\infty dx' \tilde{K}(x, x') \left( \int_{-\infty}^t ds e^{-\frac{t-s}{x}} \hat{\xi}_{\rho(y); \hat{\lambda}(s)}^P(s) \right) \left( \int_{-\infty}^t du e^{-\frac{t-u}{x'}} \hat{\xi}_{\rho(y); \hat{\lambda}(u)}^P(u) \right), \tag{273}
\end{aligned}$$

where we have exchanged the order of integration. By considering the Markovian SPDE

$$\frac{\partial \hat{z}(t, x)}{\partial t} = -\frac{\hat{z}(t, x)}{x} + \hat{\xi}_{\rho(y); \hat{\lambda}(t)}^P(t), \tag{274a}$$

whose explicit solution is given by

$$\hat{z}(t, x) = \int_{-\infty}^t ds e^{-\frac{t-s}{x}} \hat{\xi}_{\rho(y); \hat{\lambda}(s)}^P(s), \tag{274b}$$

we obtain

$$\hat{\lambda}(t) = \lambda_0 + \int_0^\infty dx \tilde{L}(x) \hat{z}(t, x) + \int_0^\infty dx \int_0^\infty dx' \tilde{K}(x, x') \hat{z}(t, x) \hat{z}(t, x'), \tag{274c}$$

which is equivalent to the original QHawkes processes (10). This means that the QHawkes dynamics has been converted into a Markovian dynamics described by the set of equations (274) in terms of the field variable  $\{\hat{z}(t, x)\}_x$ . Correspondingly, we obtain the field ME for the QHawkes processes as

$$\frac{\partial P_t[z]}{\partial t} = \int_0^\infty dx \frac{\delta}{\delta z(x)} \left( \frac{z(x)}{x} P_t[z] \right) + \int_{-\infty}^\infty dy \rho(y) \{G[z - y] P_t[z - y] - G[y] P_t[z]\} \tag{275}$$

with the indicator function  $1(x) = 1$  for any  $x \in (0, \infty)$  and the functional intensity map  $G[z]$  defined by

$$G[z] := \lambda_0 + \int_0^\infty dx \tilde{L}(x) z(x) + \int_0^\infty dx \int_0^\infty dx' \tilde{K}(x, x') z(x) z(x'). \tag{276}$$

#### *Further formal generalisation*

Obviously, this method can be readily generalised for any functional series expansion, at least formally, such as

$$\begin{aligned}
\hat{\lambda}(t) &= \lambda_0 + \sum_{j=1}^J \int_{-\infty}^t ds_1 \cdots \int_{-\infty}^t ds_j K_j(t - s_1, \dots, t - s_j) \hat{\xi}_{\rho(y); \hat{\lambda}(s_1)}^P \cdots \hat{\xi}_{\rho(y); \hat{\lambda}(s_j)}^P \\
&= \lambda_0 + \sum_{j=1}^J \int_0^\infty ds_1 \cdots \int_0^\infty ds_j \tilde{K}_j(x_1, \dots, x_j) \hat{z}(t, x_1) \cdots \hat{z}(t, x_j). \tag{277}
\end{aligned}$$

Here, we have introduced the field variable  $\{\hat{z}(t, x)\}$  obeying the Markovian SPDE

$$\frac{\partial \hat{z}(t, x)}{\partial t} = -\frac{\hat{z}(t, x)}{x} + \hat{\xi}_{\rho(y); \hat{\lambda}(t)}^P, \tag{278}$$

and the Laplace decomposition

$$K_j(t - s_1, \dots, t - s_j) = \int_0^\infty dx_1 \cdots \int_0^\infty dx_j e^{-\sum_{l=1}^j \frac{t-s_l}{x_l}} \tilde{K}_j(x_1, \dots, x_j). \tag{279}$$

The corresponding field ME can be derived in the same manner. This implies that our formulation has the potential to cover a wide variety of NLHawkes families beyond the QHawkes processes. We leave to future studies the derivation of explicit analytical solutions for general QHawkes processes and beyond, based on our formulation.

## X. CONCLUSION

In this article, we have studied various analytical solutions to NLHawkes processes by generalising the field ME approach recently developed in Refs. [16, 17]. We have derived the field ME for the general NLHawkes processes and have formulated its functional KM expansion and the corresponding diffusive approximation. We then proceeded with deriving various exact solutions of the steady-state intensity distributions for an exponential memory kernel in the absence and presence of inhibitory effects. Some of the robust asymptotic solutions have been generalised for a wide class of memory kernels, such as (i) the non-universal power law with an arbitrary exponent for the ramp Hawkes process in the absence of inhibitory effects, (ii) the robust Zipf law for the superexponential intensity family in the presence of symmetric inhibitory and excitatory effects, and (iii) the ubiquitous power law for the fast-acceleration intensity (FAI) Hawkes models in the regime of zero or negative mean mark.

The summary table I exemplifies our systematic analysis of the NLHawkes processes. However, there are two missing items in the last column of the table for general memory kernels  $h(t)$ . This is because our focus has been mainly on FAI Hawkes processes and the ramp Hawkes process with non-positive mean mark and the QHawkes processes with symmetric mark are not FAI Hawkes processes. It is likely that different perturbative solutions are needed to solve these two cases for general memory kernels  $h(t)$ , which we leave for a future work.

While only a few analytical solution for limited cases have been derived in the past for NLHawkes processes due to their nonlinear and non-Markovian nature, we have significantly extended the set of solutions, obtaining exact and robust asymptotic expressions with the help of our formulation in terms of a field ME. This demonstrates the power of this approach in addressing non-Markovian stochastic processes. It would be interesting to generalise this framework for more general non-Markovian stochastic processes, such as non-Markovian point processes that have arbitrary intensities depending on the full past history. In addition, our results imply that the NLHawkes family can accommodate various power law relations in the intensity distribution, which could be useful for data calibration in various complex systems.

### Acknowledgments

This work was supported by (i) JST, PRESTO Grant Number JPMJPR20M2, Japan, (ii) the Japan Society for the Promotion of Science KAKENHI (Grant No. 20H05526 and No. 22H04830), (iii) Intramural Research Promotion Program in the University of Tsukuba, and (iv) the National Natural Science Foundation of China under grant No. U2039202. We thank Y. Terada and J.-P. Bouchaud for fruitful discussions.

## Appendix A: Formal properties of the Dirac delta function

### 1. Dirac delta function

The Dirac delta function is formally defined by the following relationships for real numbers  $x, y$ :

$$\delta(x - y) = \begin{cases} 0 & (x \neq y) \\ \infty & (x = y) \end{cases}, \quad \int_{-\infty}^{\infty} f(x)\delta(x - y)dx = f(y), \quad (\text{A1})$$

which is the continuous analogue to the Kronecker delta, defined by

$$\delta_{ij} = \begin{cases} 0 & (i \neq j) \\ 1 & (i = j) \end{cases}, \quad \sum_i f_i \delta_{ij} dx = f_j, \quad (\text{A2})$$

for integer  $i, j$ .

There are several formal methods to construct the Dirac delta function. In this paper, we construct the Dirac delta function via a formal continuous limit from the discrete picture. Let us consider the lattice coordinate  $x_i := idx$  for an integer  $i$  with the lattice interval  $dx$ . The Dirac delta function can be formally introduced by

$$\delta(x_i - x_j) = \lim_{dx \downarrow 0} \frac{1}{dx} \delta_{i,j}, \quad (\text{A3})$$

which satisfies

$$\int_{-\infty}^{\infty} f(x)\delta(x - y)dx := \lim_{dx \downarrow 0} \sum_i f(x_i) \left[ \frac{1}{dx} \delta_{ij} \right] dx = f(y) \quad \text{for } y = x_j. \quad (\text{A4})$$

## 2. Functional derivative

The functional derivative is an analogue to the partial differential, such that

$$\frac{\delta}{\delta z(x)} z(y) = \delta(x - y), \quad (\text{A5})$$

which is similar to  $(\partial/\partial z_i)z_j = \delta_{ij}$ . The functional derivative can be constructed via a formal continuous limit,

$$\frac{\delta}{\delta z(x_k)} [\dots] = \lim_{dx \downarrow 0} \frac{1}{dx} \frac{\partial}{\partial z_k} [\dots]. \quad (\text{A6})$$

Indeed, this definition satisfies the relationship (A5), such that

$$\frac{\delta}{\delta z(x)} z(y) = \lim_{dx \downarrow 0} \frac{1}{dx} \frac{\partial}{\partial z_i} z_j = \lim_{dx \downarrow 0} \frac{1}{dx} \delta_{ij} = \delta(x_i - x_j) \quad (\text{A7})$$

for  $x = x_i$  and  $y = x_j$ .

