A Multifractal Wavelet Model with Application to Network Traffic*

Rudolf H. Riedi, *Member, IEEE,* Matthew S. Crouse, *Student Member, IEEE,* Vinay J. Ribeiro, *Student Member, IEEE,* and Richard G. Baraniuk *Member, IEEE*†

Abstract— **In this paper, we develop a new multiscale modeling framework for characterizing positive-valued data with long-range-dependent correlations ($1/f$ noise). Using the Haar wavelet transform and a special multiplicative structure on the wavelet and scaling coefficients to ensure positive results, the model provides a rapid $O(N)$ cascade algorithm for synthesizing N-point data sets. We study both the second-order and multifractal properties of the model, the latter after a tutorial overview of multifractal analysis. We derive a scheme for matching the model to real data observations and, to demonstrate its effectiveness, apply the model to network traffic synthesis. The flexibility and accuracy of the model and fitting procedure result in a close fit to the real data statistics (variance-time plots and moment scaling) and queuing behavior. Although for illustrative purposes we focus on applications in network traffic modeling, the multifractal wavelet model could be useful in a number of other areas involving positive data, including image processing, finance, and geophysics.**

Index Terms—**Long-range dependence, multifractals, network traffic positive $1/f$ noise, wavelets.**

1 Introduction

1.1 Fractal signal models

The discovery of the *fractal*, *self-similar*, or $1/f$ nature of many phenomena has led to exciting breakthroughs in a variety of scientific disciplines, including physics, chemistry, astronomy, biology, meteorology, hydrology, and soil science [1, 2]. In signal and image processing, fractals have been applied in fields such as computer graphics, texture modeling, image compression, and pattern recognition [3, 4].

Fractal models have made a major impact in the area of communications recently, particularly in the area of computer data networks. As the work of Leland et al. [5] and subsequent studies have demonstrated, network traffic loads exhibit fractal properties such as self-similarity, burstiness, and *long-range dependence* (LRD). Inadequately modeled by classical Poisson or Markov models, these properties strongly influence network performance [5]. For instance, performance predictions based on classical traffic models are often far too optimistic when compared against actual performance with real data. Fractal traffic models have provided exciting new insights into network behavior and promise new algorithms for network data prediction and control.

The *fractional Brownian motion* (fBm) $B(t)$ has been the most broadly applied fractal signal model [5–7]. Its power lies in its simplicity: fBm is statistically self-similar[1]

$$B(at) \stackrel{fd}{=} a^H B(t). \tag{1}$$

Thus, while it has rich statistical properties, it remains amenable to a tractable analysis. The fBm is not stationary, but its increments form the stationary *fractional Gaussian noise* (fGn) process. When the *Hurst parameter* $H > 1/2$, fGn exhibits LRD.

N samples of fGn can be simulated exactly via direct Cholesky factorization ($O(N^3)$ computational complexity) [4] or Levinson's recursion ($O(N^2)$ complexity) [8]. These costs can become overbearing, especially in networking applications where often $N \gg 10^6$. For such large problems, approximate synthesis techniques ($O(N)$ complexity) based on wavelets have been developed.

The discrete wavelet transform represents a 1-D real signal $X(t)$ in terms of shifted and dilated versions of a prototype bandpass wavelet function $\psi(t)$ and shifted versions of a low-pass scaling function $\phi(t)$ [9, 10]. For special choices of the wavelet and scaling functions, the atoms

$$\psi_{j,k}(t) \;:=\; 2^{j/2}\,\psi\Big(2^j t - k\Big), \tag{2}$$

$$\phi_{j,k}(t) \;:=\; 2^{j/2}\,\phi\Big(2^j t - k\Big), \quad j,k \in \mathbb{Z} \tag{3}$$

form an orthonormal basis, and we have the signal representation [9, 10]

$$X(t) = \sum_k U_{J_0,k}\,\phi_{J_0,k}(t) \;+\; \sum_{j=J_0}^{\infty}\sum_k W_{j,k}\,\psi_{j,k}(t), \tag{4}$$

with[2]

$$W_{j,k} \;:=\; \int X(t)\,\psi_{j,k}(t)\,\mathrm{d}t, \tag{5}$$

$$U_{j,k} \;:=\; \int X(t)\,\phi_{j,k}(t)\,\mathrm{d}t. \tag{6}$$

*Manuscript received February 1998; revised September 1998. This work was supported by the National Science Foundation under Grant MIP–9457438, the Office of Naval Research under Grant N00014–95–1–0849, DARPA/AFOSR under Grant F49620-97-1-0513, and Texas Instruments. The material of this paper was presented in part at the IEEE–SP International Symposium on Time–Frequency and Time–Scale Analysis, Pittsburgh, PA, October 1998.

†The authors are with the Department of Electrical and Computer Engineering, Rice University, Houston, TX 77251-1892 USA (e–mail: riedi@rice.edu, mcrouse@ece.rice.edu, vinay@ece.rice.edu, richb@ece.rice.edu).

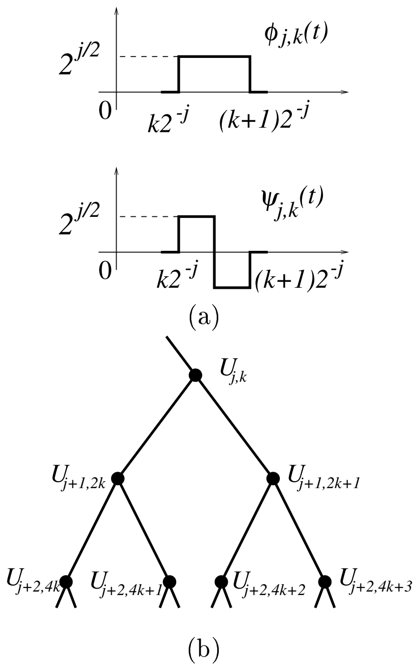


Figure 1: (a) The Haar scaling and wavelet functions $\phi_{j,k}(t)$ and $\psi_{j,k}(t)$. (b) Binary tree of scaling coefficients from coarse to fine scales.

For a wavelet $\psi(t)$ centered at time zero and frequency f_0, the *wavelet coefficient* $W_{j,k}$ measures the signal content around time $2^{-j}k$ and frequency $2^j f_0$. The *scaling coefficient* $U_{j,k}$ measures the local mean around time $2^{-j}k$. In the wavelet transform, j indexes the *scale* of analysis: J_0 indicates the coarsest scale or lowest resolution of analysis, and larger j correspond to higher resolutions of the analysis.

The Haar scaling and wavelet functions (see Figure 1(a)) provide the simplest example of an orthonormal wavelet basis. Because of (3), the supports of the fine-scale scaling functions nest inside the supports of those at coarser scales; this can be neatly represented by the binary tree structure of Figure 1(b). Row (scale) j of this scaling coefficient tree contains an approximation to $X(t)$ of resolution 2^{-j}. Row j of the complementary wavelet coefficient tree (not shown) contains the details in scale $j+1$ of the scaling coefficient tree that are suppressed in scale j. In fact, the $U_{j+1,k}$ consist simply of scaled sums and differences of the $U_{j,k}$ and $W_{j,k}$.

The wavelet transform closely approximates the Karhunen-Loève transform for fBm and fGn [11–13]. This fact has been leveraged into efficient approximate fBm and fGn models [14]: we posit that the wavelet coefficients $W_{j,k}$ are simply independent, zero-mean Gaussian random variables with power-law decaying variance $\mathrm{var}(W_{j,k}) \propto 2^{-j\gamma}$, with $\gamma = 2H+1$ for fBm and $\gamma = 2H-1$ for fGn.

Unfortunately, despite their great simplicity, fractal models such as fBm and fGn have significant limitations for modeling certain types of natural and man-made processes. First, fBm and fGn are Gaussian models, whereas many LRD processes, including network traffic, turbulence, financial data, and images, are inherently *positive* and often *spiky*. Both of these qualities are explicitly nonGaussian. Second, many signals exhibit LRD but display short-term correlations and scaling behavior inconsistent with the strict self-similarity of (1).

1.2 A Multifractal Wavelet Model (MWM)

In this paper, we develop a new wavelet-based signal model for positive, stationary, LRD data. While characterizing positive data in the wavelet domain is problematic for general wavelets, for the Haar wavelet, we have the simple condition: $X(t)$ is positive if and only if $|W_{j,k}| \leq U_{j,k}$ for all j,k.

In the *multifractal wavelet model* (MWM), we ensure a positive signal output by modeling the wavelet coefficients as $W_{j,k} = A_{j,k}U_{j,k}$, with the multipliers $A_{j,k}$ independent random variables supported on $[-1,1]$. For simplicity, we choose β (beta) and simple point mass distributions for the multipliers.

The MWM flows as a multiscale, coarse-to-fine synthesis down the tree in Figure 1(b): Given the approximation to $X(t)$ at resolution 2^{-j} (the $U_{j,k}$), we compute the wavelet coefficients $W_{j,k} = A_{j,k}U_{j,k}$ with random $A_{j,k}$. The approximation to $X(t)$ at resolution $2^{-(j+1)}$ (the $U_{j+1,k}$) is then obtained from scaled sums and differences of the $U_{j,k}$ and $W_{j,k}$. This process can be iterated until any desired resolution/signal-length is reached; the total cost is a meager $O(N)$ operations for an N-point output.

Like fGn models, the MWM can closely model the power spectrum, and hence the LRD, of a set of training data if the variances of the multipliers $A_{j,k}$ are chosen appropriately. Unlike fGn models, the MWM can also match positivity and higher-order statistics due to its multiplicative construction.

For example, Figure 2 compares real data (Bellcore Ethernet packet interarrival data, August 1989) with synthetic MWM and fGn data, at different aggregation levels. Both models match the mean, variance, and correlation decay of the real data. Evident from the figure are the large number of (unacceptable) negative values of fGn, caused by the real data having a high standard-deviation-to-mean ratio. The MWM data much more closely matches the characteristics of the real data. Moreover, a length-2^{18} MWM synthesis required just eight seconds of workstation run time, in contrast to eighteen hours for a Levinson fGn synthesis.

1.3 Cascades and multifractals

The multiplicative construction of the MWM process is reminiscent of the *binomial measure*, a classical *multifractal* process. Multifractals were first introduced to model dissipation of energy in turbulence [15, 16] and have proved well-suited to modeling non-homogeneous phenomena [17, 18]. More recently, the multifractal nature of network traffic has been demonstrated convincingly, first in [19] and subsequently

[1] The equality is in the sense of finite-dimensional distributions.

[2] We consider the signal $X(t)$ to be random and so use capital letters for all quantities derived from it.

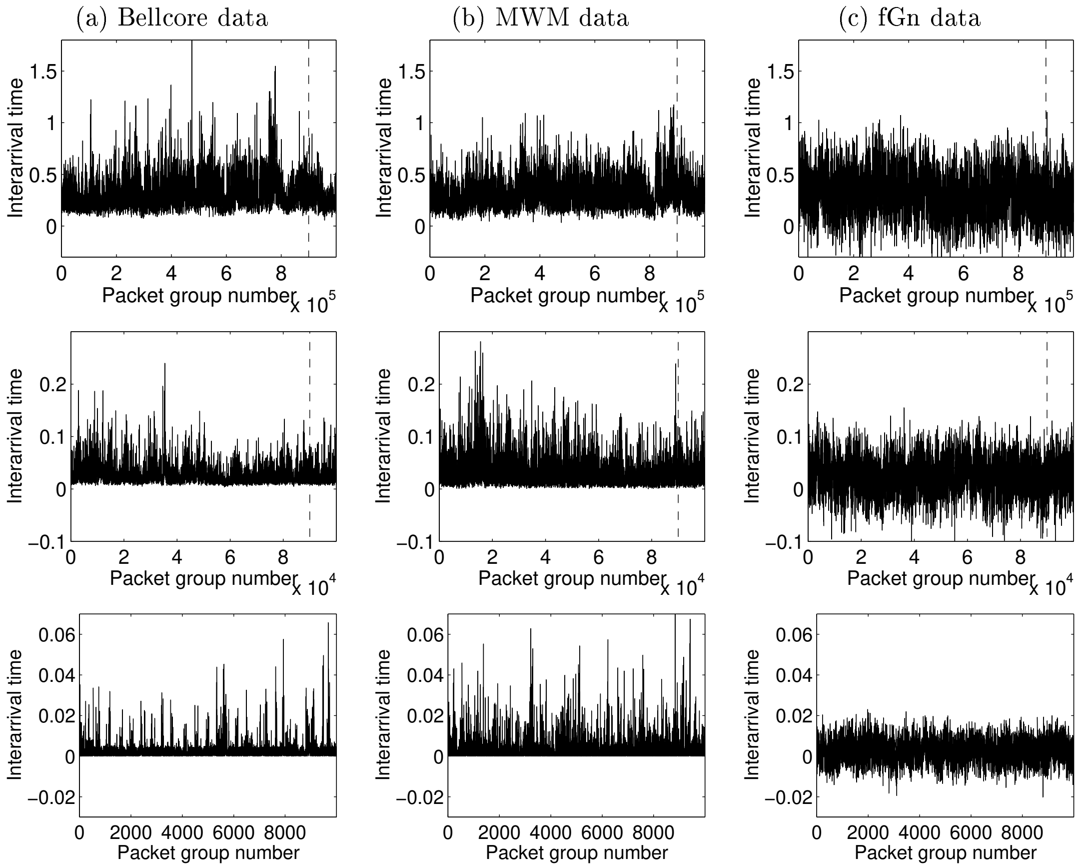


Figure 2: Interarrival times of groups of packets of (a) Bellcore August 1989 *pAug* data [5], (b) one realization of the multifractal wavelet model (MWM) synthesis, and (c) one realization of fGn synthesis. The top, middle, and bottom plots correspond to interarrivals of one-hundred packets, ten packets, and one packet, respectively. The ten-packet and one-packet plots correspond to the last tenth of the data from the one-hundred-packet and ten-packet plots, respectively, as indicated by the vertical dotted lines. Approximately 30% of the fGn values are negative.

in [20,21]. The beauty of the multifractal formalism has motivated considerable research effort in mathematics [22–32]; however, few multifractal data models have been developed to date.

In the most simple terms, multifractals possess a local smoothness H_t that depends on t in an erratic way. Equivalently, multifractals have moments that scale non-linearly. By matching the multifractal properties of training data, the MWM can capture and synthesize rare events in addition to global behavior. Random products are "usually" small but "sometimes" extremely large. This results in the burstiness seen in Figure 2(b). Models based on fBm/fGn, on the other hand, exhibit a non-varying behavior in both H_t and moments — they are "monofractal."