## 3. Functional Taylor expansion

For a finite-dimensional vector  $\mathbf{z} := (z_1, \dots, z_N)$ , the Taylor expansion is given by

$$f(\mathbf{z}) = \sum_{n=1}^{\infty} \frac{1}{n!} \left( \sum_{i=1}^N z_i \frac{\partial}{\partial x_i} \right)^n f(\mathbf{x}) \Big|_{\mathbf{x}=\mathbf{0}}. \quad (\text{A8})$$

As its continuous analogue, the functional Taylor expansion for a function  $f[z] := f(\{z(x)\}_x)$  reads

$$f[z] = \sum_{n=1}^{\infty} \frac{1}{n!} \left( \int_{-\infty}^{\infty} dx z(x) \frac{\delta}{\delta y(x')} \right)^n f[y] \Big|_{y=0}. \quad (\text{A9})$$

## Appendix B: Another derivation of the field master equation (17)

Here, we provide another derivation of the field ME (17) by direct manipulation of PDFs. Let us consider the time evolution of any functional  $f[\hat{z}] := f(\{\hat{z}(t, x)\}_x)$ , given by

$$\begin{aligned} df(\{\hat{z}(t, x)\}_x) &:= f(\{\hat{z}(t + dt, x)\}_x) - f(\{\hat{z}(t, x)\}_x) \\ &= \begin{cases} -dt \int_0^\infty dx \frac{\hat{z}(t, x)}{x} \frac{\delta f[\hat{z}(t, x)]}{\delta \hat{z}(t, x)} & \text{(No jump during } [t, t + dt]\text{: probability} = 1 - \hat{\lambda}(t)dt) \\ f(\{\hat{z}(t, x) + \hat{y}\tilde{h}(x)\}_x) - f(\{\hat{z}(t, x)\}) & \text{(Jump in } [t, t + dt]\text{ with } \hat{y} \in [y, y + dy]\text{: probability} = \hat{\lambda}(t)\rho(y)dtdy) \end{cases} \end{aligned} \quad (\text{B1})$$

with intensity

$$\hat{\lambda}(t) = G[\hat{z}] := g \left( \int_0^\infty dx \hat{z}(t, x) \right). \quad (\text{B2})$$

By taking the ensemble average on both sides of the equation, we obtain

$$\int \mathcal{D}z f[z] \frac{\partial P_t[z]}{\partial t} dt = \int \mathcal{D}z \left[ -dt \int_0^\infty dx \frac{z(x)}{x} \frac{\delta f[z]}{\delta z(x)} + dt \int dy \rho(y) G[z] \{ f[z + y\tilde{h}] - f[z] \} \right] P_t[z]. \quad (\text{B3})$$

By integration by parts and performing a variable transformation  $z + y\tilde{h} \rightarrow z$ , we obtain an identity

$$\int \mathcal{D}z f[z] \frac{\partial P_t[z]}{\partial t} = \int \mathcal{D}z f[z] \left[ \int dx \frac{\delta}{\delta z} \frac{z}{x} P_t[z] + \int dy \rho(y) G[z - y\tilde{h}] P_t[z - y\tilde{h}] - G[z] P_t[z] \right]. \quad (\text{B4})$$

Since this identity holds for any functional  $f[z]$ , we obtain Eq. (17).

## Appendix C: Integral identities for exponential mark distributions

Here we provide the detailed derivation of the identities (58) and (81).

### 1. For positive contribution

Let us consider the following quantity

$$\mathcal{I}_+(\nu) := \int_0^\infty dy \frac{e^{-y/y^*}}{y^*} f(\nu - \eta y/\tau) \quad (\text{C1})$$

with the boundary condition  $\lim_{\nu \rightarrow \pm\infty} f(\nu) = 0$ . Let us differentiate both hand sides as

$$\frac{d}{d\nu} \mathcal{I}_+(\nu) = \int_0^\infty dy \frac{e^{-y/y^*}}{y^*} \frac{d}{d\nu} f(\nu - \eta y/\tau). \quad (\text{C2})$$

The identity

$$\frac{d}{dy} f(\nu - \eta y/\tau) = -\frac{\eta}{\tau} \frac{d}{d\nu} f(\nu - \eta y/\tau), \quad (\text{C3})$$

leads to

$$\frac{d}{d\nu} \mathcal{I}_+(\nu) = -\frac{\tau}{\eta} \int_0^\infty dy \frac{e^{-y/y^*}}{y^*} \frac{d}{dy} f(\nu - \eta y/\tau) = -\frac{\tau}{\eta} \left[ \frac{e^{-y/y^*}}{y^*} f(\nu - \eta y/\tau) \right]_0^\infty - \frac{\tau}{\eta y^*} \int_0^\infty dy \frac{e^{-y/y^*}}{y^*} f(\nu - \eta y/\tau), \quad (\text{C4})$$

where we have performed an integration by part. This means that

$$\frac{d}{d\nu} \mathcal{I}_+(\nu) = c f(\nu) - c \mathcal{I}_+(\nu), \quad c := \frac{\tau}{\eta y^*}, \quad (\text{C5})$$

which is equivalent to

$$\left( 1 + \frac{1}{c} \frac{d}{d\nu} \right) \mathcal{I}_+(\nu) = f(\nu). \quad (\text{C6})$$

We thus have the identity (58).

### 2. For negative contribution

Let us consider the following quantity

$$\mathcal{I}_-(\nu) := \int_{-\infty}^0 dy \frac{e^{y/y^*}}{y^*} f(\nu - \eta y/\tau) \quad (\text{C7})$$

with the boundary condition  $\lim_{\nu \rightarrow \pm\infty} f(\nu) = 0$ . Let us differentiate both hand sides as

$$\begin{aligned} \frac{d}{d\nu} \mathcal{I}_-(\nu) &= \int_{-\infty}^0 dy \frac{e^{y/y^*}}{y^*} \frac{d}{d\nu} f(\nu - \eta y/\tau) \\ &= -\frac{\tau}{\eta} \int_{-\infty}^0 dy \frac{e^{y/y^*}}{y^*} \frac{d}{dy} f(\nu - \eta y/\tau) \\ &= -\frac{\tau}{\eta} \left[ \frac{e^{y/y^*}}{y^*} f(\nu - \eta y/\tau) \right]_{-\infty}^0 + \frac{\tau}{\eta y^*} \int_0^\infty dy \frac{e^{y/y^*}}{y^*} f(\nu - \eta y/\tau), \end{aligned} \quad (\text{C8})$$

where we have performed an integration by part. This means that

$$\frac{d}{d\nu} \mathcal{I}_-(\nu) = -c f(\nu) + c \mathcal{I}_-(\nu), \quad c := \frac{\tau}{\eta y^*}, \quad (\text{C9})$$

which is equivalent to

$$\left( 1 - \frac{1}{c} \frac{d}{d\nu} \right) \mathcal{I}_-(\nu) = f(\nu). \quad (\text{C10})$$

We thus have the identity (81).

## Appendix D: Summary of special functions

Here we summarise special functions used in this paper.

### 1. Modified Bessel functions

The modified Bessel functions of the first and second kinds, denoted by  $I_\gamma(x)$  and  $K_\gamma(x)$ , are defined by

$$I_\gamma(x) := \sum_{k=0}^{\infty} \frac{1}{k! \Gamma(k + \gamma + 1)} \left(\frac{x}{2}\right)^{2k+\gamma}, \quad (\text{D1a})$$

$$K_\gamma(x) := \frac{\pi}{2} \frac{I_{-\gamma}(x) - I_\gamma(x)}{\sin(\gamma\pi)}. \quad (\text{D1b})$$

### 2. Confluent hypergeometric function

The confluent hypergeometric functions of the first and second kinds are defined by

$${}_1F_1(a, b; x) := \frac{\Gamma(b)}{\Gamma(b-a)\Gamma(a)} \int_0^1 dt e^{xt} t^{a-1} (1-t)^{b-a-1}, \quad (\text{D2a})$$

$${}_1U_1(a, b; x) := \frac{1}{\Gamma(a)} \int_0^\infty dt e^{-xt} t^{a-1} (1+t)^{b-a-1}, \quad (\text{D2b})$$

respectively. For positive  $c$  and  $\beta$ , an asymptotic formula is available for large  $x$

$${}_1U_1\left(1 + \frac{c}{\beta}, 1 + \frac{2c}{\beta}; x\right) \propto x^{-1-c/\beta}. \quad (\text{D3})$$

### 3. Generalised Laguerre function

The generalised Laguerre function  $L_a^b(x)$  is defined as the solution of the following differential equation:

$$x \frac{d^2}{dx^2} L_a^b(x) + (b+1-x) \frac{d}{dx} L_a^b(x) + a L_a^b(x) = 0. \quad (\text{D4})$$

For positive  $c$  and  $\beta$ , an asymptotic formula is available for large  $x$

$$L_{-1-c/\beta}^{2c/\beta}(x) \simeq \frac{\Gamma(c/\beta)}{\Gamma(-c/\beta)\Gamma(1+c/\beta)} e^x x^{-c/\beta}. \quad (\text{D5})$$

### 4. Hypergeometric function

The hypergeometric function is defined as the analytic function whose expansion is given by

$${}_2F_1(a, b, c; z) := \sum_{k=0}^{\infty} \frac{(a)_k (b)_k}{(c)_k} \frac{z^k}{k!} \quad (\text{D6})$$

for  $|z| < 1$  with the Pochhammer symbol  $(a)_k := \Gamma(a+k)/\Gamma(a)$  and  $(a)_0 = 1$ . The hypergeometric function has the integral representation

$${}_2F_1(a, b, c; z) := \frac{\Gamma(c)}{\Gamma(b)\Gamma(c-b)} \int_0^1 t^{b-1} (1-t)^{c-b-1} (1-tz)^{-a} dt \quad (\text{D7})$$

for  $0 < b < c$  and  $|z| < 1$  with real numbers  $a$ ,  $b$ ,  $c$ , and  $z$ . There is an identity

$${}_2F_1(a, b, c; z) = (1-z)^{-a} {}_2F_1\left(a, c-b, c; \frac{z}{z-1}\right). \quad (\text{D8})$$

a. Useful identities

We state the following useful asymptotic formulas: For  $n > 2$  and large  $\nu \rightarrow +\infty$ , the asymptotic formula holds,

$$\nu^2 {}_2F_1\left(1, \frac{2}{n}, 1 + \frac{2}{n}; -\frac{k\nu^n}{\nu_0}\right) = \underbrace{\frac{\frac{2\pi}{n}}{\sin \frac{2\pi}{n}} \left(\frac{\nu_0}{k}\right)^{\frac{2}{n}}}_{\text{const.}} + o(\nu^0). \quad (\text{D9a})$$

In addition, for  $n = 2$  and large  $\nu$ , we obtain

$$\nu^2 {}_2F_1\left(1, 1, 2; -\frac{k\nu^2}{\nu_0}\right) = \frac{\nu_0}{k} \log\left(\frac{k\nu^2}{\nu_0}\right) + o(\nu^0), \quad (\text{D9b})$$

where we have used the identity  ${}_2F_1(1, 1, 2; x) = -x^{-1} \log(1 - x)$ .

b. Crossover between  $n = 2$  and  $n > 2$

The formulas (D9) are qualitatively different for  $n = 2$  and  $n > 2$  because of the analytical singularity of  ${}_2F_1$  at  $n = 2$ . Here we consider the crossover between  $n = 2$  and  $n > 2$ . Since the expansion holds for  $n > 2$

$${}_2F_1\left(1, \frac{2}{n}, 1 + \frac{2}{n}; -x\right) = \frac{\frac{2\pi}{n}}{\sin \frac{2\pi}{n}} x^{-2/n} - \frac{\frac{2}{n}}{\frac{2}{n} - 1} x^{-1} + O(x^{-2}) \quad (\text{D10})$$

for large  $x$ , by the substitution  $x = k\nu^n/\nu_0$ , we obtain the following expansion in terms of the small parameter  $1/x$ , valid for any  $n > 2$ :

$$\nu^2 {}_2F_1\left(1, \frac{2}{n}, 1 + \frac{2}{n}; -\frac{k\nu^n}{\nu_0}\right) \simeq \nu^2 \frac{\frac{2\pi}{n}}{\sin \frac{2\pi}{n}} x^{-1} \left(x^{1-2/n} - \frac{\sin \frac{2\pi}{n}}{\pi(\frac{2}{n} - 1)}\right) = \frac{\nu_0}{k} \nu^{-2\epsilon/(1-\epsilon)} \frac{\pi(1-\epsilon)}{\sin \pi\epsilon} \left(x^{-\epsilon} - \frac{\sin \pi\epsilon}{\pi\epsilon}\right) \quad (\text{D11})$$

with  $\epsilon := 1 - 2/n > 0$ . By taking the limit  $\epsilon \downarrow 0$  for a large but fixed  $\nu$ , we can apply

$$\frac{x^{-\epsilon} - 1}{\epsilon} \simeq -\log x + \frac{\epsilon}{2} (\log x)^2 + \dots \quad (\text{D12})$$

to obtain

$$\lim_{\epsilon \downarrow 0} \left\{ \nu^2 {}_2F_1\left(1, \frac{2}{n}, 1 + \frac{2}{n}; -\frac{k\nu^n}{\nu_0}\right) \right\} \simeq -\frac{\nu_0}{k} \log\left(\frac{k\nu^2}{\nu_0}\right), \quad (\text{D13})$$

recovering formula (D9b).

In contrast, even if  $\epsilon > 0$  is small, the first-order truncation of expansion (D12) is not applicable for too large  $\nu$ 's. Indeed, the truncation is only valid for

$$\frac{\epsilon}{2} (\log x)^2 \ll |\log x| \implies \frac{k}{\nu_0} \nu^n \ll e^{2/\epsilon}. \quad (\text{D14})$$

We thus obtain the characteristic intensity of this crossover as  $\lambda^* := \nu_0 e^{2/\epsilon}$ .

## Appendix E: Solution of an integral equation

Here we study the solution of the integral equation with the following form

$$\int_{-\infty}^{\infty} dy \rho(y) \phi(\nu - y) - \phi(\nu) = 0, \quad (\text{E1})$$

which repeatedly appear in this paper, with the assumption that  $\phi(\nu)$  is nonnegative. We assume that the mark distribution is two-sided

$$p_+ := \int_0^{\infty} \rho(y) dy > 0, \quad p_- := \int_{-\infty}^0 \rho(y) dy > 0 \quad (\text{E2})$$

and that the mean mark is nonpositive

$$m := \int_{-\infty}^{\infty} y \rho(y) dy \leq 0. \quad (\text{E3})$$

### 1. For negative mean mark $m < 0$

Let us first consider the case with negative mean mark  $m < 0$ . Let us assume that a special solution of Eq. (E1) is given by an exponential,

$$\phi(\nu) \simeq C_0 e^{-c\nu} \quad \text{for large } \nu. \quad (\text{E4})$$

By substituting this solution into Eq. (E1), we obtain

$$\Phi(c) = 0, \quad \Phi(x) := \int_{-\infty}^{\infty} \rho(y)(e^{xy} - 1)dy. \quad (\text{E5})$$

Based on this fact, we decompose the general solution of Eq. (E1) as the superposition of exponentials:

$$\phi(\nu) \simeq \sum_i C_i e^{-c_i \nu} \quad \text{for large } \nu, \quad (\text{E6})$$

where  $c_i$  is the  $i$ -th root of  $\Phi(x) = 0$ . Because  $\phi(\nu)$  is nonnegative, oscillatory solutions (corresponding to  $c_i$  being a complex number) are excluded and thus  $c_i$  must be a real number. According to Appendix F, the solutions of  $\Phi(x) = 0$  are given by  $x = 0$  and  $x = c^* > 0$ . Thus, we obtain

$$\phi(x) \simeq C_0 + C_1 e^{-c^* \nu} \quad \text{for large } \nu. \quad (\text{E7})$$

### 2. For zero mean mark $m = 0$

We next consider the case with zero mean mark  $m \uparrow 0$ . According to Appendix F, the roots of  $\Phi(x) = 0$  are given by  $x = 0$  and  $x = c^* > 0$ .

In the zero mean mark limit  $m \uparrow 0$ , the positive root  $c^*$  approaches zero as shown here. Let us assume that the mean mark is negative but very small, such that  $m = -\epsilon$  with a small positive parameter  $\epsilon > 0$ . The moment-generating function is expanded around  $x = 0$  as

$$\Phi(x) = -\epsilon x + \frac{\langle y^2 \rangle x^2}{2} + \dots, \quad \langle y^2 \rangle := \int_{-\infty}^{\infty} y^2 \rho(y) dy. \quad (\text{E8})$$

For small  $\epsilon$ , the positive root of  $\Phi(x) = 0$  is given by

$$c^* = \frac{2\epsilon}{\langle y^2 \rangle} + o(\epsilon) > 0, \quad (\text{E9})$$

which converges to zero for small  $\epsilon$  limit:  $\lim_{\epsilon \downarrow 0} c^* = 0$ .

Assuming a small positive  $\epsilon$ , let us expand the solution (E7) to obtain

$$\phi(x) \simeq C'_0 + C'_1 x + O(C_1 \epsilon^2), \quad C'_0 := C_0 + C_1, \quad C'_1 := -C_1 c^* = O(C_1 \epsilon). \quad (\text{E10})$$

Since  $C_0$  and  $C_1$  are arbitrary constants, we can assume  $C_1 = O(\epsilon^{-1})$ ,  $C'_0 = O(\epsilon^0)$ , and  $C'_1 = O(\epsilon^0)$ , where the divergence of  $C_1$  is absorbed by an appropriate selection of  $C_0$ . Under this assumption, the solution is given by

$$\phi(x) \simeq C'_0 + C'_1 \nu + O(\epsilon^1), \quad C'_0 = O(\epsilon^0), \quad C'_1 = O(\epsilon^0). \quad (\text{E11})$$

By taking the zero mean limit  $\epsilon \downarrow 0$ , we obtain

$$\phi(x) \simeq C'_0 + C'_1 \nu \quad \text{for large } \nu \quad (\text{E12})$$

as the general solution.

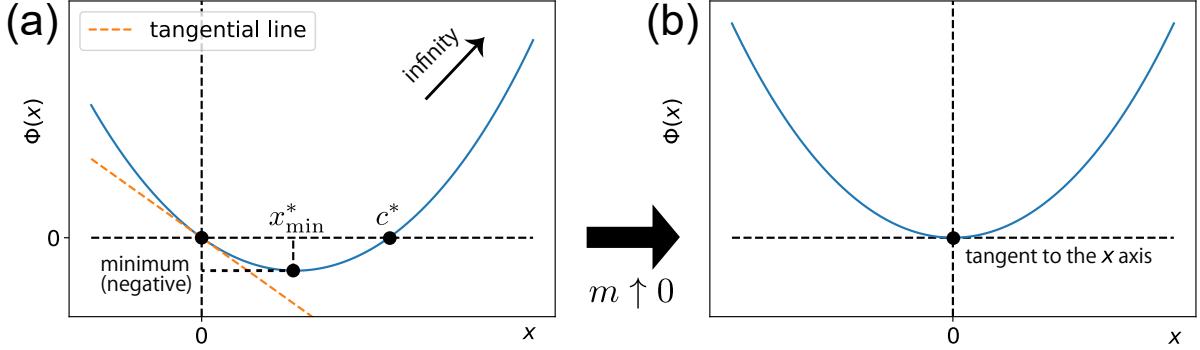


FIG. 12: Schematic of the shape of the moment-generating function  $\Phi(x)$ .  $\Phi(x)$  is strictly convex and takes specific values  $\Phi(0) = 0$ ,  $\Phi(+\infty) = \infty$ , and  $d\Phi(0)/dx = m$ . (a) For negative  $m$ , the tangential line at  $x = 0$  has a negative slope  $d\Phi(0)/dx = m < 0$ . The minimum occurs at  $x = x^*_{\min} > 0$  and the roots of  $\Phi(x) = 0$  are given by  $x = 0$  and  $x = c^* > 0$ . (b) In the zero mean mark limit  $m \uparrow 0$ ,  $c^*$  approaches zero and thus the only root of  $\Phi(x) = 0$  is given by  $x = 0$ .

#### Appendix F: Analytical properties of the moment-generating function $\Phi(x)$

Here we summarise the analytical properties of the moment-generating function

$$\Phi(x) := \int_{-\infty}^{\infty} dy \rho(y) (e^{xy} - 1) \quad (\text{F1})$$

in the regime where the mean mark is non positive

$$m := \int_{-\infty}^{\infty} y \rho(y) dy \leq 0 \quad (\text{F2})$$

and the mark distribution is two-sided

$$p_+ := \int_0^{\infty} \rho(y) dy > 0, \quad p_- := \int_{-\infty}^0 \rho(y) dy > 0. \quad (\text{F3})$$

The moment-generating function  $\Phi(x)$  is a strictly convex function because

$$\frac{d^2 \Phi(x)}{dx^2} = \int_{-\infty}^{\infty} y^2 \rho(y) e^{xy} dy > 0, \quad (\text{F4})$$

implying that  $\Phi(x)$  has no more than one minimum.