With regards to network traffic, self-similar, additive schemes model traffic arrivals as a mean rate with superimposed fGn fluctuations. This agrees with the conception of traffic as the superposition of individual components and is accurate on large time scales. Multiplicative models, on the other hand, represent traffic arrivals as the product of random multipliers, which mimicks the partitioning of total traffic throughput into parts. This point-of-view is appealing when considering small time scales [33].

1.4 Organization

After some background on fractals and wavelets in Section 2, we provide the construction and basic properties of the MWM in Section 3. In Section 4, we develop the modeling framework and provide a procedure for fitting the MWM to actual data measurements. Section 5 reviews multiplicative cascades and reveals the relationship between the MWM and the binomial cascade. We give a brief introduction to multifractal analysis (MFA), relate the MFA to wavelets and LRD, and perform an MFA of the MWM in Section 6. To illustrate the effectiveness of the MWM, in Section 7, we employ it to generate high-quality synthetic network traffic data. We confirm the accuracy of the synthesis in terms of both statistical measures and queuing behavior and comment on possible physical reasons for the presence of multiplicative processes in network traffic. We close with a discussion and conclusions in Section 8. In Appendix A, we give a tutorial review of the MFA. The proof of the multifractal formalism for the MWM appears in Appendix B.

2 Fractals, Scaling, and Wavelets

Fractals are geometric objects exhibiting an intricate, highly irregular appearance on all resolutions [34]. The *fractal dimension* $\dim(E)$ [35] measures the degree of irregularity or roughness of a set E. Here, we are mainly interested in fractal signals, i.e. signals having a fractal graph. Most known fractals are *self-similar;* if we "zoom" (in or out) of the fractal, we obtain a picture similar to the original. In a deterministic setting, this imposes strong restrictions on the fractal, and the easiest way to obtain such an object is to apply a simple geometrical rule iteratively to obtain details up to infinitely fine resolution. Consequently, deterministic fractals consist of highly repetitive patterns. Real-world phenomena can rarely be described using such simple models. Nevertheless, "similarity on all scales" sometimes holds in a statistical sense, leading to the notion of random fractals.

2.1 Fractional Brownian motion and fractional Gaussian noise

For processes, the notion of "similarity on all scales" can be made precise in various ways. A very strict one is that of *self-similar with stationary increments:* A process Y is H-sssi if it has stationary increments and for all $a > 0$ (cf. (1))

$$Y(at) \stackrel{fd}{=} a^H Y(t) \tag{7}$$

The pre-eminent random fractal signal model at present is the *fractional Brownian motion* (fBm) $B(t)$. This process is uniquely defined through two properties: H-sssi and Gaussianity [7,36]. The Hurst parameter lies in the range $0 < H < 1$; smaller H corresponds to fBm's with "wilder" or rougher-looking local behavior.

Although fBm is useful for theoretical analysis, its increments process (for finite increment Δt)

$$G[n] := B(n\Delta t) - B\left((n-1)\Delta t\right), \tag{8}$$

known as *fractional Gaussian noise* (fGn), is often more useful in practice. While fBm is nonstationary, fGn is stationary.

For fBm, self-similarity (7) is equivalent to its autocorrelation function $r_B(t,s) := \mathbb{E}\left[B(t)B(s)\right]$ having the form

$$r_B(t,s) = \frac{\sigma^2}{2}\left(|t|^{2H} + |s|^{2H} - |t-s|^{2H}\right) \tag{9}$$

or its (generalized) power spectral density behaving as $\Gamma_B(f) \propto |f|^{-(2H+1)}$ [12]. It follows from (9) that fGn has an autocorrelation function

$$r_G[\tau] = \frac{\sigma^2}{2}\left|\Delta t\right|^{2H}\left(|\tau+1|^{2H} + |\tau-1|^{2H} - 2|\tau|^{2H}\right). \tag{10}$$

As with fBm, fGn has a discrete-time power spectrum that behaves as $\Gamma_G(f) \propto |f|^{-(2H-1)}$ for f near 0. Thus fBm and fGn are often called $1/f$ *noise.*

2.2 Long-range dependence

While the rigid correlation structure of fGn is somewhat restrictive for modeling purposes, the tail decay of $r_G[\tau]$ has proven to be of importance in itself. In particular, it inspires weaker notions of "similarity on all scales" in terms of second-order statistics only.

It is easy to see that (10) decays like $r_G[\tau] \simeq \tau^{2H-2}$. For $1/2 < H < 1$, the correlation is strictly positive and decays so slowly that it is non-summable. A process Z with this property ($\sum_\tau r_Z[\tau] = \infty$) is said to exhibit *long range dependence* (LRD), since it possesses strong correlations at large lags. LRD can be equivalently characterized in terms of the behavior of the aggregated processes

$$Z^{(m)}[n] := \frac{1}{m}\sum_{i=(k-1)m+1}^{km} Z[i]. \tag{11}$$

The fGn with $1/2 < H < 1$ has proven useful for signal modeling, because it has LRD yet permits tractable theoretical analysis due to (7). In particular, the H-sssi property (7) together with (8) imply that

$$G[n] \stackrel{fd}{=} m^{1-H} G^{(m)}[n]. \tag{12}$$

Processes for which $\mathrm{var}(Z[n]) = m^{2-2H}\mathrm{var}(Z^{(m)}[n])$ are termed *second-order self-similar processes* [2]. For such processes, a log-log plot of the variance of $Z^{(m)}[n]$ as a function of m — the *variance-time plot* — is strictly linear with a slope of $2-2H$ [5]. The variance-time plot can be used to detect the self-similarity and LRD of a trace and can be applied to nonGaussian, non-zero-mean data as well.[3]

2.3 Wavelets and $1/f$ processes

The inherent scaling property of the wavelet basis is well-suited for analyzing self-similar processes. Wavelets serve as an approximate Karhunen-Loève transform for $1/f$ processes [11], including fBm [12] and fGn [13]. These highly correlated, LRD signals become nearly uncorrelated in the wavelet domain. This property has lead to the widespread use of wavelets for the analysis and synthesis of fractal and LRD signals [14].

In particular, the energy of the wavelet coefficients of a continuous fBm exhibits a power-law decay with scale [12]. The variance progression of the wavelet transform of sampled fBm and fGn does not follow a strict power-law, but rather includes scale-dependent factors [12,13]. Kaplan and Kuo [13] have shown that for the Haar wavelet, the variance progression of the wavelet transform of fGn satisfies

$$\mathrm{var}(W_{j,k}) \propto 2^{-j(2H-1)}. \tag{13}$$

Moreover, the wavelet coefficients of fGn are typically much less correlated than those of the underlying sampled fBm process. They use these facts to develop a robust wavelet-based estimator for the H of an fGn submerged in additive white Gaussian noise. Similar wavelet-based estimators for H compare favorably with standard estimation techniques [37]

[3] Although the Hurst parameter H is sometimes used strictly in the context of fGn, we will view H as a variance-time plot parameter to characterize LRD processes in general.

and have been applied to practical problems such as network traffic analysis [14].

Wavelets can also be used to synthesize approximate $1/f$ processes with generalized spectra of the form $\Gamma(f) \propto |f|^{-\gamma}$, $0 < \gamma < 2$, which includes fBm and fGn.[4] Playing off the Karhunen-Loève property of the wavelet transform, Wornell generates zero-mean, independent Gaussian random variables $W_{j,k}$ with power scaling according to [11]

$$\operatorname{var}(W_{j,k}) \propto 2^{-j\gamma}. \tag{14}$$

He then inverts the wavelet transform to obtain the synthesized process. Even though the mean and variance of the synthesized signal are stationary, this approach generally results in a non-stationary Gaussian process with time-varying correlation function (see Section 3.4). However, the *time-averaged* correlation and spectrum do approximate that of a $1/f$ process [11]. Though only approximate, this method's $O(N)$ computational cost compares favorably with the $O(N^2)$ cost of the Levinson algorithm for exact synthesis [8] and the $O(N^3)$ cost of direct Cholesky factorization [4].

2.4 Moving beyond fBm

Although fBm and fGn are powerful and tractable signal models, their strict self-similarity is too restrictive to adequately characterize many types of signals [19, 38]. For instance:

1. many signals possess significant LRD, but display short-term correlations and scaling behavior inconsistent with strict self-similarity;
2. in many signals, the scaling behavior of moments as the signal is aggregated is a non-trivial (nonlinear) function of the moment order; and
3. many signals have increments that are inherently positive and hence nonGaussian.

Signals with these properties fall naturally into the class of *multifractal processes*. Multifractal signal models are positive measures or distributions possessing self-similarity but non-homogeneous scaling. The goal of this paper is a multifractal extension of traditional fBm and fGn signal models suitable for analyzing, characterizing, and synthesizing positive processes with LRD. As with fractals, we will find the wavelet transform useful for constructing and analyzing our model.

3 A Multifractal Wavelet Model

The primary goal of this paper is to develop a wavelet-domain model for a positive, stationary, LRD signal $C(t)$ and its integral $D(t)$. (The integral will be more convenient for the analysis in Section 5).

In practice, we will work with a discrete-time signal $C^{(n)}[k]$ that approximates $C(t)$ at resolution 2^{-n}. To reflect this in the wavelet transform, we replace the semi-infinite sum in (4) with a sum over the finite number of scales $0 \le j < n$, $j, n \in \mathbb{Z}_+$. Here we also set, without loss of generality, the coarsest scale $J_0 = 0$, meaning that the first sum in (4) reduces to the single term $U_{0,0}\,\phi_{0,0}$. This corresponds to a single scaling coefficient tree approximating $C(t)$ on the interval $[0, 1]$. While we will emphasize this case in the sequel, in certain cases (as in Section 4.4 below), we will find it convenient to employ a forest of R trees rooted at R scaling coefficients $U_{0,k}$, $k = 0, 1, \ldots, R-1$. In this case, the process $C(t)$ is assumed to lie in the interval $[0, R]$.

Using the Haar wavelet, the discrete process $C^{(n)}[k]$ takes values that correspond to the integral of $C(t)$ in the interval $[k2^{-n}, (k+1)2^{-n}[$. Such processes have a natural interpretation as an increment process:

$$\begin{aligned} C^{(n)}[k] \;&:=\; D((k+1)2^{-n}) - D(k2^{-n}) \\ &=\; \int_{k2^{-n}}^{(k+1)2^{-n}} C(t)\,\mathrm{d}t \;=\; 2^{-n/2} U_{n,k} \end{aligned} \tag{15}$$

for $k = 0, \ldots, 2^{-n} - 1$. Equation (15) is similar to (8) with $\Delta t = 2^{-n}$.

To be useful in real applications, our model must be simple, produce a fast analysis and synthesis, and closely match the process's positive, nonGaussian marginals and its LRD. We will now show how this is possible using a simple Haar wavelet construction of the increments process $C^{(n)}[k]$.

3.1 Positivity through multiplication

Wavelet-domain modeling of positive processes is complicated by the fact that the wavelet coefficient constraints required to ensure a positive output are nontrivial. Quite the contrary for the Haar wavelet, however. For the Haar wavelet, the scaling and wavelet transform coefficients can be recursively computed using

$$U_{j,k} \;=\; 2^{-1/2}(U_{j+1,2k} + U_{j+1,2k+1}) \quad \text{and} \tag{16}$$

$$W_{j,k} \;=\; 2^{-1/2}(U_{j+1,2k} - U_{j+1,2k+1}). \tag{17}$$

Furthermore, in the Haar transform of positive data, we know that all $U_{j,k} \ge 0$, since each $U_{j,k}$ equals a scaled local mean. Rearranging (16) and (17) to

$$\begin{aligned} U_{j+1,2k} \;&=\; 2^{-1/2}(U_{j,k} + W_{j,k}) \quad \text{and} \\ U_{j+1,2k+1} \;&=\; 2^{-1/2}(U_{j,k} - W_{j,k}), \end{aligned} \tag{18}$$

we thus find a simple constraint to guarantee that the process is positive:

$$|W_{j,k}| \le U_{j,k}. \tag{19}$$

Although we have derived (19) as a necessary condition, it is easy to see that it is also sufficient. For more general wavelet systems (with longer, overlapping wavelets), the conditions are considerably more complex.

We wish to build a statistical model for the $W_{j,k}$'s that automatically incorporates (19). This leads us to a simple *multiplicative* signal model. Let $A_{j,k}$ be a random variable

[4] Processes corresponding to a wider range of γ's can also be synthesized, using wavelets with regularity greater than two [12].

supported on the interval $[-1,1]$ and define the wavelet coefficients by

$$W_{j,k} = A_{j,k}\, U_{j,k}. \tag{20}$$

In Section 3.4.1 we will place some additional constraints on the $A_{j,k}$.

The *multifractal wavelet model* (MWM) consists of the Haar wavelet transform and the structure constraint (20).

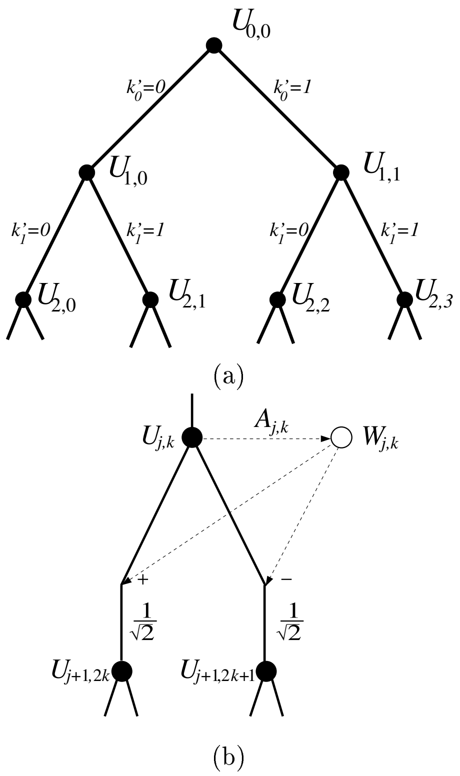


Figure 3: (a) More detailed tree structure of scaling coefficients. (b) MWM construction: At scale j, we form the wavelet coefficient as the product $W_{j,k} = A_{j,k}U_{j,k}$, with $A_{j,k}$ a random variable distributed in $[-1,1]$. Then, at scale $j+1$, we form the scaling coefficients $U_{j+1,2k}$ and $U_{j+1,2k+1}$ as sums and differences of $U_{j,k}$ and $W_{j,k}$ (normalized by $1/\sqrt{2}$).

3.2 Synthesis procedure

The MWM can be interpreted as a simple coarse-to-fine synthesis running as follows (see Figure 3):

1. Set $j=0$. Fix or compute the coarsest (root) scaling coefficient $U_{0,0}$ (modeling of $U_{0,0}$ is discussed in Section 4.4).
2. At scale j, generate the random multipliers $A_{j,k}$ and calculate each $W_{j,k}$ via (20) for $k = 0,\ldots,2^j-1$.
3. At scale j, use $U_{j,k}$ and $W_{j,k}$ in (18) to calculate $U_{j+1,2k}$ and $U_{j+1,2k+1}$, the scaling coefficients at scale $j+1$ for $k=0,\ldots,2^j-1$.
4. Iterate steps 2 and 3, replacing j by $j+1$ until the finest scale $j=n$ is reached.

Since we generate the scaling coefficients simultaneously with the wavelet coefficients, there is no need to invert the wavelet transform. The finest-scale scaling coefficients are in fact the MWM output process, i.e. $C^{(n)}[k] = 2^{-n/2}U_{n,k}$, $k=0,\ldots,2^n-1$. The total cost for computing N MWM signal samples is $O(N)$.

Because of the simple structure of the Haar transform, Steps 2 and 3 above can be combined, eliminating the wavelet coefficients altogether:

$$\begin{aligned} U_{j+1,2k} &= \left(\frac{1+A_{j,k}}{\sqrt{2}}\right) U_{j,k} \quad \text{and} \\ U_{j+1,2k+1} &= \left(\frac{1-A_{j,k}}{\sqrt{2}}\right) U_{j,k}. \end{aligned} \tag{21}$$

3.3 Closed-form coefficient expressions

Because of its simplicity, we can easily obtain explicit formulas for the MWM's fine-scale Haar wavelet and scaling coefficients in terms of the scaling coefficients and multipliers at coarser scales. We begin by defining an indexing scheme to relate the coarsest-scale scaling coefficient $U_{0,0}$ to its "descendants" at finer scales, the scaling coefficients $U_{j,k}$, $j>0$ (see Figure 3(a)). Let k_j, $j>0$, be the variable indexing the possible shifts of the descendants of $U_{0,0}$ at scale j. We can relate the shift k_j of a scaling coefficient to the shift of one of its two direct descendants (children) k_{j+1} via $k_{j+1} = 2k_j + k'_j$, with $k'_j = 0$ corresponding to the left descendant and $k'_j = 1$ the right descendant (see Figure 3(a)). From this we can express k_j as a binary expansion in terms of the k'_i $(i=0,\ldots,j-1)$,

$$k_j = \sum_{i=0}^{j-1} k'_i\, 2^{j-1-i}. \tag{22}$$

Moreover, $k_j = \left\lfloor \frac{k_{j+1}}{2} \right\rfloor$ and $k'_j = k_{j+1} - 2\left\lfloor \frac{k_{j+1}}{2}\right\rfloor$, with $\lfloor x \rfloor$ the largest integer less than or equal to x. Note that fixing a sequence k'_i specifies not only k_j, but a "line of descendants" of U_{i,k_i} $(i=0,\ldots,j)$ from $U_{0,0}$ down to U_{j,k_j}.

Using this notation, we can derive closed-form expressions for the MWM wavelet and scaling coefficients.

Proposition 1 *Define the wavelet coefficients of the Haar wavelet system through (20), with the random variables $A_{j,k}$ supported on $[-1,1]$. We then have the general relations*

$$U_{j,k_j} = 2^{-j/2}\, U_{0,0} \prod_{i=0}^{j-1} \left[1 + (-1)^{k'_i} A_{i,k_i}\right] \tag{23}$$

and

$$W_{j,k_j} = 2^{-j/2}\, A_{j,k_j}\, U_{0,0} \prod_{i=0}^{j-1} \left[1 + (-1)^{k'_i} A_{i,k_i}\right]. \tag{24}$$

3.4 Properties of the MWM

3.4.1 Additional constraints on the multipliers

The Haar wavelet coefficients of a stationary signal will be, using (5), identically distributed within each scale with $\mathbb{E}[W_{j,k}] = 0$. To model these properties in the MWM, we will assume that, within each scale j, we have the following:

1. the multipliers $A_{j,k}$, $k = 0, 1, \ldots, 2^{j-1}$, are *identically* distributed according to some random variable $A_{(j)} \in [-1, 1]$,
2. the $A_{(j)}$ are *symmetric* about 0, and
3. (simplifying assumption) the $A_{j,k}$ are *independent* of both the coarsest scaling coefficient $U_{0,0}$ and the $A_{l,k}$ on finer scales $l > j$.[5]

3.4.2 Marginal density and stationarity

Under the above assumptions, Proposition 1 leads us to the marginal density and stationarity properties of $C^{(n)}[k]$. Setting $j = n$ in (22) and (23), and setting $k = k_n$ in (15) yields[6]

$$\begin{aligned} C^{(n)}[k] &= 2^{-n}\, U_{0,0} \prod_{i=0}^{n-1} \left(1 + (-1)^{k_i'} A_{i,k_i}\right) \\ &\stackrel{d}{=} 2^{-n}\, U_{0,0} \prod_{j=0}^{n-1} (1 + A_{(j)}). \end{aligned} \qquad (25)$$

Thus, $C^{(n)}[k]$ is first-order stationary and identically distributed. Note that without the requirement that $A_{(j)}$ be symmetric, the marginal distribution of $C^{(n)}[k]$ would depend on k and (25) would not hold. Hence, symmetry of the multipliers is key for modeling stationary processes.

However, $C^{(n)}[k]$ will not be second-order stationary in general. Due to the dyadic structure of the wavelet transform, wide-sense stationarity of $C^{(n)}[k]$ is unattainable using a wavelet-domain model with uncorrelated wavelet coefficients (except in the trivial case of white noise). In the MWM, for a fixed shift m, $\mathbb{E}\left[C^{(n)}[k+m]C^{(n)}[k]\right]$ will vary as a function of k in relation to the size of the smallest subtree containing both $C^{(n)}[k+m]$ and $C^{(n)}[k]$. If the $A_{j,k}$ multipliers are independent and identically distributed (iid), then the smaller the subtree, the stronger the potential correlation.

Given our independence assumptions, the moments of $C^{(n)}[k]$ are readily calculable from (25) via

$$\mathbb{E}\left[C^{(n)}[k]^q\right] = \mathbb{E}\left[U_{0,0}^q\right] \prod_{j=0}^{n-1} \mathbb{E}\left[\left(\frac{1 + A_{(j)}}{2}\right)^q\right]. \qquad (26)$$

As we increase the number of scales in the wavelet transform ($n \to \infty$), an appropriately scaled version of $C^{(n)}[k]$ converges to a log-normal random variable as long as $\mathbb{E}\left[\log(A_{(j)})^3\right]$ is bounded for $j \geq 0$. This follows from the application to $\log(C^{(n)}[k])$ of the Berry-Esseen theorem [39], a Central Limit Theorem for non-identically distributed random variables.

3.4.3 Wavelet-domain dependency structure

If we assume that the $A_{j,k}$'s are independent both between scales and within scales, then the wavelet coefficients will be dependent, but uncorrelated. This lack of correlation follows from the fact that terms of the form $\mathbb{E}[A_{j,k}]$ factor out of any correlation calculation, with $\mathbb{E}[A_{j,k}] = 0$. However, a higher-order dependency structure remains, which is of course key for preserving signal positivity.

While a dependency structure with no correlations between wavelet coefficients may at first seem somewhat unnatural, such models are not entirely unrealistic. Wavelet coefficients of random signals can exhibit minimal second-order correlations (approximately decorrelated via the Karhunen-Loève transform), yet still have strong dependencies in higher-order moments. For instance, many real-world data sets exhibit strong dependencies in the energy of the wavelet coefficients, corresponding to fourth-order cross-moments [40, 41].

3.5 Related work

Constructions similar to the MWM were developed earlier in [42, 43]. A similar multiplicative model for wavelet coefficients has been developed in [44, 45], where it is applied to wavelet-domain Bayesian estimation of the intensity of a Poisson process. There, the $A_{j,k}$'s are independent multipliers that, within each scale, are identically-distributed as mixtures of β random variables. The primary difference with this work is that we model the data directly, whereas [44, 45] models a wavelet-domain prior density for the intensity function of a Poisson process.

In other related work, [46] models the wavelet coefficients using a context-based hidden Markov model. It can be shown that this model corresponds to (20), again with the $A_{j,k}$'s identically-distributed within each scale, but with each $A_{j,k}$ distributed according to a mixture density dependent on the value of $U_{j,k}$. Although this model proves to be quite flexible and accurate for characterizing positive LRD data, it requires iterative, maximum-likelihood (expectation-maximization) training, has numerous parameters, and is difficult to characterize analytically.

4 Data Modeling using the MWM

To complete our model, we now specify probability density functions (pdfs) for the coarsest scaling coefficient $U_{0,0}$ and for the $A_{(j)}$ multipliers at each scale. We can use the degrees of freedom in these pdfs in order to control two key signal properties. First, we control the correlations and LRD of the output signal $C^{(n)}[k]$ through the wavelet energy decay. Second, we control the higher-order moments and marginal pdf of $C^{(n)}[k]$ through the scaling coefficient moments.

4.1 Controlling the Wavelet Energy Decay

To approximate the correlation behavior of a target signal, we vary the wavelet energy decay across scale. We choose the pdfs for the $A_{(j)}$'s to control the wavelet coefficients' scaling behavior via (24). The fact that this scaling behavior allows us to model correlations can be explained as follows.

Consider the Karhunen-Loève properties of the wavelet

[5] Strictly speaking, for our development we need only assume independence along "lines of descendants." That is, multipliers on different scales can be dependent as long as one is not a descendant of the other.

[6] The symbol "$\stackrel{d}{=}$" denotes equality in distribution

transform. Previous work [11, 12, 47] has demonstrated that the wavelet transform approximately decorrelates or whitens a general class of LRD signals, including $1/f$ processes. If the decorrelation were exact, then specifying the correct variances of the wavelet coefficients would fully capture the correlation structure of the signal. Since this decorrelation is approximate, we can approximately control the correlation behavior by appropriately setting the second moments (energies) of the wavelet coefficients at each scale.

The simplest way to control energy scaling is to fix the energy at the coarsest scale ($j = 0$) and then set the ratios of energy for the other scales with $\eta_j := \frac{\mathrm{var}(W_{j-1,k})}{\mathrm{var}(W_{j,k})}$, $0 \leq j < n$. For a stationary $1/f$ process, we see from (13) that $\eta_j = 2^{2H-1}$ is constant. Using Proposition 1 we can calculate the η_j's of the MWM via

$$\begin{aligned} \eta_j &= \frac{\mathbb{E}\left[W_{j-1,k}^2\right]}{\mathbb{E}\left[W_{j,k}^2\right]} \\ &= \frac{2\,\mathbb{E}\left[A_{(j-1)}^2\right]\,\mathbb{E}\left[U_{j-1,k}^2\right]}{\mathbb{E}\left[A_{(j)}^2\right]\,\mathbb{E}\left[(1+A_{(j-1)})^2\right]\,\mathbb{E}\left[U_{j-1,k}^2\right]} \\ &= 2\frac{\mathbb{E}\left[A_{(j-1)}^2\right]}{\mathbb{E}\left[A_{(j)}^2\right]\left(1+\mathbb{E}\left[A_{(j-1)}^2\right]\right)}. \end{aligned} \quad (27)$$

To match a given variance decay, we can recursively solve (27) for $\mathbb{E}\left[A_{(j)}^2\right]$ in terms of η_j and $\mathbb{E}\left[A_{(j-1)}^2\right]$ for $j = 1, 2, \ldots, n-1$. We initialize the calculation at the coarsest scale ($j = 0$) through

$$\mathbb{E}\left[A_{(0)}^2\right] = \frac{\mathbb{E}\left[W_{0,0}^2\right]}{\mathbb{E}\left[U_{0,0}^2\right]}. \quad (28)$$

4.2 Controlling the moments of the scaling coefficients

It is easily shown that the moments of the scaling coefficients scale according to

$$\frac{\mathbb{E}\left[U_{j-1,k}^q\right]}{\mathbb{E}\left[U_{j,k}^q\right]} = 2^{q/2}\,\mathbb{E}\left[(1+A_{(j-1)})^q\right]^{-1}. \quad (29)$$

Through (29) we can control the scaling of the higher-order (and even negative) moments of the scaling coefficients — and thus of $C^{(n)}[k]$ — through the moments of the $A_{(j)}$'s.

4.3 Distributions for the multipliers

We will investigate two distributions for the multipliers, the symmetric β distribution and a symmetric point-mass distribution. Both of these distributions are compactly supported, easily shaped, and amenable to closed-form calculations.

4.3.1 Symmetric beta distribution

A $\beta(p,p)$ random variable A, symmetrically distributed over $(-1, 1)$, has pdf [48]

$$g_A(a) = \frac{(1+a)^{p-1}(1-a)^{p-1}}{\mathcal{B}(p,p)\;2^{2p-1}}. \quad (30)$$

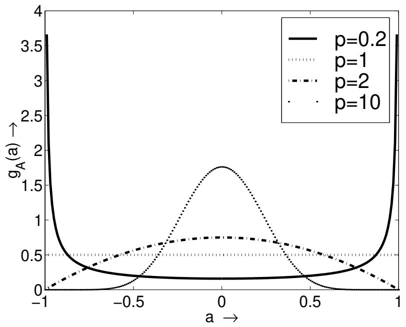


Figure 4: Examples of the pliable pdf $g_A(a)$ of the $\beta(p,p)$ random variable A, for different values of p. For $p = 0.2$, A resembles a binomial random variable, and for $p = 1$ it has a uniform density. For $p > 1$ the density resembles a truncated Gaussian density, with the resemblance increasing with p.

Here $\mathcal{B}(\cdot,\cdot)$ is the beta function, and $p > 0$ is a shape factor (see Figure 4). For large p, the $\beta(p,p)$ approximates a Gaussian distribution [48]. The variance is given by

$$\mathrm{var}(A) = \mathbb{E}\left[A^2\right] = \frac{1}{2p+1}. \quad (31)$$

Combined with (27), (31) tells us how to choose the p's to obtain the desired scaling behavior as parameterized via η_j. Denoting by $p_{(j)}$ the beta parameter at scale j, we find that

$$p_{(j)} = \frac{\eta_j}{2}\left(p_{(j-1)} + 1\right) - 1/2. \quad (32)$$

When we use β-distributed multipliers, we call the model the *β multifractal wavelet model* (βMWM).

4.3.2 Point-mass distribution

The point mass distribution we consider is non-zero at three points

$$\begin{aligned} &\Pr[A = c] = \Pr[A = -c] = r \\ &\Pr[A = 0] = 1 - 2r \end{aligned} \quad (33)$$

with $0 \leq r, c \leq 1$. Although seemingly not as rich as the β, this distribution has two parameters and thus can match an additional higher-order moment of the signal.