##### 1. For negative mean mark $m < 0$

Because we have assumed  $p_+ = \int_0^{\infty} dy \rho(y) > 0$ ,  $\rho(y) \neq 0$  in some finite region of  $y > 0$ . Therefore, the relation

$$m_+ := \int_0^{\infty} y \rho(y) dy > 0 \quad (\text{F5})$$

must hold. Based on this fact, the following three properties hold:

(i) The tangential line of the curve  $\Phi(x)$  at  $x = 0$  has a negative slope:

$$\frac{d\Phi(x)}{dx} \Big|_{x=0} = \int_{-\infty}^{\infty} y\rho(y)dy = m < 0. \quad (\text{F6})$$

(ii) The moment-generating function is zero at  $x = 0$ :

$$\Phi(0) = 0. \quad (\text{F7})$$

(iii) The moment-generating function diverges to infinity for large  $x$

$$\lim_{x \rightarrow \infty} \Phi(x) = \infty \quad (\text{F8})$$

because

$$\begin{aligned} \Phi(x) &= \int_0^{\infty} dy\rho(y)(e^{xy} - 1) + \int_{-\infty}^0 dy\rho(y)(e^{xy} - 1) \\ &\geq \int_0^{\infty} dy\rho(y)xy + \int_{-\infty}^0 dy\rho(y)(0 - 1) \\ &= m_+x - p_- \rightarrow \infty, \end{aligned} \quad (\text{F9})$$

where we have used  $e^{xy} \geq xy + 1$  for  $x \geq 0$  and  $e^{xy} > 0$  for all  $x$ . These properties imply that the minimum of  $\Phi(x)$  exists at some  $x_{\min}^* > 0$  as depicted in Fig. 12a and that the root of  $\Phi(x) = 0$  are given by  $x = 0$  and  $x = c^* > 0$ .

## 2. For zero mean mark $m = 0$

Let us consider the case where the mean mark is zero  $m = 0$ , which is realised typically for symmetric mark distributions  $\rho(y) = \rho(-y)$ . Under this condition, the minimum occurs at  $x = 0$  because

$$\frac{d\Phi(x)}{dx} \Big|_{x=0} = 0. \quad (\text{F10})$$

This implies that the solution of  $\Phi(x) = 0$  is  $x = 0$ , which is a double root (see Fig. 12b). The appearance of the double root can be understood as the zero mean mark limit of the negative mean mark regime. While  $c^*$  is positive for  $m < 0$ , it approaches zero in the zero mean mark limit:  $\lim_{m \uparrow 0} c^* = 0$  (see the discussion in Appendix. E 2).

In the special case of symmetric mark distributions  $\rho(y) = \rho(-y)$ , the moment-generating function can be transformed into

$$\Phi(x) = \int_{-\infty}^{\infty} \rho(y)(\cosh xy - 1)dy \quad (\text{F11})$$

Also, considering the geometrical shape (see Fig. 12b), the equation

$$\Phi(x) = ax \quad (\text{F12})$$

has a single positive root for positive  $a$ .

## 3. Special cases

For reference, we summarise several results for specific mark distributions.

a. Gaussian mark distribution

Let us consider the Gaussian mark distribution

$$\rho(y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(y-m)^2/(2\sigma^2)} \quad (\text{F13})$$

with mean mark  $m$  and variance  $\sigma^2$ . The corresponding moment-generating function is given by

$$\Phi(x) = e^{mx+\sigma^2 x^2/2} - 1, \quad (\text{F14})$$

which leads to the explicit formula for the root of  $\Phi(c^*) = 0$  as

$$c^* = -\frac{2m}{\sigma^2}. \quad (\text{F15})$$

b. Two-sided exponential asymmetric mark distribution

We next consider the case of the two-sided exponential asymmetric mark distribution

$$\rho(y) = \begin{cases} \frac{p_+}{y_+^*} e^{-y/y_+^*} & (y \geq 0) \\ \frac{p_-}{y_-^*} e^{y/y_-^*} & (y < 0) \end{cases}, \quad (\text{F16})$$

where  $p_+ + p_- = 1$ ,  $y_+^* > 0$ , and  $y_-^* > 0$ . The mean mark is given by

$$m := y_+^* p_+ - y_-^* p_- < 0. \quad (\text{F17})$$

We obtain

$$\Phi(x) = \frac{p_+ y_+^* x}{1 - y_+^* x} - \frac{p_- y_-^* x}{1 + y_-^* x}, \quad (\text{F18})$$

which leads to

$$c^* = \frac{p_- y_-^* - p_+ y_+^*}{y_+^* y_-^*} = -\frac{m}{y_+^* y_-^*} > 0 \quad (\text{F19})$$

as the unique positive root of  $\Phi(c^*) = 0$ .

## Appendix G: Proofs of mathematical properties of $\mathbf{H}$ (203)

Here, we summarize the proofs of the main mathematical properties of  $\mathbf{H}$  defined by Eq. (203).

### 1. Proof that eigenvalues are real

We show that all eigenvalues of  $\mathbf{H}$  are real numbers as follows. First,  $\mathbf{H}$  can be symmetrized as  $\bar{\mathbf{H}}$ , defined by

$$\bar{\mathbf{H}} := \mathbf{A} \mathbf{H} \mathbf{A}^{-1} = \begin{pmatrix} \frac{1}{\tau_1} - \tilde{h}_1, & -\sqrt{\tilde{h}_1 \tilde{h}_2}, & \dots & -\sqrt{\tilde{h}_1 \tilde{h}_K} \\ -\sqrt{\tilde{h}_2 \tilde{h}_1}, & \frac{1}{\tau_2} - \tilde{h}_2, & \dots & -\sqrt{\tilde{h}_2 \tilde{h}_K} \\ \vdots & \vdots & \ddots & \vdots \\ -\sqrt{\tilde{h}_K \tilde{h}_1}, & -\sqrt{\tilde{h}_K \tilde{h}_2}, & \dots & \frac{1}{\tau_K} - \tilde{h}_K \end{pmatrix}, \quad \mathbf{A} := \begin{pmatrix} \sqrt{\tilde{h}_1}, & 0, & \dots & 0 \\ 0, & \sqrt{\tilde{h}_2}, & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0, & 0, & \dots & \sqrt{\tilde{h}_K} \end{pmatrix}. \quad (\text{G1})$$

Indeed, by representing all the matrices by their elements  $\bar{\mathbf{H}} := (\bar{H}_{ij})$ ,  $\mathbf{H} := (H_{ij})$ , and  $\mathbf{A} := A_{ij}$ , we obtain

$$\bar{H}_{ij} = \sum_{k,l} A_{ik} H_{kl} A_{lj}^{-1} = \sum_{k,l} \sqrt{\tilde{h}_i} \delta_{ik} \left( \frac{\delta_{kl}}{\tau_k} - \tilde{h}_l \right) \sqrt{\frac{1}{\tilde{h}_j}} \delta_{lj} = \frac{\delta_{ij}}{\sqrt{\tau_i \tau_j}} - \sqrt{\tilde{h}_i \tilde{h}_j}. \quad (\text{G2})$$

Since  $\bar{\mathbf{H}}$  is a symmetric matrix, all their eigenvalues are real. We obtain

$$\mathbf{H}\mathbf{e}_i = \lambda_i \mathbf{e}_i \iff \bar{\mathbf{H}}(\mathbf{A}\mathbf{e}_i) = \lambda_i (\mathbf{A}\mathbf{e}_i). \quad (\text{G3})$$

This relationship implies that any the eigenvalues of  $\mathbf{H}$  are the same as that of  $\bar{\mathbf{H}}$ . Thus, all the eigenvalues of  $\mathbf{H}$  are real likewise.

## 2. Determinant

The determinant  $\det \mathbf{H}$  is derived as follows. Let us recall the invariance of determinants

$$\det \mathbf{H} = \det \begin{pmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \vdots \\ \mathbf{a}_j \\ \vdots \\ \mathbf{a}_K \end{pmatrix} = \det \begin{pmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \vdots \\ \mathbf{a}_j + c\mathbf{a}_k \\ \vdots \\ \mathbf{a}_K \end{pmatrix} \quad (\text{G4})$$

for any constant  $c$ . This implies

$$\det \mathbf{H} = \det \begin{pmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \mathbf{a}_3 \\ \vdots \\ \mathbf{a}_K \end{pmatrix} = \det \begin{pmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 - \mathbf{a}_1 \\ \mathbf{a}_3 \\ \vdots \\ \mathbf{a}_K \end{pmatrix} = \det \begin{pmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 - \mathbf{a}_1 \\ \mathbf{a}_3 - \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_K \end{pmatrix} = \dots = \det \begin{pmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 - \mathbf{a}_1 \\ \mathbf{a}_3 - \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_K - \mathbf{a}_1 \end{pmatrix} := \det \begin{pmatrix} \mathbf{a}'_1 \\ \mathbf{a}'_2 \\ \mathbf{a}'_3 \\ \vdots \\ \mathbf{a}'_K \end{pmatrix} \quad (\text{G5})$$

and

$$\det \mathbf{H} = \det \begin{pmatrix} \mathbf{a}'_1 \\ \mathbf{a}'_2 \\ \vdots \\ \mathbf{a}'_K \end{pmatrix} = \det \begin{pmatrix} \mathbf{a}'_1 + \tau_2 \tilde{h}_2 \mathbf{a}'_2 \\ \mathbf{a}'_2 \\ \vdots \\ \mathbf{a}'_K \end{pmatrix} = \det \begin{pmatrix} \mathbf{a}'_1 + \tau_2 \tilde{h}_2 \mathbf{a}'_2 + \tau_3 \tilde{h}_3 \mathbf{a}'_3 \\ \mathbf{a}'_2 \\ \vdots \\ \mathbf{a}'_K \end{pmatrix} = \dots = \det \begin{pmatrix} \mathbf{a}'_1 + \sum_{k=2}^K \tau_k \tilde{h}_k \mathbf{a}'_k \\ \mathbf{a}'_2 \\ \vdots \\ \mathbf{a}'_K \end{pmatrix} \quad (\text{G6})$$

with constants  $\{\tau_k \tilde{h}_k\}_k$ . Using these relations, the determinant of  $\mathbf{H}$  is given by

$$\begin{aligned} \det \mathbf{H} &= \det \begin{pmatrix} -\tilde{h}_1 + 1/\tau_1, & -\tilde{h}_2, & \dots, & -\tilde{h}_K \\ -\tilde{h}_1, & -\tilde{h}_2 + 1/\tau_2, & \dots, & -\tilde{h}_K \\ \vdots & \vdots & \ddots & \vdots \\ -\tilde{h}_1, & -\tilde{h}_2, & \dots, & -\tilde{h}_K + 1/\tau_K \end{pmatrix} \leftarrow \mathbf{a}_1 \\ &= \det \begin{pmatrix} (1 - \tau_1 \tilde{h}_1)/\tau_1, & -\tilde{h}_2, & \dots, & -\tilde{h}_K \\ -1/\tau_1, & 1/\tau_2, & \dots, & 0 \\ \vdots & \vdots & \ddots & \vdots \\ -1/\tau_1, & 0, & \dots, & 1/\tau_K \end{pmatrix} \leftarrow \mathbf{a}'_2 = \mathbf{a}_2 - \mathbf{a}_1 \\ &= \det \begin{pmatrix} (1 - \sum_{k=1}^K \tau_k \tilde{h}_k)/\tau_1, & 0, & \dots, & 0 \\ -1/\tau_1, & 1/\tau_2, & \dots, & 0 \\ \vdots & \vdots & \ddots & \vdots \\ -1/\tau_1, & 0, & \dots, & 1/\tau_K \end{pmatrix} \leftarrow \mathbf{a}''_2 = \mathbf{a}'_2 + \sum_{k=2}^K \tau_k \tilde{h}_k \mathbf{a}'_k \\ &= \frac{1 - \sum_{k=1}^K \tau_k \tilde{h}_k}{\tau_1 \dots \tau_K}. \end{aligned} \quad (\text{G7})$$