The point-mass distribution has variance $\mathrm{var}(A) = 2rc^2$. The higher order moments of $\left(\frac{1+A}{2}\right)$, which are useful for characterizing the scaling coefficient moments (see (26)), are given by

$$\begin{aligned} \mathbb{E}\left[\left(\frac{1+A}{2}\right)^q\right] &= 2^{-q}r\left((1-c)^q + (1+c)^q\right) \\ &\quad + 2^{-q}(1-2r). \end{aligned} \quad (34)$$

4.4 Distribution for the root scaling coefficient

What remains is to model the density of $U_{0,0}$, the root of the tree in Figure 3. In theory, this distribution should be strictly positive. However, if there are enough scales in the wavelet transform, we can appeal to Central Limit Theorem-type arguments (although LRD makes precise analysis somewhat cumbersome) that the root scaling coefficient is approximately Gaussian, thus characterized only through its mean

Table 1: Asymptotic values for the shape p and variance $\mathbb{E}(A^2)$ of the β multipliers $A_{j,k}$ as a function of H.

H	0.55	0.6	0.65	0.7	0.75	0.8	0.85	0.9	0.95
p	0.077	0.175	0.301	0.470	0.707	1.06	1.66	2.86	6.47
$\mathbb{E}[A^2]$	0.866	0.741	0.625	0.516	0.414	0.320	0.231	0.149	0.072

$E[U_{0,0}]$ and variance $\mathrm{var}(U_{0,0})$. Crucial to this assumption is that the mean greatly outweighs the variance so that the probability of a negative value is negligible.

Although our development has focused on a single wavelet tree with a single scaling coefficient $U_{0,0}$, in certain synthesis applications it is useful for the MWM to employ several wavelet trees with one root scaling coefficient per tree. For instance, we may wish to synthesize a trace of length 2^{n_0}, but have only enough coarse-scale information to form a model over $n < n_0$ scales. In this case, we can concatenate 2^{n_0-n} length-2^n traces, which corresponds to an MWM with 2^{n_0-n} iid coarsest-scale scaling coefficients $U_{0,k}$. Of course, an iid assumption for the $U_{0,k}$ is suboptimal in that it destroys LRD over time lags greater than 2^n. This problem, along with a potential solution, is discussed further in Section 4.6.

4.5 Modeling positive $1/f$ noise

We next investigate how to parameterize the MWM in order to model a stationary positive-valued $1/f$ increments process with Hurst parameter H, or spectrum decay $\propto f^{-(2H-1)}$. It is easily seen from (13) that we should choose $\eta_j = 2^{2H-1}$ independently of scale. This leads to:

Proposition 2 *Assume that the $A_{j,k}$ in (20) are iid within each scale j (distributed as $A_{(j)}$), supported on $[-1,1]$, symmetric about 0, and such that*

$$\mathbb{E}\left[A^2_{(j)}\right] = \frac{2^{2-2H}\,\mathbb{E}\left[A^2_{(j-1)}\right]}{1+\mathbb{E}\left[A^2_{(j-1)}\right]}. \tag{35}$$

Then the MWM output process $C^{(n)}[k] = 2^{-n/2}U_{n,k}$ is positive and exhibits power-law behavior of the wavelet coefficient energies (14) with exponent $2H-1$. Moreover,

$$\lim_{j\to\infty} \mathbb{E}\left[A^2_{(j)}\right] = 2^{2-2H}-1, \qquad 1/2 < H < 1. \tag{36}$$

The first part, i.e. (35), follows from (27). By solving (35) for the fixed-point, we obtain (36). A simple analysis of (36) shows that for $1/2 < H < 1$ the iteration is well-defined on all scales, since the variance of $A_{(j)}$ must lie in $[0,1]$ for all j.

If we use a β distribution for the multipliers, the fixed point formula for the variance $\mathbb{E}[A^2]$ leads to a fixed point for p of the form

$$p = \lim_{j\to\infty} p_{(j)} = \frac{2^{2H-1}-1}{2-2^{2H-1}}, \qquad 1/2 < H < 1. \tag{37}$$

Table 1 provides typical fixed-point values for p and the variance $\mathbb{E}[A^2]$ given the desired H. There is no such expression for the point-mass distribution, since even though the variance converges, an extra degree of freedom remains available for matching higher-order moments.

We conclude that the MWM can approximate a positive-valued $1/f$ process with Hurst parameter $1/2 < H < 1$ to infinitely fine resolution.

4.6 Fitting the MWM to data measurements

We now develop a procedure for fitting the MWM to actual data measurements. The first step in the fitting is a wavelet analysis: we compute the wavelet coefficients of the measurements (a length-N signal) using a Haar wavelet transform algorithm (filter bank, etc. [9, 10]) The number of wavelet scales in the transform, n, is chosen as mentioned below.

We require $\mathrm{var}(W_{j,k})$, $j = 0,\ldots,n-1$ and $\mathbb{E}\left[U^2_{0,0}\right]$ to fit the MWM via (27) and (28). (Values for the higher-order scaling coefficient moments (29) may also be useful if the multiplier densities have more than one free parameter.) There exist two reasonable approaches for selecting these values. We can either plug in the empirical wavelet variances directly, or we can assume a parametric model for the variances and use the measured data to fit the model.

If we plug the empirical moments directly into (27) and (28), we must ensure that we have enough data to collect reliable statistics. This problem is most pressing for the coarsest-scale wavelet and scaling coefficients, of which we have the fewest. In practice, we set the number of levels n of the Haar transform such that the number $\lfloor N2^{-n}\rfloor$ of coarsest-scale wavelet and scaling coefficients is sufficient for estimating $\mathbb{E}\left[W^2_{0,0}\right]$ and $\mathbb{E}\left[U^2_{0,0}\right]$.

A parametric model for the moment scaling would allow us to extrapolate the coarse-scale scaling and wavelet coefficient moments that we have difficulty measuring due to lack of data. It would also render the modeling more robust and provide a more concise representation of the data's behavior. Parametric models for η_j as a function of scale are currently under investigation.

In some cases, it may be impossible to exactly match the moment scaling of the data using the MWM. The scaling of moments of the actual data may be inconsistent with the possible moments of the $A_{j,k}$ multipliers. For instance, the positive moments of $A_{j,k}$ are bounded above by those of a random variable with point masses of weight 1/2 at -1 and at 1. The moment scaling of certain data may lead to multiplier moment constraints outside these bounds that cannot be fit exactly. This could occur, for example, if the data exhibited dependencies between the $A_{j,k}$ and $U_{j,k}$.

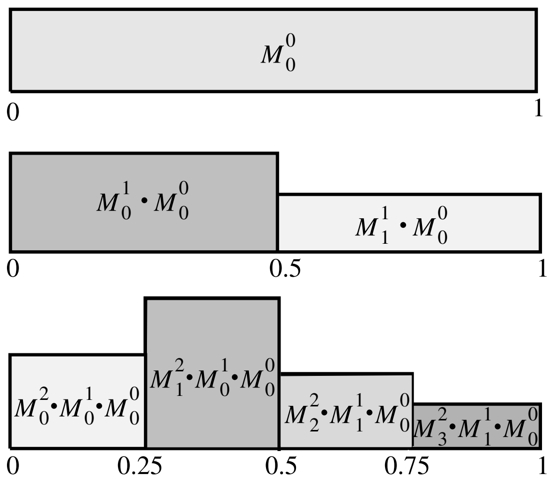


Figure 5: Iterative construction of the binomial cascade. In the second image, the products $M_i^1 \cdot M_0^0$ give the area of the respective shaded region, i.e. the increment $C_i^{(1)}$ of D over I_i^1 etc. The height of a rectangle of length 2^{-n} is thus $2^n \cdot C_i^{(1)}$. Relation (39) guarantees that the areas add up in the right way. In particular, the height of the shaded rectangles is $2M_k^n$ times the height of the respective "parent rectangle". This is how "spikiness" is created: small heights give rise to at least one even smaller child, while large ones produce at least one even larger child. A more precise statement can be found in Appendix A.3 from which it can be inferred that in the limit the spikes will actually be infinitely large on a rather large, dense subset of $[0, 1]$.

5 Multiplicative Cascades

Multiplicative cascades generalize the self-similarity of fBm by offering greater flexibility and richer scaling properties. Identifying the MWM algorithm with a multiplicative cascade allows us to benefit from the accumulated theoretical and practical knowledge of the field of multifractals, including a precise understanding of the convergence of the algorithm, properties of the marginal distributions, advantages over monofractal fGn models, and a range of possible refinements and extensions [15, 16, 22–32, 49–57]. The theory of cascades comes with a dedicated set of tools for analysis, both theoretical and numerical, that we will outline in the next two sections (see Appendices A and B for more details).

At this point, our discussion will become decidedly more technical, mainly because we wish to extend the MWM to a continuous-time process. Though indispensable for a true understanding of multiplicative processes, readers may, at least at first reading, wish to bypass the following two sections for Section 7, where we present an application of the MWM framework to computer network traffic modeling.

5.1 The MWM is a binomial cascade

The MWM extends the simple, classical multifractal — the *binomial measure* μ [22, 53, 54, 57] — in a natural fashion. This measure μ is most conveniently constructed iteratively through a so-called *cascade* structure, whence it is often addressed as a *binomial cascade*. As we will show, its distribution function $D_b(t) := \mu([0, t))$ coincides with the integral D of the MWM signal $C[k]$.

The iterative cascade construction is illustrated in Figure 5. Starting from a uniform distribution on the unit interval of total mass M_0^0, we "redistribute" this mass by splitting it between the two subintervals of half size in the ratio M_0^1 to M_1^1, with $M_0^1 + M_1^1 = 1$. Proceeding iteratively, we obtain after n steps a distribution that is uniform on intervals $I_k^n := [k2^{-n}, (k+1)2^{-n})$ and assigns to these intervals the mass

$$\begin{aligned} C_b^{(n)}[k_n] &:= D_b((k_n+1)2^{-n}) - D_b((k_n)2^{-n}) = \mu(I_k^n) \\ &= M_{k_n}^n \cdot M_{k_{n-1}}^{n-1} \cdots M_{k_1}^1 \cdot M_0^0. \end{aligned} \tag{38}$$

Here we again use the notation (22) at scale $j = n$. The tree structure of Figure 3 translates easily into the present situation: the interval $I_{k_n}^n$ lies within the intervals $I_{k_i}^i$ ($i = 0, \ldots, n-1$) which form a nested sequence. If $k_i' = 0$, then $I_{k_{i+1}}^{i+1}$ is the left subinterval of its parent interval $I_{k_i}^i$; if $k_i' = 1$, it lies on the right.

To generate a random D_b, we choose the various M_l^i to be random variables. Their distributions may depend on i and l and are arbitrary, as long as they are positive and provided that for all j and k_{j-1}

$$M_{2k_{j-1}}^j + M_{2k_{j-1}+1}^j = 1 \tag{39}$$

almost surely. This introduces a strong dependency between "siblings", i.e. the multipliers at the two child nodes sharing the same parent. We will require for all j and k_j that all multipliers appearing in (38) are mutually independent. We will call this property *independence along lines of descendants.* A compact way of writing this is

$$\text{if } I_k^n \subset I_l^j, \text{ then } M_k^n \text{ and } M_l^j \text{ are independent.} \tag{40}$$

As long as the two dependency requirements (39) and (40) are satisfied, we are completely free to introduce additional correlation structure.

Comparison with Proposition 1 (applied with $j = n$) or, more pointedly, with (25) reveals that the MWM is a random binomial cascade. Indeed, setting $M_0^0 = D_b(1) - D_b(0) := U_{0,0}$ and

$$M_{k_n}^n = \frac{1 + (-1)^{k_{n-1}'} A_{n-1,k_{n-1}}}{2}, \tag{41}$$

the increments $C_b^{(n)}[k]$ of this binomial distribution function D_b (c.f. (38)) coincide with the increments $C^{(n)}[k]$ of integral D of the MWM signal (c.f. (25)). Thus, we drop the subscript "$_b$" in the sequel.

5.2 Additional properties of the MWM

Since the MWM is a binomial cascade, known results on cascades transfer immediately.

5.2.1 Ordinary convergence of $D(t)$

In the limit the above iterative construction will converge, meaning that D is well defined for all t. This is due essentially to two simple properties of distribution functions such as D: they are increasing and continuous from the right. Thus, it is enough to define D at all dyadic points, and to take limits from the right at non-dyadic points. At stage n, we define $D(k2^{-n})$ through (38) with the convention $D(0) = 0$. At later steps of the construction, these values remain unchanged due to (39). This completes the argument.

Let us note that the increment $C^{(n)}[k_n]$ of D between dyadic points tends to zero as $n \to \infty$ due to (38) and the fact that the multipliers are less than 1. Consequently, D is continuous.

5.2.2 Distributional convergence of $C(t)$

We have constructed D through its dyadic increments $C^{(n)}[k]$ and by passing to the limit of infinitely fine resolution ($n \to \infty$). Later, we will be mainly interested in the increments $C^{(n)}[k]$. Nevertheless, defining D itself is handy, since it is a continuous-time process and provides a compact representation of the increment processes (15) and (38) at various resolutions n.

Moreover, we cannot define a "process" $C(t)$ with $D(t) = \int_0^t C(s)\,ds$ in the usual sense. Indeed, the approximations $C(k2^{-n}) \sim 2^n C^{(n)}[k_n]$ (plotted in Figure 5) tend either to 0 or ∞. (c.f. Appendix A.3). In particular, the derivative D' of D is zero almost everywhere, as follows from (75) in Appendix A.1. Thus, the essential growth of D happens "at" the points where D' does not exist. This explains the spiky appearance of the increments $C^{(n)}[k]$ for large n.

The proper way to define $C(t)$ is in the distributional sense:

$$\begin{aligned}\int g(t)\,C(t)\,dt &:= \lim_{n\to\infty} \sum_{k=0}^{2^n-1} g(k2^{-n})\,C^{(n)}[k] \\ &= \lim_{n\to\infty} \sum_{k=0}^{2^n-1} g(k2^{-n})\,\mu(I_k^n) \\ &= \int g(t)\,\mathrm{d}\mu(t). \end{aligned} \tag{42}$$

As a particular case, the wavelet and scaling coefficients of $C(t)$ are properly defined, and it is an easy task to check that they are indeed given by (23) and (24).

To emphasize the fact that $C(t)$ is not a proper function in the cases of interest here, let us show that the l^2-norm of its wavelet coefficients is infinite, at least in expectation. Indeed, using Proposition 1 we find after a short calculation that $\mathbb{E}\left[\sum_{j,k} |W_{j,k}|^2\right] = \sum_{j=0}^n \sum_k \mathrm{var}(W_{j,k}) = \mathbb{E}[U_{0,0}^2] \sum_{j=0}^n \mathbb{E}\left[A_{(j)}^2\right] \prod_{i=0}^{j-1} \mathbb{E}\left[(1+A_{(i)})^2\right]$. For this expression to remain finite as $n \to \infty$, $\mathbb{E}\left[A_{(j)}^2\right]$ would have to decay to 0 (as $j \to \infty$) due to $\mathbb{E}\left[(1+A_{(j)})^2\right] \geq 1$. This requirement, however, leads to processes with uninteresting fine scale behavior, and it certainly does not hold in the presence of LRD (see (36)).

The fact that the MWM algorithm does not furnish an L^2-signal[7] in the limit $n \to \infty$ provides a further strong argument towards leaving the usual framework of wavelets when performing multiplicative iteration schemes. Given the decay of the wavelet coefficients (c.f. Section 6.3) we can determine in which *Besov spaces* the limiting MWM signal $C(t)$ lives (see Appendix A.2.6).

5.2.3 Marginals of $C^{(n)}[k]$

Our next observation concerns the marginals of the discrete approximation $C^{(n)}[k]$ to the MWM signal $C(t)$. If we assume that the multipliers M appearing in (38) are mutually independent with finite third moments, then the logarithms of the increments $C^{(n)}[k]$ of D are approximately Gaussian due to the Law of Large Numbers (LLN). A cascade process has, thus, approximately log-normal marginals $C^{(n)}[k]$. Note that these marginals have finite moments of the same order of the multipliers appearing in (38).

The theory of cascades, which in mathematics are addressed as T-martingales [22, 52], provides a wealth of possible generalizations. Softening the conservation condition "$M_0 + M_1 = 1$ almost surely" to "$\mathbb{E}[M_0 + M_1] = 1$" (consistency in the mean), we can use multipliers M with log-normal distribution. Then, the marginals of the increment process are exactly log-normal on all scales. In this case, convergence is guaranteed by martingale arguments.

Also of considerable importance is the possibility to go beyond the binary structure imposed by the Haar wavelet system and to introduce randomness in the geometry of the construction [24,25] and — as a particular case — wide sense stationarity in the signal. To describe such systems is, however, beyond the scope of this paper.

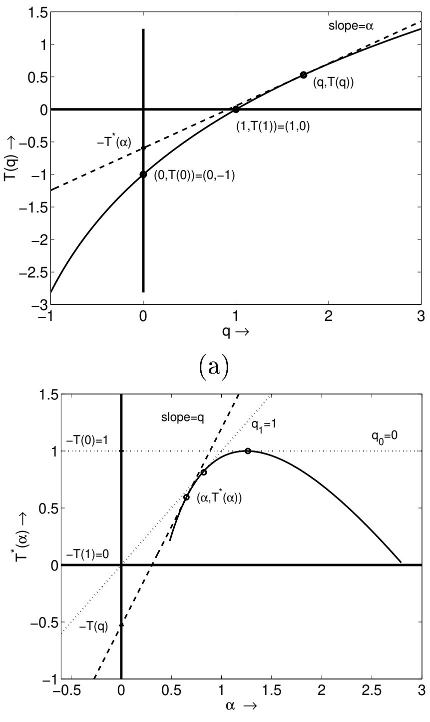
6 Multifractal Analysis of the MWM

So far we have noted two attractive properties of cascades: their increment processes are spiky and have nonGaussian marginals. Surprisingly, these two properties are strongly related, and much effort has been expended connecting them rigorously under various assumptions [23–32]. The scaling of moments, which is captured with the simple and efficient *partition function* $T(q)$, acts as the bridge. This function can be viewed as a concise way of describing various features of cascades and of processes in general.

After introducing the various *multifractal spectra* $f(\alpha)$ (measures of spikiness) and relating them to $T(q)$, we show that fBm has a degenerate multifractal structure. It is, thus, of limited use for modeling purposes in view of higher-order moments. Next, we relate the multifractal analysis (MFA) to the wavelet transform of a signal and unravel the connection between MFA and LRD. We end this section by computing the multifractal spectrum of the MWM explicitly.

A thorough review of the key features of multifractal analysis is given in Appendix A.

[7] Since the Haar transform is orthonormal, the l^2 norm of the wavelet



(a)

(b)

Figure 6: The Legendre transform $T \mapsto T^*$ in the simple case of a concave, differentiable function such as the spectrum of a βMWM ((66) with $p = 1.66$, $H = 0.85$). Set $\alpha = T'(q)$; then $T^*(\alpha)$ is such that the tangent at $(q, T(q))$ passes through $(0, -T^*(\alpha))$. In other words, $-T^*(\alpha) + q\alpha = T(q)$ (see (51)). By symmetry, the tangent at $(\alpha, T^*(\alpha))$ has slope q and passes through $(0, -T(q))$. There are two special values of q. Trivially, $T(0) = -1$, hence the maximum of T^* is 1. In addition, every positive increment process has $T(1) = 0$, hence T^* touches the bisector.

6.1 Multifractal spectra

6.1.1 Spikiness

The strength of growth, also called the *degree of Hölder continuity,* of an increasing process Y at time t can be characterized by

$$\alpha(t) := \lim_{k_n 2^{-n} \to t} \alpha_{k_n}^n, \tag{43}$$

with

$$\alpha_{k_n}^n := -\frac{1}{n} \log_2 \Delta_{k_n}^n[Y], \tag{44}$$

$$\Delta_{k_n}^n[Y] := \left| Y((k_n+1)2^{-n}) - Y(k_n 2^{-n}) \right|, \tag{45}$$

and $k_n = 0, \ldots, 2^n - 1$.

The smaller the $\alpha(t)$, the faster Y grows at t. Considering only $t \in [0,1]$ for simplicity, the frequency of occurrence of a given strength α at coarse scales can be measured by the *coarse (grained) multifractal spectrum*:

$$f_{\mathrm{G}}(\alpha) := \lim_{\varepsilon \to 0} \lim_{n \to \infty} \frac{1}{n} \log_2 \#\{\alpha_{k_n}^n \in (\alpha - \varepsilon, \alpha + \varepsilon)\}. \tag{46}$$

In this setting, f_{G} takes values between 0 and 1 and is often shaped like a $\cap$ (concave). The smaller $f_{\mathrm{G}}(\alpha)$ is, the "fewer" points t act like $\alpha(t) \simeq \alpha$. If $\overline{\alpha}$ denotes the value $\alpha(t)$ assumed by "most" points t, then $f_{\mathrm{G}}(\overline{\alpha}) = 1$ (c.f. Appendix A.3).

6.1.2 NonGaussianity and higher-order moments

Like any Gaussian process, fBm is completely determined by its second-order statistics. Things are quite the contrary for cascades such as the MWM. Being especially interested in the *scaling* of moments, we define the *partition function*

$$T(q) := \lim_{n \to \infty} \frac{1}{-n} \log_2 \mathbb{E}\left[\sum_{k_n=0}^{2^n-1} (\Delta_{k_n}^n[Y])^q\right]. \tag{47}$$

Note that T is always concave. For a typical plot of f_{G} and T, see Figure 6.

6.1.3 The multifractal formalism

The multifractal spectrum $f_{\mathrm{G}}(\alpha)$ and $T(q)$ are closely related, as the following quick and dirty argument shows. Omitting in the sum of (47) all terms but the ones with $\alpha_{k_n}^n \approx \alpha$ and using (46), we obtain

$$\begin{aligned} \sum_{k_n=0}^{2^n-1} (\Delta_{k_n}^n[Y])^q &\geq \sum_{\alpha_n \sim \alpha} (2^{-n\alpha})^q \\ &\simeq 2^{n f_{\mathrm{G}}(\alpha)} 2^{-nq\alpha} \\ &= 2^{-n(q\alpha - f_{\mathrm{G}}(\alpha))}. \end{aligned} \tag{48}$$

We conclude that we should "expect" $T(q)$ to be smaller than $q\alpha - f_{\mathrm{G}}(\alpha)$, or equivalently $f_{\mathrm{G}}(\alpha) \leq q\alpha - T(q)$. Since this holds for all α and q, we find

$$T(q) \leq f_{\mathrm{G}}{}^*(q) := \inf_{\alpha}(q\alpha - f_{\mathrm{G}}(\alpha)) \tag{49}$$

and

$$f_{\mathrm{G}}(\alpha) \leq T^*(\alpha) := \inf_{q}(q\alpha - T(q)). \tag{50}$$

This relation is established rigorously in Appendix A.2.

The transform $T^*(\alpha)$ appearing in (50) is called the *Legendre transform.* If $T''(q) < 0$, then we find by simple calculus that

$$\begin{aligned} T^*(\alpha) &= q\alpha - T(q) \quad \text{and} \\ (T^*)'(\alpha) &= q \ \text{ at } \alpha = T'(q). \end{aligned} \tag{51}$$

We may write this equivalently as the dual formula $T(q) = q\alpha - T^*(\alpha)$, $T'(q) = \alpha$ at $q = (T^*)'(\alpha)$. This is illustrated in Figure 6. Since T is typically differentiable and always concave, (51) is sufficient for our purposes. More details on the Legendre transform are given in Appendix A.2.

This relation via the Legendre transform is typical in the theory of Large Deviations [58], which establishes relations such as equality in (50) under the weakest possible assumptions. In proper terminology, f_{G} is the *rate function* of a so-called *Large Deviation Principle* (LDP): it measures how frequently or how likely the observed $\alpha_{k_n}^n$ deviates from the "expected value" $\overline{\alpha}$. We will elaborate on this, especially the use of a theorem of Gärtner-Ellis [59] towards an improvement of (50) in Appendix A (c.f. Theorem 6 and 9).

coefficients equals the L^2 norm of the output signal.

6.2 Numerical estimation of $T(q)$

For the MWM, we have $\Delta_k^n[D] = C^{(n)}[k] = 2^{-n/2}U_{n,k}$, and the sum in (47) becomes

$$S_n(q) = \sum_{k_n=0}^{2^n-1} (\Delta_{k_n}^n[D])^q = \sum_{k=0}^{2^n-1} |2^{-n/2}U_{n,k}|^q. \quad (52)$$

In order to numerically estimate $T(q)$, we will first ignore the expectation, which is fair for large n under an ergodicity assumption. (This procedure is also viable in more general circumstances, as we show in Appendix A.2.4.) Then, we seek a relation of the form $2^{(-n)T(q)} \simeq S_n(q)$, which we obtain numerically from a linear plot of $\log_2 S_j(q)$ against j ($j = 1, \ldots, n$).

6.3 Multifractal analysis and wavelets

Wavelet decompositions contain considerable information on the singularity behavior of a process Y. Indeed, adapting the argument of [60, p. 291] and correcting for the L^2 wavelet normalization used in this paper, it is easily shown that $|Y(s) - Y(t)| = O(|s-t|^\alpha)$ implies that

$$2^{n/2} \left| \int Y(s)\, \psi_{n,k_n}^*(s)\, \mathrm{d}s \right| = O\Big(2^{-n\alpha(t)}\Big) \quad (53)$$

if k_n is chosen as usual to satisfy $k_n 2^{-n} \le t \le (k_n+1)2^{-n}$. This holds for any $\alpha > 0$ and any compactly supported wavelet. Given knowledge on the decay of the maximum of the wavelet coefficients in the vicinity of t and sufficient wavelet regularity, this relation can be inverted. For a precise statement, see [60] and [9, Thm. 9.2]. This suggests that replacing the increments in the definition (43) of $\alpha(t)$ by the left hand side of (53) would produce an alternative description of the local behavior of Y. In nice cases, we expect the resulting scaling exponent to be equal to $\alpha(t)$. This could prove particularly useful for more general classes of processes.

Let us rejoin the MWM. By construction, we actually know the wavelet coefficients of the MWM signal C, which is the distributional derivative of the increasing process D. Following the above recipe we may define, thus, a *multifractal scaling exponent based on wavelets* for C:

$$\begin{aligned} \widetilde{\alpha}_{[C]}(t) &:= \lim_{n\to\infty} \widetilde{\alpha}_{k_n}^n[C] \quad \text{as } k_n 2^{-n} \to t. \quad (54) \\ \widetilde{\alpha}_{k_n}^n[C] &= -\frac{1}{n}\log_2\left(2^{n/2}|W_{n,k_n}|\right), \end{aligned}$$

Since $D(t) = \int_0^t C(s)\,\mathrm{d}s$, we expect $\widetilde{\alpha}_{[C]}(t)$ to be closely related to $\alpha(t)$. Adapting (47) to (54) results in

$$\widetilde{T}_{[C]}(q) := \lim_{n\to\infty} \frac{1}{-n} \log_2 \mathbb{E}\left[\sum_{k=0}^{2^n-1} 2^{nq/2}|W_{n,k}|^q\right]. \quad (55)$$

An analysis using (55) is of particular interest in the context of *Besov spaces,* as is explained further in Appendix A.2.6.

All general results on the multifractal formalism hold also with $\widetilde{\alpha}$ and $\widetilde{T}$, in particular (50) and Lemma 5, Theorem 6, Lemma 7, and Corollary 8 of the Appendix A. We should mention that [21] uses this fact in its analysis of cascades.

For the Haar wavelet coefficients of an MWM, we have $W_{n,k} = A_{n,k}U_{n,k}$. Provided that the $A_{n,k}$ converge in distribution as $n \to \infty$, they do not contribute to the scaling law $\widetilde{T}_{[C]}(q)$. For the sum in (55), we have then that $\mathbb{E}\Big[\widetilde{S}_n(q)\Big] = 2^{nq}\,\mathbb{E}\Big[|A_{(n)}|^q\Big]\,\mathbb{E}[S_n(q)]$ using (52). Hence,

$$\widetilde{T}_{[C]}(q) = -q + T(q). \quad (56)$$

Let us assume in addition that there exists $\varepsilon > 0$ such that $|A_{n,k}| \ge \varepsilon$ for all n, k. Then, $(1/n)\log_2 A_{n,k_n} \to 0$ for all t, and using again $2^{-n/2}U_{n,k} = \Delta_k^n[D]$ we find

$$\widetilde{\alpha}_{[C]}(t) = -\lim_{n\to\infty} \frac{1}{n}\log_2\left(2^{n/2}|U_{n,k_n}|\right) = -1 + \alpha(t). \quad (57)$$

This is exactly the relation we expect between the scaling exponents of a process and its (distributional) derivative, unless the process contains more complex oscillatory behavior such as chirps [61].

Differentiating (56), we find $\widetilde{T}'_{[C]}(q) = -1 + T'(q)$, which is by Legendre transform (51) in agreement with (57). From this it becomes clear that all results on the MFA of the MWM process $D(t)$ translate directly into a scaling analysis of its distributional derivative $C(t)$. In particular, see Corollary 3, (65) and (66) below, as well as Theorems 9 and 11, and (103).

6.4 Multifractal scaling of moments and LRD

The multifractal scaling exponent $T(2)$ of a process Y is closely related to LRD parameter H, since both measure the power-law behavior of second-order statistics.[8] More precisely, $T(2)$ captures the scaling behavior of the second sample moments, while H captures the decay of the covariances.