Notably,  $\det \mathbf{H} = 0$  at criticality  $\eta = 1$ . This singularity is consistent with the singularity of the inverse matrix  $\mathbf{H}^{-1}$ , as discussed in Appendix G 3.

### 3. Inverse matrix

The inverse matrix of  $\mathbf{H}$  is derived from the method of row reduction:

$$\begin{aligned}
& \left( \begin{array}{cccc|cc} -\tilde{h}_1 + 1/\tau_1, & -\tilde{h}_2, & \dots & -\tilde{h}_K & 1, 0, \dots, 0 \\ -\tilde{h}_1, & -\tilde{h}_2 + 1/\tau_2, & \dots & -\tilde{h}_K & 0, 1, \dots, 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \vdots \ddots \vdots \\ -n_1/\tau_1, & -n_2/\tau_2, & \dots, & -n_K/\tau_K + 1/\tau_K & 0, 0, \dots, 1 \end{array} \right) \leftarrow \mathbf{b}_1 \\
\rightarrow & \left( \begin{array}{cccc|cc} (1 - \tau_1 \tilde{h}_1)/\tau_1, & -\tilde{h}_2, & \dots & -\tilde{h}_K & 1, 0, \dots, 0 \\ -1/\tau_1, & 1/\tau_2, & \dots & 0 & -1, 1, \dots, 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \vdots \ddots \vdots \\ -1/\tau_1, & 0, & \dots & 1/\tau_K & -1, 0, \dots, 1 \end{array} \right) \leftarrow \mathbf{b}'_1 = \mathbf{b}_1 - \mathbf{b}_1 \\
\rightarrow & \left( \begin{array}{cccc|cc} (1 - \eta)/\tau_1, & 0, & \dots & 0 & 1 - \sum_{k=2}^K \tau_k \tilde{h}_k, & \tau_2 \tilde{h}_2, & \dots, & \tau_K \tilde{h}_K \\ -1/\tau_1, & 1/\tau_2, & \dots & 0 & -1, & 1, & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ -1/\tau_1, & 0, & \dots & 1/\tau_K & -1, & 0, & \dots & 1 \end{array} \right) \leftarrow \mathbf{b}''_1 = \mathbf{b}_1' + \sum_{k=2}^K \tau_k \tilde{h}_k \mathbf{b}'_k \\
\rightarrow & \left( \begin{array}{cccc|cc} 1, & 0, & \dots & 0 & \tau_1 + \tau_1 n_1/(1 - \eta), & \tau_1 n_2/(1 - \eta), & \dots, & \tau_1 n_K/(1 - \eta) \\ -\tau_2/\tau_1, & 1, & \dots & 0 & -\tau_2, & \tau_2, & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ -\tau_K/\tau_1, & 0, & \dots & 1 & -\tau_K, & 0, & \dots & \tau_K \end{array} \right) \leftarrow \mathbf{b}'''_1 = \frac{\tau_1}{(1 - \eta)} \mathbf{b}''_1 \\
\rightarrow & \left( \begin{array}{cccc|cc} 1, & 0, & \dots & 0 & \tau_1 + \tau_1^2 \tilde{h}_1/(1 - \eta), & \tau_1 \tau_2 \tilde{h}_2/(1 - \eta), & \dots, & \tau_1 \tau_K \tilde{h}_K/(1 - \eta) \\ 0, & 1, & \dots & 0 & \tau_2 \tau_1 \tilde{h}_1/(1 - \eta), & \tau_2 + \tau_2^2 \tilde{h}_2/(1 - \eta), & \dots, & \tau_2 \tau_K \tilde{h}_K/(1 - \eta) \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0, & 0, & \dots & 1 & \tau_K \tau_1 \tilde{h}_1/(1 - \eta), & \tau_K \tau_2 \tilde{h}_2/(1 - \eta), & \dots, & \tau_K + \tau_K^2 \tilde{h}_K/(1 - \eta) \end{array} \right) \leftarrow \mathbf{b}''''_1 = \mathbf{b}'''_1 + \frac{\tau_2}{\tau_1} \mathbf{b}'''_2 \\
& \vdots \\
& \left( \begin{array}{cccc|cc} 1, & 0, & \dots & 0 & \tau_1 + \tau_1^2 \tilde{h}_1/(1 - \eta), & \tau_1 \tau_2 \tilde{h}_2/(1 - \eta), & \dots, & \tau_1 \tau_K \tilde{h}_K/(1 - \eta) \\ 0, & 1, & \dots & 0 & \tau_2 \tau_1 \tilde{h}_1/(1 - \eta), & \tau_2 + \tau_2^2 \tilde{h}_2/(1 - \eta), & \dots, & \tau_2 \tau_K \tilde{h}_K/(1 - \eta) \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0, & 0, & \dots & 1 & \tau_K \tau_1 \tilde{h}_1/(1 - \eta), & \tau_K \tau_2 \tilde{h}_2/(1 - \eta), & \dots, & \tau_K + \tau_K^2 \tilde{h}_K/(1 - \eta) \end{array} \right) \leftarrow \mathbf{b}''''_K = \mathbf{b}''''_1 + \frac{\tau_2}{\tau_1} \mathbf{b}''''_2
\end{aligned} \tag{G8}$$

with the branching ratio  $\eta$  defined by

$$\eta := \sum_{k=1}^K \tau_k \tilde{h}_k. \tag{G9}$$

This relation implies

$$\mathbf{H}^{-1} = \begin{pmatrix} \tau_1 + \tau_1^2 \tilde{h}_1/(1 - \eta), & \tau_1 \tau_2 \tilde{h}_2/(1 - \eta), & \dots, & \tau_1 \tau_K \tilde{h}_K/(1 - \eta) \\ \tau_2 \tau_1 \tilde{h}_1/(1 - \eta), & \tau_2 + \tau_2^2 \tilde{h}_2/(1 - \eta), & \dots, & \tau_2 \tau_K \tilde{h}_K/(1 - \eta) \\ \vdots & \vdots & \ddots & \vdots \\ \tau_K \tau_1 \tilde{h}_1/(1 - \eta), & \tau_K \tau_2 \tilde{h}_2/(1 - \eta), & \dots, & \tau_K + \tau_K^2 \tilde{h}_K/(1 - \eta) \end{pmatrix} \tag{G10}$$

or equivalently

$$H_{ij}^{-1} = \tau_i \delta_{ij} + \frac{\tau_i \tau_j \tilde{h}_j}{1 - \eta} \tag{G11}$$

in the representation by matrix elements. The above calculation can be directly confirmed as follows:

$$\mathbf{H} \mathbf{H}^{-1} = \mathbf{I} \iff \sum_{j=1}^K H_{ij} H_{jk}^{-1} = \sum_{j=1}^K \left( -\tilde{h}_j + \frac{1}{\tau_j} \delta_{ij} \right) \left( \tau_j \delta_{jk} + \frac{\tau_j \tau_k \tilde{h}_k}{1 - \eta} \right) = \delta_{ik}. \tag{G12}$$

The inverse matrix has a singularity at  $\eta = 1$ , corresponding to the criticality of the ramp Hawkes process.

#### 4. Eigenvectors of $\mathbf{H}$

Since  $\mathbf{H}$  is directly associated with the real symmetric matrix  $\tilde{\mathbf{H}}$ ,  $\mathbf{H}$  can be diagonalised, such that

$$\mathbf{P} := (\mathbf{e}_1, \dots, \mathbf{e}_K), \quad \mathbf{P}^{-1} \mathbf{H} \mathbf{P} = \begin{pmatrix} \lambda_1, & 0, & \dots & 0 \\ 0, & \lambda_2, & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0, & 0, & \dots & \lambda_K \end{pmatrix} \quad (\text{G13})$$

with eigenvectors  $\{\mathbf{e}_k\}_{k=1,\dots,K}$  and corresponding eigenvalues  $\{\lambda_k\}_{k=1,\dots,K}$ .