For a process Y with *zero-mean* increments, this relation can be made precise. To this end we use the fact that H can be measured through a scaling of the sample variance as derived from (12) [2]. Therefore, let $Z[k] = \Delta_k^n[Y]$ denote the increment process of Y at some given (finest) resolution 2^{-n}. Following (11), we let then $Z^{(m)}[k]$ be the aggregated increment process, i.e. at aggregation level $m = 2^i$ the process $mZ^{(m)}[k] = \Delta_k^{n-i}[Y]$ is the increment of Y at resolution $m/2^n = 2^{i-n}$. According to (12) the variance of $Z^{(m)}$ scales as $\mathrm{var}(Z^{(m)})/\mathrm{var}(Z) \simeq m^{2H-2} = 2^{i(2H-2)}$ for an LRD process Y. On the other hand, $\mathrm{var}(Z^{(m)}) = m^{-2}\,\mathbb{E}\Big[|mZ^{(m)}|^2\Big] = 2^{-2i}\,\mathbb{E}\Big[|\Delta_k^{n-i}[Y]|^2\Big] \simeq 2^{-2i}\,2^{(i-n)(1+T(2))}$ according to (47). Comparing the scaling terms 2^i, we find that $2H - 2 = -2 + (1 + T(2))$, or

$$H = \frac{T(2)+1}{2} \quad (58)$$

for zero-mean processes. For fBm, this is in agreement with (60) below.

[8] While we may define an MFA for an arbitrary process as in (43), the interpretation in terms of Hölder continuity is valid only for increasing processes with positive increments. Moreover, here we neglect the fact that $T(2)$ is defined through a limit of arbitrary fine resolutions while LRD is an asymptotic law for large scales. In other words, we assume that scaling is perfect on all relevant scales.

Multifractal measures such as the MWM signal $C^{(n)}[k]$ are not second-order stationary. Hence, LRD cannot be defined through the decay of the auto-covariances. However, alternative fractal properties, such as the decay of aggregate variances (11) or wavelet coefficients (13) — which are equivalent to LRD in the presence of second-order stationarity — can still be defined and calculated.

As a further difficulty, processes obtained from cascades have positive increments $Z[k]$, so that the above argument using the variances has to be corrected to read $2^{-2i}\,2^{(i-n)(1+T(2))} \simeq \mathbb{E}\Big[|Z^{(m)}|^2\Big] = \mathrm{var}(Z^{(m)}) + \mathbb{E}\Big[Z^{(m)}\Big]^2 \simeq \mathrm{var}(Z)2^{i(2H-2)} + \mathbb{E}[Z]^2$, noting that $\mathbb{E}\Big[Z^{(m)}\Big]$ is independent of the scale m. Since $2H-2<0$ we may, thus, still expect the same relation (58), at least in the limit of very fine resolution (small m and i).

The variance-time plot method above is known to be an unreliable (but simple) estimator of LRD behavior [8], while the wavelet method of [37] is more robust. Since we are dealing with increment processes, we need to apply (13): $\mathrm{var}(W_{j,k}) \simeq 2^{-j(2H-1)}$. Recalling that we can obtain $T(2)$ through (55) and (56), we find by stationarity $\mathrm{var}(W_{j,k}) = 2^{-j}\,\mathbb{E}[2^j|W_{j,k}|^2] \simeq 2^{-j}\,2^{-j(1+\tilde{T}_{[C]}(2))} = 2^{-jT(2)}$, and the same relation (58) follows again in the limit to fine resolution $j\to\infty$.

Finally, checking the value $T(2)$ predicted by theory in (66), we again find agreement with (58). The same is actually true for much larger classes of cascade multifractals.

6.5 The multifractal spectrum of fBm

We now show that fBm does not possess a rich multifractal structure. Stationarity of increments and self-similarity yield immediately that

$$\mathbb{E}\sum_{k_n=0}^{2^n-1}(\Delta^n_{k_n}[B])^q = 2^n\mathbb{E}\left[|B(2^{-n})|^q\right] = 2^{n-nqH}\mathbb{E}[|B(1)|^q], \tag{59}$$

and thus

$$\text{fBm:}\qquad T(q) = \begin{cases} qH-1 & \text{for } q>-1 \\ -\infty & \text{for } q\le -1, \end{cases} \tag{60}$$

$$T^*(\alpha) = \begin{cases} -\infty & \text{for } \alpha<H \\ 1+H-\alpha & \text{for } \alpha\ge H. \end{cases} \tag{61}$$

This means that there are no values $\alpha(t)<H$ to be observed. This is somewhat in agreement with a result of Adler [62] that states that the degree of Hölder continuity[9] of fBm is H everywhere in $[0,1]$ with probability one. The formula also indicates that $\alpha(t)>H$ will be observed. This is due to the fact that the increments of fBm are zero-mean Gaussian on all scales, hence there is a considerable probability of finding small increments, i.e. large α^n_k. In other words, $\alpha^n_{k_n}$ converges very non-uniformly to $\alpha(t)=H$.

[9]Since fBm is a not an increasing process, the notion of Hölder regularity H_t we introduced in Appendix A.1 has to replace $\alpha(t)$. A wavelet-based analysis using $\widetilde{\alpha}$ and $\widetilde{S}$ usually reflects Adler's result more closely.

In conclusion, the $T(q)$ of fBm is linear, i.e. a degenerate concave function. This captures the monofractal structure of fBm in simple terms. Real-world signals such as network traffic, however, exhibit truly multifractal behavior, i.e. they possess a strictly concave $T(q)$ (see Figure 9).

6.6 The multifractal spectrum of MWM

We begin by stating a corollary to Theorem 9.

Corollary 3 *Consider an MWM as given in (25) or (38), with multipliers $A_{j,k}$ symmetrical and identically distributed within scale and independent along any line of descendants (c.f. (40)). Assume furthermore that the $A_{(j)}$ converge in distribution as $j\to\infty$. Then, we have with probability one that*

$$f_{\mathrm{G}}(\alpha) = T^*(\alpha) \tag{62}$$

on the entire interval $\{\alpha : T^*(\alpha)>0\}$, i.e. on $\{\alpha = T'(q) : qT'(q) > T(q)\}$, which corresponds to the q-interval bounded by the two values $\underline{q}$ and $\overline{q}$ where the tangent at $T(q)$ passes through the origin.

This result follows as a consequence of the work of [22, 24, 25, 32] together with (100) under the additional assumption that the $A_{(n)}$ are all identically distributed. With Theorem 11 we show in Appendix B how to generalize the argument of [24] to our case.

Let the assumptions of the Corollary be in force for the remainder of this section. Then, the multipliers M^n_k generating a binomial cascade equivalent to the MWM (c.f. (41)) are independent along lines of descendants (40). Also, they are identically distributed within scale due to the symmetry of $A_{(n)}$:

$$M^n_{k_n} \overset{d}{=} M^{(n)} := \frac{1+A_{(n-1)}}{2}. \tag{63}$$

These two facts allow the following calculation, which is the basic step towards calculating $T(q)$. We denote by $\sum'$ the sum over all $k_n = 0,\ldots,2^n-1$ and use again the notation of (22). Then,

$$\begin{aligned}\mathbb{E}[S_n(q)] &= \sum{}'\,\mathbb{E}[(M^n_{k_n})^q]\cdots\mathbb{E}\Big[\Big(M^0_0\Big)^q\Big] \\ &= \sum{}'\,\mathbb{E}\Big[\Big(M^{(n)}\Big)^q\Big]\cdots\mathbb{E}\Big[\Big(M^0_0\Big)^q\Big] \\ &= \mathbb{E}\Big[\Big(M^0_0\Big)^q\Big]\;2^n\prod_{i=1}^{n}\mathbb{E}\Big[\Big(M^{(i)}\Big)^q\Big]. \end{aligned} \tag{64}$$

Let us add now the fact that the $A_{(n)}$ (respectively $M^{(n)}$) converge in distribution to a random variable, say A (respectively $M=(1+A)/2$). Then, we find

$$\begin{aligned}\text{MWM:}\qquad T(q) &= -1-\log_2\mathbb{E}[M^q] \\ &= q-1-\log_2\mathbb{E}[(1+A)^q]. \end{aligned} \tag{65}$$

As an example, consider the βMWM defined in Proposition 2 with symmetrical β multipliers $A_{(n)}$. Since the variance of the multipliers converges by (36), so does the only parameter $p_{(n)}$ and, hence, the whole distribution. The limiting random variable A has the standard symmetrical β distribution, supported on $[0, 1]$. Its parameter is $p = (2^{2H-1} - 1)/(2 - 2^{2H-1})$ for $1/2 \le H < 1$ by (37). Using the well-known formula for the moments of a β distribution we find for $q > -p$

$$\beta\text{MWM:} \qquad T(q) = -1 - \log_2 \frac{\Gamma(p+q)\Gamma(2p)}{\Gamma(2p+q)\Gamma(p)}, \tag{66}$$

with $T(q) = -\infty$ for $q \le -p$. For the point mass (δ)-distribution (33) the obvious formula results using (34):

$$\delta\text{MWM:} \qquad T(q) = q - 1 - \log_2 \left[(1-2r) + r\Big((1-c)^q + (1+c)^q\Big)\right].$$

Having now two parameters available provides more flexibility. This will be used in Section 7.1 to match not only the energy decay, i.e. $T(2)$ as is done with the βMWM, but also the first negative moment, i.e. $T(-1)$. Fitting negative moments results in better matching of small values. These correspond to large α (43), i.e. to negative q and the decreasing part of T^* (c.f. Figure 9).

More generally, in a mixture model the moments are convex combinations of the moments of the mixing distributions. Thus, T is readily available for such cases using (65). The additional parameters introduced in this way allow for even greater flexibility.

In conclusion, the partition function $T(q)$ displays a diverse array of statistical properties of a signal in a concise way. The parameters of the MWM, however, should not be looked for among the $T(q)$ but rather among the parameters of the underlying distributions of the multipliers.

7 Application to Network Traffic

Let us now turn to a problem of considerable current practical interest — computer network traffic modeling. Data traffic models are an invaluable asset to the network analyst. In network analysis, model parameters are used to capture and summarize important characteristics of data traffic. With simple models, the impact of various parameters on network performance can be studied through analytical means [6, 63–67]. In cases where theoretical analysis is intractable, models are routinely used to synthesize test data traces for simulation purposes [68]. Here, computational efficiency of the synthesis becomes as important as the accuracy.

We begin with some historical remarks. Although LRD models have long been known to characterize a variety of phenomena, only recently has LRD been discovered in data network traffic [5]. This has lead to new insights about traffic and network performance [5], primarily that high levels of LRD lead to poor network performance and that classical models like Markov and Poisson processes are too optimistic in their performance predictions. As a consequence, incorporating LRD in traffic models for network analysis has lead to more realistic results, and self-similar models like fGn have been suggested for modeling LRD traffic.

Norros [66] surveys the theoretical bounds for the queuing performance of self-similar traffic. Here, the total traffic arriving up to time t is modeled by

$$Z(t) = \lambda t + \sqrt{a\lambda}\, B_H(t), \tag{67}$$

where B_H is fBm (with Hurst exponent H and $\mathrm{var}(B_H(1)) = 1$), and a and λ are constants. In other words, the incoming traffic $Z(t+h) - Z(t)$ is assumed to arrive with a mean rate λ superimposed on a colored Gaussian noise (fGn) process. The parameter a controls the overall variance.

The successes of self-similar models such as (67) have lain mainly in their ability to capture LRD while permitting tractable theoretical analysis. However, self-similar models like fBm/fGn have three severe drawbacks: (1) Gaussian marginals, meaning the process must take negative values, (2) computational inefficiency for exact synthesis, (3) degenerate multifractal properties. While the first two clearly limit the use of self-similar models for synthesis, it is the object of ongoing research to establish the importance of the third for queuing performance. The MWM exhibits power spectra, marginals, and multifractal behavior consistent with actual traffic while providing an $O(N)$ synthesis algorithm for N-point output traces.

In this section, we synthesize network traffic data by training the MWM on real data. This data-fitting exercise demonstrates the accuracy of the model not only in statistical terms (multifractal properties) but also through queuing experiments. Though we are not claiming to present a physical model for network traffic, the close fit of the multiplicative process underlying the MWM to the real data provides valuable insight into the mechanisms of buffering and multiplexing of network traffic.

Interesting quantities for simulation include packet interarrival times, packets-per-time, and bytes-per-time. Packet interarrival times can be converted directly into packets-per-time by binning the packet arrivals into time bins of the required size, whereas bytes-per-time includes the additional information of packet size. Here, we train on a set of $1/f$-like packet interarrival time data, since interarrival times, being continuous-valued, are most natural for the MWM. In addition, analysis of interarrival times avoids the problem of choosing an appropriate time unit as in packets-per-time and bytes-per-time. However, we could as well apply the MWM to approximate discrete-valued packet-per-time or bytes-per-time. For these cases, we could quantize the MWM's continuous-valued output into discrete-valued data or follow the approach of [45].

7.1 Synthesis via matching

7.1.1 Real data

We focus on the August 1989 Bellcore Ethernet trace *pAug* of 10^6 interarrival times (Figure 2(a)), as measured by Leland et al. [5]. Although slightly dated, this data set provides a well-known benchmark useful for examining the fractality and LRD of network traffic.

First, we analyze the properties of the trace. Recognizing its limitations as an LRD estimator, we use the variance-time plot (Figure 7) to obtain a qualitative characterization of the correlations present in the data. From the plot, we find the trace exhibits LRD with $H \approx 0.79$. Since the plot is somewhat "kinked," the trace most likely does not exhibit a strict second-order scaling. As Figure 2 plainly shows, modeling $pAug$ as an fGn process with $H = 0.79$ and the same mean and variance leads to nearly 30% of the synthesized data being negative. The culprit is the large standard deviation to mean ratio of 1.8 of $pAug$. The oft used but ad hoc procedure of setting all negative points to zero would clearly result in a process with very different statistics to those required. In general, fGn models are of limited utility for positive data with small mean and large variance.

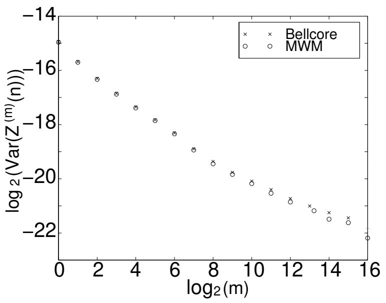
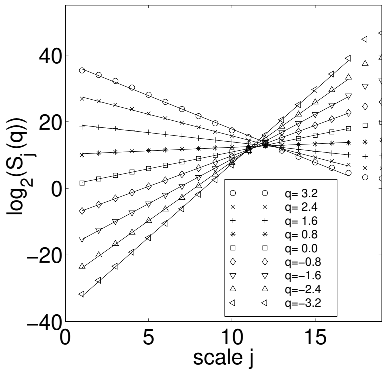
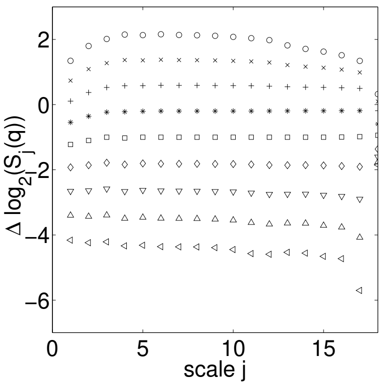


Figure 7: Variance-time plot of the Bellcore $pAug$ data "×" and one realization of the βMWM synthesis "○". Here m denotes the level of aggregation and $Z^{(m)}(n)$, the aggregated process defined through (11).



(a)



(b)

Figure 8: (a) Scaling of log-moments $\log_2(S_j(q))$ vs. scale j for the 10^6 Bellcore interarrival times $pAug$ with q ranging from -3.2 to 3.2 and j ranging from 1 to 19, with $j = 19$ the finest scale. (To compare with Figure 7, note $\log_2(m) = 19 - j$). (b) Increments of the log-log scaling shown in (a). The closeness of the linear fits in (a), as indicated by the stable behavior of the increments in (b), indicates that the interarrival times are indeed multifractal.

Moving beyond second-order statistics, we measure the multifractal properties of $pAug$. As discussed in Section 6.2 we estimate $T(q)$ as the slope of a linear fit of the log-log plot of the sample moments $S_j(q)$ at resolution 2^{-j} against the scale j (52). In Figure 8(a) the only noticeable deviation from linearity is at the very finest resolution of analysis — a fact that is enhanced in Figure 8(b), where the increments of the log-log plot are displayed. With 16 octaves (5 decades) of excellent scaling, we can be confident in concluding that $pAug$ is multifractal.[10] The only noticeable deviation from linearity is at the very finest resolution of analysis. The linearity of the log-log plots can be more closely verified in Figure 8(b), which displays the increments of the log-log plot from Figure 8(a).

Extracting $T(q)$ using (52) from Figure 8 and applying the Legendre transform (51), we obtain the multifractal spectrum $f_G(\alpha)$ of Figure 9. As indicated by the multifractal formalism (see Section 6.1.3, Corollary 3, and Theorems 6 and 9), this function gives the large deviations from the "most frequent" singularity exponent and thus displays valuable information about the occurrence of rare events such as bursts (small α). Figure 9 reveals a rich multifractal spectrum. In contrast, fBm has a trivial spectrum consisting only of one point indicating that it has the same "burstiness" $\alpha(t) = H$ everywhere [70].

7.1.2 Synthetic data

Having established the LRD and multifractal characteristics of the $pAug$ trace, we will next model these properties using the βMWM. To train the βMWM, we use the approach outlined in Section 4.6. We choose the number of wavelet scales $n = 16$ to synthesize data sets of 2^{16} points. This allows us to collect multiple realizations of the wavelet coefficients and root scaling coefficient, and thus form reliable mean and variance estimates. For the root scaling coefficient, we use the Gaussian assumption discussed in Section 4.4.

With trained βMWM in hand, we synthesize 15 length-2^{16}

[10] Since we characterize traffic interarrival times, our result does not conflict with that of [69], which concluded that the *bytes-per-time* and *packets-per-time* of the August 1989 Bellcore traces were not multifractal. Multifractal scaling of similar quality over 5 decades has been reported for several TCP traces in [19].

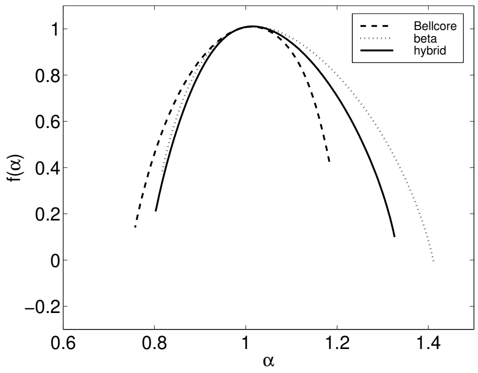


Figure 9: Multifractal spectra (51) of the Bellcore *pAug* data, βMWM synthesis, and a hybrid MWM employing beta distributions at coarse scales and point masses at fine scales. The spectra were obtained through the Legendre transform of the scaling of the moments (see Figures 8 and 10). The close match in the upper left part, which corresponds to q values (=slopes of tangents to the spectrum) between 0 and 2, indicates that the βMWM matches these low (qth) order moments very well. The divergence of the spectra on the right indicates that the chance of observing large α in the βMWM data is somewhat too high. This behavior is improved significantly by adding point mass multipliers in the fine scales.

sub-traces and concatenate them to form a trace of approximately length-10^6, the size of the real data set. We now apply the same battery of tests to this trace as we applied to the actual Bellcore *pAug* data. Figure 2(b) shows that the synthesized data captures much of the gross structure of the Bellcore data at different aggregation levels, including the one-sided marginal density. In addition, the variance-time plots of Figure 7 depict an excellent match of the correlation structure.[11]

We next measure the multifractal properties of the synthetic trace. ¿From the linearity of the log-log plots in Figure 10(a), we see that the synthetic trace exhibits a multifractal scaling, except for q strongly negative and j large. In converting these plots into the multifractal spectrum of Figure 9, we see that spectrum of the synthesized data closely matches the *pAug* spectrum for α near one. The close match in the upper left part, which corresponds to q values (=slopes of tangents to the spectrum) between 0 and 2, indicates that the βMWM matches these low (qth) order moments very well. The divergence of the spectra on the right indicates that the chance of observing large α in the βMWM data is somewhat too high. Since large α correspond to fast decay, this means that the βMWM trace has values that are too small. In fact, the minimum value of the wavelet-synthesized trace is on the order of 10^{-12}, whereas the minimum of *pAug* is on the order of 10^{-5}. This is due to the fact that, unlike the coarser-scale β multipliers $A_{j,k}$, the fine-scale β multipliers have pdfs with significant mass near ± 1. Clearly, from (25) we see that this results in small values for the synthesized process $C^{(n)}[k]$. This may be indicative of different phenomena in the fine scales of the real data as compared to the coarse scales.

Using β distributions in the coarse scales and point mass distributions in the fine scales, we can largely correct this problem, synthesizing data with a minimum value of 10^{-6} while preserving the other features of the βMWM (see Figure 9). We choose the point mass parameters (see Section 4.3.2) to match both the wavelet energy decay and the scaling of the negative first moment of the real data in (29). We do not claim that the point mass multipliers are realistic — using point mass multipliers at all scales results in syntheses that look somewhat artificial. Here, we simply illustrate the fact that we can choose the multiplier distributions to better match higher-order or lower-order moments of the data.

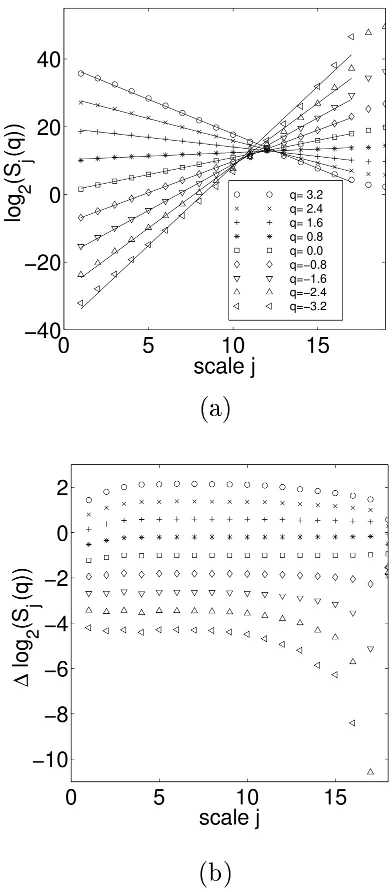
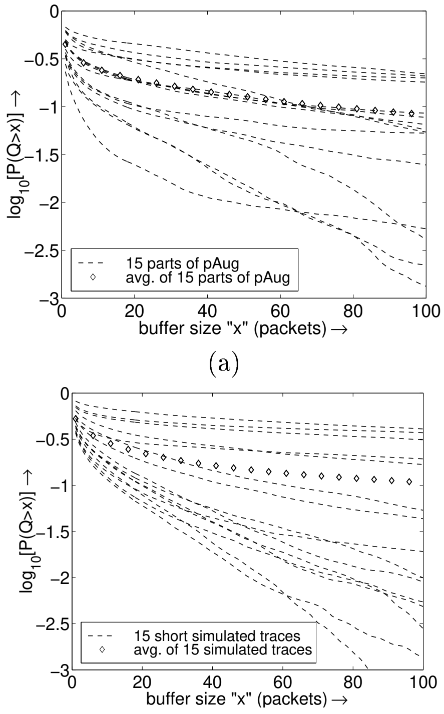


Figure 10: (a) Log-log moment scaling and (b) incremental scaling for the βMWM synthesized data. (See Figure 8 for more description.) The synthetic data exhibits a linear multifractal scaling, with the exception of strongly negative q's and large j.

7.2 Queuing behavior

As a final test of the accuracy of the match of the βMWM to the *pAug* target data, we now compare their queuing behaviors. The queuing behavior of traffic is important because of its influence on network management algorithms, such as

[11]We remind the reader that the variance-time plot must be interpreted with care due to the non-stationarity of the wavelet-synthesized data.



(b)

Figure 11: Here we partition the pAug trace into 15 sub-traces of equal number of packets and compare their queuing behavior with that of 15 synthesized traces of the same length. In (a), observe that the real sub-traces have a wide variation in tail queue behavior. In (b), observe that the synthesized traces display a similar variation in tail queue behavior.

connection admission control, that strive to support certain quality of service (QoS) demands [71, 72].

The presence of LRD in traffic has been shown to significantly affect queuing performance [65]. For stationary traffic with only short-range dependence (SRD), classical queuing results for Markov models show that the tail of the distribution of the queue-length in a single server queue with deterministic service satisfies

$$\Pr[Q > x] \simeq e^{-\delta x}, \tag{68}$$

where the positive constant δ depends on the service rate at the queue and the statistical properties of the arrivals process. Unlike (68), fBm-based models for LRD traffic exhibit Weibull tail distributions of the form

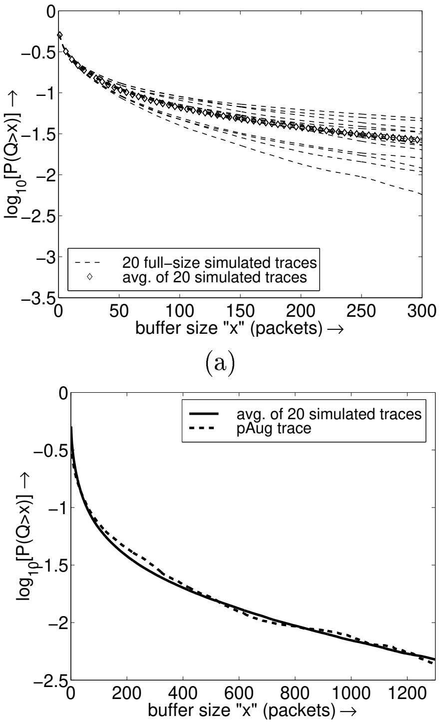
$$\Pr[Q > x] \simeq e^{-\delta x^{2-2H}}, \tag{69}$$

where H is the Hurst exponent [6,63,73]. Clearly, we see from (68) and (69) that the tail queue probability of self-similar traffic decays at a much slower rate than that of SRD traffic. With the LRD of Ethernet traffic being established beyond doubt, it is important for traffic models to incorporate LRD, without which the prediction of queuing performance can be overly optimistic. However, as mentioned earlier, fGn's Gaussian marginals makes it unsuitable for the *pAug* data set; it is meaningless to perform queuing experiments with the data of Figure 2(c).

In the simulations that follow, we consider the performance of an infinite-length single server queue with a single trace as input. We assume a constant service rate of 500 packets/sec. For simplicity we assume all packets to be of equal size.

The ideal experiment comparing the queuing behavior of real world and synthesized traces would be to compute the average tail queue behavior of several realizations of the real *pAug* process as well as several realizations of the βMWM. Unfortunately, typically only one realization of the real trace is available. To circumvent this setback, we partition the real *pAug* trace into 15 sub-traces each of length 2^{16} packets and assume that each sub-trace is an independent realization of the underlying real process. We compare the queuing performance of these *pAug* sub-traces against 15 synthetic traces obtained from the βMWM in Figure 11. Note the similarly widely varying performance of both the real and synthetic traces. This result indicates that we should expect such variations and should be cautious drawing conclusions from the average tail queue behavior.

We next compare the queuing performance of the entire



(b)

Figure 12: Comparison of the queuing behavior of *pAug* with 20 full-size synthesized traces. Displayed are the tail probabilities of buffer occupancy vs. buffer size. In (a) observe the variability of the queue performance of the synthesized traces. In (b) observe that the average queue performance of simulated traces and that of the real trace match closely.

pAug trace with that of 20 traces of approximately the same length (10^6 points) generated using the βMWM (see Figure 12). The simulated traces in Figure 12(a) exhibit a wide variation in tail queue behavior. The results of the previous experiment indicate that this is to be expected. We also observe that the average tail queue behavior of the simulated traces matches that of the real trace surprisingly well (see Figure 12(b)). However, as the previous experiment suggests, the real data cannot be expected to always exhibit the same queuing behavior as the average of several simulated traces.

In summary, these queuing experiments demonstrate that our βMWM synthesized traffic traces not only match real traffic in terms of its various statistical properties but also in its queuing behavior.

7.3 Physical interpretation

We have argued for the use of the MWM for synthesizing network traffic in terms of statistical properties (see Sections 7.1 and 7.2). The quality of the matching challenges the current understanding of networking and performance analysis by suggesting that some of the mechanisms shaping the traffic flow might carry an inherent multiplicative structure. Our motivation for providing a possible explanation for the presence of multiplicative mechanisms is twofold. First, we hope that multiplicative models will inspire research in networking and trust that they will lead to a deeper understanding of the forces shaping traffic characteristics. Promising steps in this direction have already been made in [20, 33, 38]. Second, such an explanation will further support the use of the MWM network traffic synthesizer.

It is generally agreed that today's network traffic is created by a large number of independent individual sources. A simple but powerful model assumes that these sources switch between two states, the "ON" state in which they produce traffic at constant rate and the "OFF" state in which they are silent. Aggregating these traffic loads yields the total traffic load observed at, say, a gateway. With this model, heavy-tailed ON periods lead to LRD similar to that observed in actual traffic. Convincing modeling results have made a strong case for this point-of-view [74, 75]. However, ON/OFF models are accurate only in the limit of *large time scales* (seconds and longer), and they do not account for the actual queuing and multiplexing occurring in the network.