At criticality  $\eta = 1$ , the smallest eigenvalue is zero, such that  $\lambda_1 = 0$ . In addition, the zero eigenvector  $\mathbf{e}_1$  is explicitly given by

$$\mathbf{e}_1 = \begin{pmatrix} \tau_1 \\ \tau_2 \\ \vdots \\ \tau_K \end{pmatrix} \quad (\text{G14})$$

Indeed, we can directly confirm the following relationship:

$$\mathbf{H} \mathbf{e}_1 = \begin{pmatrix} \frac{1}{\tau_1} - \tilde{h}_1, & -\tilde{h}_2, & \dots & -\tilde{h}_K \\ -\tilde{h}_1, & \frac{1}{\tau_2} - \tilde{h}_2, & \dots & -\tilde{h}_K \\ \vdots & \vdots & \ddots & \vdots \\ -\tilde{h}_1, & -\tilde{h}_2, & \dots & \frac{1}{\tau_K} - \tilde{h}_K \end{pmatrix} \begin{pmatrix} \tau_1 \\ \tau_2 \\ \vdots \\ \tau_K \end{pmatrix} = \begin{pmatrix} 1 - \eta \\ 1 - \eta \\ \vdots \\ 1 - \eta \end{pmatrix} = \mathbf{0} \quad \text{for } \eta := \sum_{k=1}^K \tau_k \tilde{h}_k = 1. \quad (\text{G15})$$

We next consider the representation based on the eigenvectors:

$$\mathbf{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_K \end{pmatrix} := \mathbf{P}^{-1} \mathbf{s}, \quad \mathbf{P}^{-1} = \begin{pmatrix} \mathbf{g}_1 \\ \mathbf{g}_2 \\ \vdots \\ \mathbf{g}_K \end{pmatrix}. \quad (\text{G16})$$

On the basis of this representation, we obtain

$$\frac{dX_1}{dl} = 0 + O(\mathbf{X}^2), \quad \frac{dX_k}{dl} = -\lambda_k X_k + O(\mathbf{X}^2) \quad \text{for } k \geq 2. \quad (\text{G17})$$

This implies that the leading-order contribution comes from the  $X_1$  direction, such that  $|X_1| \gg |X_k|$  with  $k \geq 2$ . We can approximate

$$\mathbf{X} = \begin{pmatrix} X_1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + O(\mathbf{X}^2) \implies \mathbf{s} = \mathbf{P} \mathbf{X} \simeq X_1 \mathbf{e}_1 + O(\mathbf{X}^2). \quad (\text{G18})$$

By direct substitution, we can confirm that  $\mathbf{g}_1$  is given by

$$\mathbf{g}_1 = \left( \frac{\tau_1 \tilde{h}_1}{\langle \tau \rangle}, \dots, \frac{\tau_K \tilde{h}_K}{\langle \tau \rangle} \right), \quad \langle \tau \rangle := \sum_{k=1}^K \tau_k^2 \tilde{h}_k. \quad (\text{G19})$$

Indeed, this implies that  $X_1$  is given by

$$X_1 = \mathbf{g}_1 \cdot \mathbf{s} = \frac{1}{\langle \tau \rangle} \sum_{k=1}^K \tau_k \tilde{h}_k s_k, \quad (\text{G20})$$

which leads to

$$\frac{dX_1}{dl} = \frac{1}{\langle \tau \rangle} \sum_{k=1}^K \tau_k \tilde{h}_k \frac{ds_k}{dl} = 0 + O(\mathbf{X}^2). \quad (\text{G21})$$

Thus, we find that the first-order contribution is absent in Eq. (G21), confirming the correctness of the representation of Eq. (G19). In addition, this representation (G19) is consistent with the following identity:

$$\mathbf{P}^{-1}\mathbf{P} = \begin{pmatrix} \mathbf{g}_1 \\ \mathbf{g}_2 \\ \vdots \\ \mathbf{g}_K \end{pmatrix} (\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_K) = \begin{pmatrix} \tau_1 \tilde{h}_1 / \langle \tau \rangle, & \tau_2 \tilde{h}_2 / \langle \tau \rangle, & \dots & \tau_K \tilde{h}_K / \langle \tau \rangle \\ \bigcirc, & \bigcirc, & \dots & \bigcirc \\ \vdots & \vdots & \ddots & \vdots \\ \bigcirc, & \bigcirc, & \dots & \bigcirc \end{pmatrix} \begin{pmatrix} \tau_1, & \bigcirc, & \dots & \bigcirc \\ \tau_2, & \bigcirc, & \dots & \bigcirc \\ \vdots & \vdots & \ddots & \vdots \\ \tau_2, & \bigcirc, & \dots & \bigcirc \end{pmatrix} = \begin{pmatrix} 1, & 0, & \dots & 0 \\ 0, & 1, & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0, & 0, & \dots & 1 \end{pmatrix}, \quad (\text{G22})$$

where  $\bigcirc$  represents some unspecified value<sup>6</sup>. Thus, we find that Eq. (G19) is the correct and consistent representation.

## Appendix H: Summary of asymptotic forms of the Laplace transform

Here we summarise the asymptotic forms of the Laplace transform, in particular for power law distributions. Let us first recall the Tauberian theorem for the Laplace transform of asymptotic power law functions [55]:

Let us consider a function  $f(x)$  satisfying the asymptotic form

$$f(x) \simeq x^{\rho-1} L(x) \quad \text{for large } x \quad (\text{H1})$$

with  $0 < \rho < \infty$  and slowly varying function  $L(x)$ . By definition, a slowly varying function satisfies  $\lim_{x \rightarrow \infty} (L(Cx)/L(x)) = 1$  for any positive constant  $C$ . The Laplace transform of  $f(x)$  has the asymptotic form

$$\tilde{f}(s) := \mathcal{L}_1[f(x); s] \simeq \Gamma(\rho) s^{-\rho} L(1/s) \quad \text{for small } s. \quad (\text{H2})$$

Using this theorem, let us consider the Laplace transform of power law functions  $f(x) \simeq Ax^{-1-a}$  for various  $a$  and positive constant  $A$ .

### 1. Negative case: $a < 0$

For  $a < 0$ , the Tauberian theorem can be readily applied to obtain

$$\tilde{f}(s) \simeq A' s^a, \quad A' := A \Gamma(-a) > 0 \quad \text{for small } s. \quad (\text{H3})$$

### 2. Positive case: $0 < a < 1$

Let us consider the following relation

$$\tilde{f}(s) = \int_0^\infty f(x) e^{-sx} dx = \left[ -F^{(0)}(x) e^{-sx} \right]_0^\infty - s \int_0^\infty F^{(0)}(x) e^{-sx} dx = F^{(0)}(0) - s \mathcal{L}_1[F^{(0)}(x); s], \quad F^{(0)}(x) := \int_x^\infty f(x') dx'. \quad (\text{H4})$$

Here we notice that the asymptotic tail of  $F^{(0)}(x)$  satisfies the condition of the Tauberian theorem, such that

$$F^{(0)}(x) \simeq \int_x^\infty \frac{Adx}{x^{1+a}} = \frac{A}{a} x^{-a} \quad \text{for large } x. \quad (\text{H5})$$

---

<sup>6</sup> If we set  $\mathbf{g}_1 = \left( \frac{\tau_1 \tilde{h}_1}{c}, \dots, \frac{\tau_K \tilde{h}_K}{c} \right)$  with some constant  $c \neq \langle \tau \rangle$ , the identity  $\mathbf{P}^{-1}\mathbf{P} = \mathbf{E}$  does not hold with the unit vector  $\mathbf{E}$ , while the relation  $dX_1/dl = 0 + O(s^2)$  still holds. Therefore,  $c$  must be  $\langle \tau \rangle$  for the self consistency.

By applying the Tauberian theorem, we obtain

$$\tilde{f}(s) \simeq F^{(0)}(0) - A's^a + o(s^a), \quad A' := \frac{A}{a}\Gamma(1-a) > 0. \quad (\text{H6})$$

When  $f(x)$  is a PDF,  $F^{(0)}(0) = \int_0^\infty dx f(x) = 1$  and we obtain

$$\tilde{f}(s) \simeq 1 - A's^a + o(s^a). \quad (\text{H7})$$

### 3. General positive case: non-integer $0 < a$

Let us introduce the integer  $m := \lfloor a \rfloor = \max\{k \in \mathbf{Z} \mid k \leq a\}$  with the set of integers  $\mathbf{Z}$ , satisfying  $m \leq a \leq m+1$ , and the iterated integral of  $f(x)$ :

$$F^{(l)}(x) := \int_x^\infty dx_{l+1} \int_{x_{l+1}}^\infty dx_l \cdots \int_{x_2}^\infty dx_1 f(x). \quad (\text{H8})$$

We find that we can apply the Tauberian theorem to  $F^{(m)}(x)$ , because

$$F^{(m)}(x) \simeq \frac{A(-1)^{m+1}}{(-a)(1-a)\dots(m-a)} x^{m-a} = A(-1)^{m+1} \frac{\Gamma(-a)}{\Gamma(m-a+1)} x^{m-a} \quad (\text{H9})$$

with  $-1 < m-a < 0$ . Due to the identity

$$\mathcal{L}_1 \left[ F^{(l)}(x); s \right] = F^{(l+1)}(0) - s \mathcal{L}_1 \left[ F^{(l+1)}(x); s \right], \quad (\text{H10})$$

we obtain an asymptotic relation for small  $s$ ,

$$\begin{aligned} \mathcal{L}_1 \left[ f(x); s \right] &= \sum_{k=0}^m (-s)^k F^{(k)}(0) + (-s)^{m+1} \mathcal{L}_1 \left[ F^{(m)}(x); s \right] \\ &\simeq \sum_{k=0}^m (-s)^k F^{(k)}(0) - A(-1)^m \Gamma(-a) s^a + o(s^a). \end{aligned} \quad (\text{H11})$$

Since the iterated integral satisfies the identity

$$F^{(l)}(x) = \frac{1}{l!} \int_x^\infty (x' - x)^l f(x') dx', \quad (\text{H12})$$

one finds that  $F^l(0)$  is proportional to the  $l$ -th order moment  $M_l$  when  $f(x)$  is a PDF:

$$M_l := \int_0^\infty x^l f(x) dx = l! F^{(l)}(0). \quad (\text{H13})$$

In other words,

$$\mathcal{L}_1 \left[ f(x); s \right] \simeq \sum_{k=0}^m \frac{(-1)^k M_k}{k!} s^k + A's^a + o(s^a), \quad A' := A\Gamma(-a). \quad (\text{H14})$$

Since  $A$  is positive for  $f(x)$  to be a PDF (i.e.,  $f(x) \geq 0$ ),  $A'$  must be negative (positive) for even (odd)  $m$ .

## Appendix I: Numerical implementation

This appendix describes our numerical method for the simulation of the NLHawkes processes. Our starting point is the Markovian SDE (18) for the discrete sum of exponentials. We introduce a discretised time series  $\{t_i\}_i$  and time

steps  $\Delta t_i := t_{i+1} - t_i$ , satisfying  $0 = t_0 < t_1 < \dots < t_{N_T} = T_{\text{tot}}$ . Given that the intensity of the state-dependent Poisson process is given by  $\hat{\lambda}(t) = G(\hat{z}(t))$ , the discrete version of the SDE (18) is given by

$$\hat{z}_k(t_i + \Delta t_i) = \hat{z}_k(t_i) e^{-\frac{\Delta t_i}{\tau_k}} + \begin{cases} 0 & (\text{Probability: } 1 - G(\hat{z}(t_k))\Delta t_k) \\ \tilde{h}_k \hat{y}_i & (\text{Probability: } G(\hat{z}(t_k))\Delta t_k) \end{cases} \quad (\text{I1})$$

with the IID random number sequence  $\{\hat{y}_i\}_i$  obeying the mark distribution  $\rho(y)$ . Given that  $G(\hat{z}(t_k))\Delta t_k$  must be sufficiently small such that  $G(\hat{z}(t_k))\Delta t_k \ll 1$  for a proper probability interpretation, we employ an adaptive scheme for the time discretisation

$$\Delta t_i = \min \left\{ \Delta t_{\max}^{(1)}, \frac{\Delta t_{\max}^{(2)}}{G(\hat{z}(t_i))} \right\}, \quad (\text{I2})$$

because the intensity  $G(\hat{z}(t_i))$  sometimes takes extremely large values near criticality. For this setup, we obtained an empirical intensity distribution by assuming ergodicity as

$$P_{\text{ss}}(\lambda) = \lim_{t \rightarrow \infty} \langle \delta(\lambda - \hat{\lambda}(t)) \rangle = \lim_{T_{\text{tot}} \rightarrow \infty} \frac{1}{T_{\text{tot}}} \int_0^{T_{\text{tot}}} \delta(\lambda - G(\hat{z}(t))) dt \simeq \frac{1}{T_{\text{tot}}} \sum_{i=0}^{N_T-1} \delta(\lambda - G(\hat{z}(t_i))) \Delta t_i. \quad (\text{I3})$$

For its practical implementation, we have applied a parallel computing technique for better convergence. We have obtained the empirical intensity distribution as

$$P_{\text{ss}}(\lambda) = \left\langle \lim_{T_{\text{tot}} \rightarrow \infty} \frac{1}{T_{\text{tot}}} \int_0^{T_{\text{tot}}} \delta(\lambda - G(\hat{z}(t))) dt \right\rangle \simeq \frac{1}{N_{\text{PC}}} \sum_{j=1}^{N_{\text{PC}}} \left[ \frac{1}{T_{\text{tot}}} \sum_{i=0}^{N_T^{(j)}-1} \delta(\lambda - G(\hat{z}^{(j)}(t_i^{(j)}))) \Delta t_i^{(j)} \right], \quad (\text{I4})$$

where  $\hat{z}^{(j)}(t)$  is the trajectory obtained in the  $j$ -th parallel thread and  $N_{\text{PC}}$  is the number of total parallel threads.

#### a. Ramp intensity map without inhibitory effect (Fig. 9)

We describe the setup for Fig. 9, where the intensity function is given by the ramp function and the mark takes a single value as

$$G(\hat{z}) := \max \left\{ \sum_{k=1}^K \hat{z}_k - \nu_1, \nu_0 \right\}, \quad \rho(y) = \delta(y - 1) \quad (\text{I5})$$

with any positive number  $\nu_0$  and any real number  $\nu_1$ .

*Figure 9a.* The parameters are given by  $K = 2$ ,  $(\tau_1, \tau_2) = (1, 2)$ ,  $(\tilde{h}_1, \tilde{h}_2) = (0.7, 0.14995)$ ,  $\eta = 0.9999$ ,  $\nu_0 = 0.01$ ,  $\nu_1 \simeq 0.385$ ,  $T_{\text{tot}} = 5 \times 10^6$ ,  $\Delta t_{\max}^{(1)} = 0.1$ , and  $\Delta t_{\max}^{(2)} = 0.01$  with the initial condition  $(\hat{z}_1(0), \hat{z}_2(0)) = (1, 1)$ . Since  $\alpha_2 = 1$ , we obtain the power law exponent  $a \simeq 1.0$  from Eq. (193). The total number of parallel threads is given by  $N_{\text{PC}} = 4$ .

*Figure 9b.* The parameters are given by  $K = 3$ ,  $(\tau_1, \tau_2, \tau_3) = (1, 2, 3)$ ,  $(\tilde{h}_1, \tilde{h}_2, \tilde{h}_3) = (0.5, 0.15, 0.1999/3)$ ,  $\eta = 0.9999$ ,  $\nu_0 = 0.01$ ,  $\nu_1 \simeq 0.147$ ,  $T_{\text{tot}} = 5 \times 10^6$ ,  $\Delta t_{\max}^{(1)} = 0.1$ , and  $\Delta t_{\max}^{(2)} = 0.01$  with the initial condition  $(\hat{z}_1(0), \hat{z}_2(0), \hat{z}_3(0)) = (1, 1, 1)$ . Since  $\alpha_2 = 1$ , we obtain the power law exponent  $a \simeq 0.5$  from Eq. (193). The total number of parallel threads is given by  $N_{\text{PC}} = 4$ .

*Figure 9c.* The parameters are given by  $K = 3$ ,  $(\tau_1, \tau_2, \tau_3) = (1, 2, 3)$ ,  $(\tilde{h}_1, \tilde{h}_2, \tilde{h}_3) = (0.5, 0.15, 0.1999/3)$ ,  $\eta = 0.9999$ ,  $\nu_0 = 0.01$ ,  $\nu_1 = 0$ ,  $T_{\text{tot}} = 5 \times 10^5$ ,  $\Delta t_{\max}^{(1)} = 0.1$ , and  $\Delta t_{\max}^{(2)} = 0.01$  with the initial condition  $(\hat{z}_1(0), \hat{z}_2(0), \hat{z}_3(0)) = (1, 1, 1)$ . Since  $\alpha_2 = 1$ , we obtain the power law exponent  $a = 0$  from Eq. (193). The total number of parallel threads is given by  $N_{\text{PC}} = 4$ .

#### b. MSA intensity map with inhibitory effect (Fig. 10)

We describe the setup for Fig. 10, where the intensity map and the mark distribution are given by the exponential function with finite cutoff and the normal distribution, respectively, such that

$$G(\hat{z}) = \min \left\{ \lambda_0 \exp \left[ \beta \sum_{k=1}^K \hat{z}_k \right], \lambda_{\max} \right\}, \quad \rho(y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{y^2}{2\sigma^2}}. \quad (\text{I6})$$

*Figure 10a.* The parameters are given by  $K = 3$ ,  $(\tau_1, \tau_2, \tau_3) = (1, 0.5, 2)$ ,  $(\tilde{h}_1, \tilde{h}_2, \tilde{h}_3) = (0.5, 0.6, 0.1)$ ,  $\lambda_0 = 0.01$ ,  $\beta = 6$ ,  $\lambda_{\max} = 10^6$ ,  $\sigma = 0.3$ ,  $T_{\text{tot}} = 5 \times 10^6$ ,  $\Delta t_{\max}^{(1)} = 0.1$ , and  $\Delta t_{\max}^{(2)} = 0.01$  with the initial condition  $(\hat{z}_1(0), \hat{z}_2(0), \hat{z}_3(0)) = (0, 0, 0)$ . The total number of parallel threads is given by  $N_{\text{PC}} = 8$ .

*Figure 10b.* The parameters are given by  $K = 3$ ,  $(\tau_1, \tau_2, \tau_3) = (1, 0.5, 2)$ ,  $(\tilde{h}_1, \tilde{h}_2, \tilde{h}_3) = (0.5, 0.6, 0.1)$ ,  $\lambda_0 = 0.001$ ,  $\beta = 10$ ,  $\lambda_{\max} = 10^6$ ,  $\sigma = 0.3$ ,  $T_{\text{tot}} = 5 \times 10^5$ ,  $\Delta t_{\max}^{(1)} = 0.1$ , and  $\Delta t_{\max}^{(2)} = 0.01$  with the initial condition  $(\hat{z}_1(0), \hat{z}_2(0), \hat{z}_3(0)) = (0, 0, 0)$ .

### c. Brownian motion with $\hat{\nu}$ -dependent diffusion constant (Fig. 11)

For Fig. 11, we describe the numerical method for the Brownian motion with  $\hat{\nu}$ -dependent diffusion constant governed by the SDE (261). The numerical simulation is based on the following discrete version

$$\hat{\nu}(t_i + \Delta t_i) = \begin{cases} \hat{\nu}(t_i) + \sqrt{\tilde{g}(\hat{\nu}(t_i))\Delta t_i} \hat{\xi}_i^G & (\hat{\nu}(t_i) \geq 0) \\ 0 & (\hat{\nu}(t_i) < 0) \end{cases} \quad (\text{I7})$$

with independent Gaussian random number  $\hat{\xi}_i^G$ . For Fig. 11, we employ the following model:

$$\tilde{g}(\hat{\nu}) = \lambda_0 e^{\beta \hat{\nu}}, \quad \Delta t_i = \min \left\{ \Delta t_{\max}^{(1)}, \frac{\Delta t_{\max}^{(2)}}{\tilde{g}(\hat{\nu}(t_i))} \right\} \quad (\text{I8})$$

with  $\lambda_0 = 10^{-4}$ ,  $\beta = 3$ ,  $\Delta t_{\max}^{(1)} = 0.1$ , and  $\Delta t_{\max}^{(2)} = 0.01$ . The total time of the simulation was  $T = 5 \times 10^5$  and the total number of parallel threads was  $N_{\text{PC}} = 4$ .

- [1] H. Scher and E. W. Montroll, Phys. Rev. B **12**, 2455 (1975).
- [2] H. Scher, H. G. Margolin, R. Metzler, J. Klafter, and B. Berkowitz, Geophys. Res. Lett. **29**, 5 (2002).
- [3] Y. Ogata, J. Am. stat. Assoc. **83**, 9 (1988).
- [4] Y. Ogata, Pure Appl. Geophys. **155**, 471 (1999).
- [5] A. Helmstetter and D. Sornette, J. Geophys. Res. **107** (B10), 2237 (2002).
- [6] S. Nandan, G. Ouillon, D. Sornette, and S. Wiemer, Seismological Research Letters **90** (4), 1650 (2019).
- [7] M. Feng, S.-M. Cai, M. Tang, and Y.-C. Lai, Nat. Commun. **10**, 3748 (2019).
- [8] E Errais, K Giesecke, LR Goldberg, SIAM J. Fin. Math. **1**, 642 (2010).
- [9] A. Chakraborti, I.M. Toke, M. Patriarca and F. Abergel, Quantitative Finance **11**, 991 (2011).
- [10] Z.-Q. Jiang, W.-J. Xie, W.-X. Zhou and D. Sornette, Reports on Progress in Physics **82**, 125901 (105pp) (2019).
- [11] D. Sornette, F. Deschattres, T. Gilbert, and Y. Ageon, Phys. Rev. Letts. **93** (22), 228701 (2004).
- [12] R. Crane and D. Sornette, Proc. Nat. Acad. Sci. USA **105** (41), 15649 (2008).
- [13] A. Hawkes, Journal of the Royal Statistical Society. Series B (Methodological) **33** (3), 438 (1971).
- [14] A. Hawkes, Biometrika **58** (1), 83 (1971).
- [15] A. Hawkes and D. Oakes, J. Appl. Prob. **11** (3), 493 (1974).
- [16] K. Kanazawa and D. Sornette, Phys. Rev. Lett. **125**, 138301 (2020).
- [17] K. Kanazawa and D. Sornette, Phys. Rev. Research **2**, 033442 (2020).
- [18] G.I. Barenblatt, *Scaling, self-similarity, and intermediate asymptotics* (Cambridge University Press, Cambridge, UK, 1996).
- [19] P. Brémaud and L. Massoulié, Annals of Probability, **24**, 1563 (1996).
- [20] J.-P. Bouchaud, J. Bonart, J. Donier, and M. Gould, *Trades, Quotes and Prices* (Cambridge University Press, Cambridge, UK, 2018).
- [21] D. Sornette and G. Ouillon, Phys. Rev. Lett. **94**, 038501 (2005).
- [22] G. Ouillon and D. Sornette, J. Geophys. Res. **110**, B04306 (2005).
- [23] S. Nandan, G. Ouillon, J. Woessner, D. Sornette and S. Wiemer, Journal of Geophysical Research - Solid Earth **121**, 1890-1909, doi:10.1002/2015JB012212 (2016).
- [24] D. Plenz and E. Niebur, eds., *Criticality in neural systems* (John Wiley & Sons, New York, 2014).
- [25] I. Osorio, M.G. Frei, D. Sornette, J. Milton and Y.-C. Lai, Epileptic Seizures, quakes of the brain? Phys. Rev. E **82** (2), 021919 (2010).
- [26] D. Sornette and I. Osorio, Prediction, chapter in “Epilepsy: The Intersection of Neurosciences, Biology, Mathematics, Physics and Engineering”, Editors: Osorio I., Zaveri H.P., Frei M.G., Arthurs S., CRC Press, Taylor & Francis Group, pp. 203-237 (2010).
- [27] K. Kanazawa and D. Sornette, Phys. Rev. Lett. **127**, 188301 (2021).
- [28] P. Blanc, J. Donier, and J.-P. Bouchaud, Quantitative Finance **17**, 171 (2017).

- [29] F. Gao and L. Zhu, Stochastic Processes and their Applications **128**, 4051 (2018).
- [30] S Ciliberto, A Guarino, R Scorreretti, Physica D: Nonlinear Phenomena **158** (1-4), 83-104 (2001).
- [31] R Scorreretti, S Ciliberto, A Guarino, Europhysics Letters **55** (5), 626 (2001).
- [32] A. Saichev and D. Sornette, Phys. Rev. E **71**, 016608 (2005).
- [33] A. Helmstetter and D. Sornette, J. Geophys. Res., **108**, 2482, 10.1029/2003JB002485 (2003).
- [34] C.G. Bowsher, Journal of Econometrics **141**(2), 876-912 (2007).
- [35] V.A. Filimonov and D. Sornette, Phys. Rev. E, **85**, 056108 (2012).
- [36] V.A. Filimonov and D. Sornette, Europhysics Letters **94**, 46003 (2011).
- [37] G. Soros, *The Alchemy of Finance*, Simon & Schuster (1988).
- [38] E. Bacry, J. Delour and J.-F. Muzy, Phys. Rev. E **64** (2), 026103 (2001).
- [39] A. Saichev, A. and D. Sornette, Phys. Rev. E **74** (1), 011111+ (2006).
- [40] A. Saichev, A. and V. Filimonov, J. Exp. Theor. Phys. **107** (2), 324–330 (2008).
- [41] E. Bacry, I. Mastromatteo and J.-F. Muzy, Market Microstructure and Liquidity **1** (1), 1550005 (2015).
- [42] C.W. Gardiner, *Handbook of Stochastic Methods*, 4th ed. (Springer, Berlin, 2009).
- [43] N. G. Van Kampen, *Stochastic Processes in Physics and Chemistry* (Elsevier, New York, 1992)
- [44] G. Ouillon and D. Sornette, Future Generation Computer Systems **22** (4), 500-520 (2006).
- [45] K. Kanazawa, *Statistical Mechanics for Athermal Fluctuation: Non-Gaussian Noise in Physics* (Springer, Berlin, 2017).
- [46] K. Kanazawa, T.G. Sano, T. Sagawa, and H. Hayakawa, Phys. Rev. Lett. **114**, 090601 (2015).
- [47] K. Kanazawa, T.G. Sano, T. Sagawa, and H. Hayakawa, J. Stat. Phys. **160**, 1294 (2015).
- [48] X. Gao and L. Zhu, Queueing Syst. **90**, 161 (2018).
- [49] F. Gerhard, M. Deger, and W. Truccolo, PLOS Computational Biology **13**(2), e1005390 (2017).
- [50] C. Van den Broeck, J. Stat. Phys. **31**, 467 (1983).
- [51] H. Kesten, Acta Math. **131**, 207-248 (1973).
- [52] D. Sornette, Physica A **250**, 295 (1998).
- [53] D. Sornette and R. Cont, J. Phys. I France **7**, 431 (1997).
- [54] A.S. Cherny and H.-J. Engelbert, *Singular Stochastic Differential Equations*, Lecture Notes in Mathematics 1858, pp.1-128, Springer (2005).
- [55] J. Krafter and I.M. Sokolov, *First Steps in Random Walks: From Tools to Applications* (Oxford University Press, Oxford, UK, 2011).

TABLE I: Summary of the results obtained in the present work, for both one-sided and two-sided mark distributions. The obtained steady-state intensity distributions of intensities are systematically classified for various NLHawkes processes. FAI and MSA stand for fast-accelerating intensity ( $g(\nu) \gg \nu^2$ ) and multifractal stress activation model ( $g(\nu) \propto e^{\beta\nu}$ ,  $\beta > 0$ ). In this report, we assume that the moment-generating function  $\Phi(x) := \int_{-\infty}^{\infty} \rho(y)(e^{xy} - 1)dy$  exists and  $c^*$  is the positive root of  $\Phi(c^*) = 0$  for  $m < 0$  or is equal to zero for  $m = 0$ . In addition, we define  $m := \int_{-\infty}^{\infty} y\rho(y)dy$ ,  $p_+ := \int_0^{\infty} \rho(y)dy$ , and  $p_- := \int_{-\infty}^0 \rho(y)dy$ .

Model	Mark PDF $\rho(y)$	Tension-intensity map $g(\nu)$	Critical?	PDF $P_{ss}(\lambda)$	Exponent $a$	Section for $h(t) = (\eta/\tau)e^{-t/\tau}$	Section for general $h(t)$
Linear	One sided: $\rho(y) = 0$ for $y < 0$	$\nu_0 + \nu_1$ , $\nu_0 > 0$	Yes ( $\eta \uparrow 1$ )	$\propto \lambda^{-1-a} e^{-\lambda/\lambda_{cut}}$ ( $\lim_{\eta \uparrow 1} \nu_{cut} = \infty$ )	$a < 0$ (intermediate asymptotics, non-universal)	Sec. IV	Sec. VII
	Ramp	$\max\{\nu_0, \nu - \nu_1\}$ , $\nu_0 > 0$			any-real number $a$ (non-universal)		
Ramp	Two-sided, symmetric: $\rho(y) = \rho(-y)$	$\max\{\nu_0, \nu - \nu_1\}$ , $\nu_0 > 0$	No	$\propto e^{-\lambda/\lambda_{cut}}$ ( $\lambda_{cut} < \infty$ )	Absent	Sec. V	Missing
	Quadratic	$\propto \nu^2$		$\propto \lambda^{-1-a}$	$a > 1/2$ (non-universal)		
Polynomial of order $n > 2$ (FAI)	$\propto \nu^n$ , $n > 2$				$a = 1 - 1/n$ (universal)	Sec. VIII	
	Exponential (FAI, MSA)	$\propto e^{\beta\nu}$			$a = 1$ (universal, Zipf)		
Gaussian (FAI)	$\propto e^{\beta\nu^2}$						
	FAI	$\gg \nu^2$		$\propto \lambda^{-1} \left[ \frac{dg(\nu)}{d\nu} \right]_{\nu=g^{-1}(\lambda)}^{-1}$			
Ramp	Two-sided, non positive mean: $p_+ > 0$ , $p_- > 0$ , $m \leq 0$	$\max\{\nu_0, \nu - \nu_1\}$ , $\nu_0 > 0$	No	$\propto e^{-\lambda/\lambda_{cut}}$ ( $\lambda_{cut} < \infty$ )	Absent	Sec. VI	Missing
	Exponential (FAI, MSA)	$\propto e^{\beta\nu}$		$\propto \lambda^{-1-a}$	$a \geq 1$ (non-universal)	Sec. VIII	
FAI		$\gg \nu^2$		$\propto \lambda^{-1} \left[ e^{-u\nu} \left  \frac{dg(\nu)}{d\nu} \right ^{-1} \right]_{\nu=g^{-1}(\lambda)}^{-1}$	$u := c^*/h(0)$		