A complete description of data network traffic requires understanding of its dynamic nature over not just large but also *small time scales* (hundreds of milliseconds and shorter). The flow of packets over fine time scales is shaped mainly by the protocols and end-to-end congestion control mechanisms (e.g., TCP) that regulate the complex interactions between the different connections on a network. Indeed, it is not hard to see that buffering and multiplexing can create bursts, for instance, when packets arrive at a server at a moderate rate, rest queued up, and then race off at the service rate. Since the traffic rate is strictly positive, this kind of short-term volatility (spiky nonGaussian behavior) cannot come from an additive process.

The MWM matches this small-scale behavior of traffic. Rather than modeling the traffic rate as an additive superposition of components, we model it as a multiplicative partitioning of the *rate of traffic flow*. The coarse scaling coefficient $U_{0,0}$ provides the mean traffic rate (or equivalently its inverse, the mean interarrival time) and the multiplications by $1 \pm A_{j,k}$ at each scale (c.f. (21)) provide perturbations in the arrival rates due to the effects of network phenomena at different time scales, such as speed-ups and delays due to traffic protocols, interference from competing traffic, and the like.

When trained on real network data, the behavior of the multipliers $A_{j,k}$ changes with scale, with extremely low variance at coarse scales and high variance at fine scales. Amazingly, this is consistent with both the small-scale behavior of actual traffic *and* the large-scale properties of the ON/OFF model. At fine scales, as we have already seen in Sections 5.2 and 7.1, multiplicative schemes with large variances produce bursts like those in real data (recall Figure 2). At coarse scales, the scaling coefficients (which correspond to the arrival times of large amounts of traffic) involve only a handful of low-variance multipliers $A_{j,k}$. From (25) we can write, for example, at the third-coarsest scale:

$$\begin{aligned} U_{2,k} &\overset{fd}{=} \frac{U_{0,0}}{2}\left(1+A_{(0)}\right)\left(1+A_{(1)}\right) \\ &\overset{fd}{\approx} \frac{U_{0,0}}{2}\left(1+A_{(0)}+A_{(1)}\right) \end{aligned} \tag{70}$$

Thus, for a fixed $U_{0,0}$ at the coarsest scale, to a first-order approximation, the MWM is *additive* at the coarse scales provided the random variables $A_{(i)}$ are small in amplitude. Moreover, the $A_{(i)}$ are approximately *Gaussian* for these low-variance (high-p) symmetric β multipliers [48]. Hence, coarse-resolution MWM outputs will exhibit an additive, Gaussian-like behavior consistent with that of the previously justified ON/OFF models and notions of client behavior as a superposition of sources.

Of course, this is not a rigorous physical development of how and why this multiplicative procedure takes place in reality. However, our preliminary results are promising and suggest where to look for multiplicative cascades: on small time scales, most likely in the TCP flow-control layer.

8 Conclusions

The multiplicative wavelet model (MWM) combines the power of multifractals with the efficiency of the wavelet transform in a flexible framework natural for characterizing and synthesizing positive LRD data. As our numerical experiments have shown, the MWM is particularly suited to the analysis and synthesis of network traffic loads. In addition, the MWM could find application in areas as diverse as financial time-series characterization, geophysics (using 2-d and 3-d wavelets and quadtrees and octtrees), and texture modeling. Several extensions to the model hold promise:

1. A parametric characterization of the wavelet-domain energy decay (rather than the current empirical variance measurements) would yield a more parsimonious and robust model.
2. The choice of β-distributed wavelet multipliers $A_{j,k}$ is not essential. As illustrated by our preliminary work with point mass distributions, we can use distributions with more parameters to match both wavelet energy decay and the scaling coefficient moments.
3. To model correlations in the wavelet-domain, we can introduce dependencies between the wavelet multipliers (for example, in their signs).
4. Instead of tackling the increments process directly, we could use the MWM as a model for an underlying Poisson intensity process (analogous to the work of [44]). This could be useful for fitting network traffic packets- or bytes-per-time, which are discrete-valued LRD processes.
5. Insights from the multifractal theory can be leveraged into more general (e.g., stationary and non-dyadic) multiplicative constructions.

Clearly we have not exhausted the possibilities of multiplicative multiscale modeling.

A Key concepts of Multifractal analysis

In this section we will make rigorous the points left vague in Section 6.

A.1 Introduction

The erratic behavior of a continuous process $Y(t)$ at a given time t can be characterized to a first approximation by comparison with an algebraic function. The *degree of local Hölder regularity* H_t is the best (largest) h such that there is a polynomial P_t such that $|Y(s)-P_t(s)| \le C|s-t|^h$ for s sufficiently close to t. If P_t is a constant, i.e. $P_t(s) = Y(t)$, as is the case with cascades, then

$$H_t = \liminf_{\varepsilon\to 0} \frac{1}{\log_2(2\varepsilon)} \log_2 \sup_{|s-t|<\varepsilon} |Y(s)-Y(t)|. \tag{71}$$

Fortunately, we can replace the supremum by $Y(t+\varepsilon) - Y(t-\varepsilon)$ for processes with positive increments. Furthermore, using the notation of (22), i.e.

$$t \in [k_n 2^{-n}, (k_n+1)2^{-n}), \tag{72}$$

we can then simplify by noting that $Y((k_n+2)2^{-n}) - Y((k_n-1)2^{-n}) \le Y(t+\varepsilon)-Y(t-\varepsilon) \le Y((k_{n-2}+2)2^{-n+2}) - Y((k_{n-2}-1)2^{-n+2})$ provided n is chosen such that $2^{-n+1} \le \varepsilon < 2^{-n+2}$. In summary,[12]

$$H_t = \liminf_{n\to\infty} H^n_{k_n} \tag{73}$$

$$H^n_{k_n} := -\frac{1}{n}\log_2 |(\Delta^n_{k_n-1} + \Delta^n_{k_n} + \Delta^n_{k_n+1})[Y]|.$$

Traditional *multifractal analysis* (MFA) of multiplicative cascades aims to describe the singularity structure of processes through the simpler but more restrictive[13] exponent $\alpha(t)$ from (43).

As mentioned earlier, for fBm we find $H_t = H$ for all t almost surely; this process has a degenerate multifractal structure. For the binomial measure, on the other hand, H_t and $\alpha(t)$ will depend crucially — and discontinuously — on t. To convince yourself, recall the iterative process of Section 5.1 and descend first down in the cascade to a point t by following always the smaller of the two multipliers. Then descend by following always the larger one. The decay rate of the increment D_b (38), i.e. $\alpha(t)$, will differ drastically in the two cases.

For a measure constructed using a cascade, i.e. $Y = D_b$, the range of $\alpha(t)$ will always be a positive interval containing the value 1. Values $\alpha(t)$ smaller than 1 correspond to points where D_b is not differentiable. If $\alpha(t) > 1$, on the other hand, then $D_b'(t) = 0$, i.e. D_b behaves at t like the function x^α at $x = 0$. A typical range of $\alpha(t)$ for a real-world signal might be $[0.6, 2]$ or $[0.8, 1.2]$.

The MFA structure can be given either in geometrical or statistical terms. [35, 55]. Here, we will be mainly interested in the statistical description.

Before going into details let us note a simple fact about the occurrence of $\alpha(t)$ for the *deterministic binomial* D_b. In this special case, all multipliers $M^n_{k_n}$ (see Section 5.1 and Appendix B.1) are deterministic, i.e. we assume that there are two fixed numbers m_0 and m_1 that add to 1 and that $M^n_{k_n} = m_{k'_{n-1}}$ almost surely. Referring to Figure 5 a step in the iterative construction amounts now to splitting the area of a region in the fixed proportions "the m_0-th part on the left, the m_1-th part on the right".

D_b being deterministic, we consider now t to be random in order to apply a limiting theorem from probability theory. Recall that (22) uses the binary digits k'_i for t (c.f. (72)). Choosing these digits to be 0 or 1 with equal probability amounts to picking the point t randomly with a uniform distribution. The LLN then implies that for almost all t

$$\alpha^n_{k_n} = -\frac{1}{n}\sum_{i=1}^{n} \log_2 m_{k'_i} \to \mathbb{E}_t[-\log_2 m_{k'_i}], \tag{74}$$

hence,

$$\alpha(t) = -\frac{1}{2}\left(\log_2(m_0) + \log_2(m_1)\right). \tag{75}$$

Note that this limiting value is strictly larger than 1 unless $m_0 = m_1 = 1/2$. Consequently, the deterministic binomial

[12] For general processes this does not hold. A multifractal analysis (MFA) with this simplified version will result in a different description of the singularity behavior of the process that can, nevertheless, provide useful information [32, 70]. If a process has both positive and negative increments, then the continuous-time supremum in the original version of H_t (71) cannot be estimated numerically. In this case, the wavelet modulus maxima method provides arguably the most accurate information on local Hölder regularity [76]. Adapted to detecting singularities of oscillating functions, on the other hand, wavelets have a disadvantage in the MFA of positive increment processes: they are not efficient for detecting large values of α that correspond to more regular parts in the process. This is why we restrict the discussion to positive increment processes and the simplified version of H_t.

[13] As we note later, replacing $H(t)$ by $\alpha(t)$ does not change the outcome for cascades.

measure has zero derivative at almost all points t. This brings home a point made in Section 5.2: the distribution $D_b(t) = \mu([0,t]) = \Pr[x \le t]$ associated with the binomial measure has no density, for if it had one it would have to equal zero. Again in other words, we cannot write $D_b(t)$ as $\int_0^t D_b'(s)\,ds$, since the latter is zero for all t.

Usually, one is happy with an "almost sure" result such as (74). Here, we would like to ask two additional questions: (1) can there be points t with $\alpha_{k_n}^n$ converging to a number different from (75), and (2) if so, what can we say about such points t? Indeed, we find immediately that at $t = 0$ we have $\alpha(0) = \log_2 m_0$. Actually, we will find the same limit at all dyadic points t, since their dyadic expansion shows only finitely many 1's. This certainly justifies our quest.

A.2 The Multifractal Spectra

A.2.1 Hausdorff spectrum f_{H}

Ideally, we would like to quantify the values and frequencies of limiting $\alpha(t)$. In other words, we are interested in the "sizes" of the sets

$$K_\alpha = \{t : \alpha(t) = \alpha\}. \tag{76}$$

This is the geometrical approach to MFA. For fBm, replacing $\alpha(t)$ by the more appropriate H_t, K_α is either the whole line (if $\alpha = H$) or empty. Consequently, fBm is said to be "monofractal", since it has only one fractal scaling exponent. The concatenation of K fbm-s Y^i with Hurst exponent H^i in the interval $[i/K, (i+1)/K[$ would form a process with $K_{H^i} = [i/K, (i+1)/K[$.

For more general processes, the sets K_α are highly interwoven and each of them may lie dense on the line. Consequently, the right notion of "size" is that of the fractal Hausdorff dimension, which leads us to defining the *Hausdorff multifractal spectrum*:

$$f_{\mathrm{H}}(\alpha) := \dim(K_\alpha). \tag{77}$$

Unfortunately, Hausdorff dimensions are impossible to calculate numerically in any real-world situation, and we have to rely on the multifractal formalism (coming up next) (100) and (104) to estimate f_{H} under certain assumptions.

For a definition of fractal dimensions, see [29,32,35]. Here, we only mention that $\dim(E)$ is a positive real number, and the larger it is the "larger" the set E. We explain this notion of "largeness" by comparing a plane and a line. Though a plane and a line have integer dimensions, our methods can be generalized to broken, or fractal dimensions. First, note that a randomly selected *probe* line in space will most likely intersect a given plane, but not a given line. For random fractals this generalizes to: a random *probe* fractal will intersect a second given fractal only if their fractal dimensions add up at least to the dimension of the embedding space. Second, a plane has more degrees of freedom than a line, i.e. a square can be segmented into $\sim \delta^{-2}$ pieces of size δ, an interval only into $\sim \delta^{-1}$. A fractal will ideally partition into $\delta^{-\gamma}$ pieces of size δ where γ is its fractal dimension.

A.2.2 Large deviation spectrum f_{G}

In practice, measurement of the "burstiness" of a process has to rely on numerically more accessible methods and notions than f_{H}. Enter the statistical description of multifractal structure. To this end we consider a histogram of the $\alpha_{k_n}^n$'s taken at some finite level n. (Recall (74) for a formula of $\alpha_{k_n}^n$ for the deterministic binomial measure.) The histogram will show a non-trivial distribution of values that increasingly concentrates around the expected value (75) due to the LLN: values other than the expected one must occur less and less often.

It is here that Large Deviation Principles (LDP) [58,59] turn out to be invaluable. As a generalization of the Chernoff-Cramer bound [77, Thm. 9.3], which we present below, LDPs suggest that probabilities of rare events decay exponentially fast. For a sequence of iid random variables W_n with $\mathbb{E}[W] < a$ and $\Pr[W > a] \neq 0$, set $V_n := W_1 + \ldots + W_n$. Then, we find for all $q > 0$ that

$$\begin{aligned} \Pr[(1/n)V_n \ge a] &= \Pr\Big[2^{qV_n} \ge 2^{nqa}\Big] \\ \le \mathbb{E}[2^{qV_n}]2^{-nqa} &= \Big(\mathbb{E}[2^{qW}]2^{-qa}\Big)^n. \end{aligned} \tag{78}$$

Here we have used the Tschebischev inequality and in the last step the iid property. It follows that

$$\begin{aligned} \frac{1}{n}\log_2 \Pr[(1/n)V_n \ge a] &\le \inf_{q>0}\left(\frac{1}{n}\log_2 \mathbb{E}[2^{qV_n}] - qa\right) \\ &= \inf_{q>0}\Big(\log_2 \mathbb{E}[2^{qW}] - qa\Big). \end{aligned} \tag{79}$$

Theorems on LDPs generalize such results to arbitrary sequences V_n and show when the bound is sharp in the limit $n \to \infty$ [59]. For our purposes, we set

$$V_n := \log_2 \Delta_{k_n}^n[Y], \tag{80}$$

yielding $\alpha_{k_n}^n = V_n/n$ as desired. In the special case of the random binomial or MWM, V_n can indeed be written as a sum as above with $W_n = \log_2 M_{k_n}^n$ (c.f. (38) and (74)).

It is important not to confuse the randomness relevant for the LDP with the randomness in Y. Here, we explicitly fix one realization (or path) of Y. Then, we consider the location t, encoded by k_n, as the only randomness relevant for the LDP. Since k_n can take only 2^n different values that we assume to be equally likely, probabilities in t are calculated by simple counting.[14] As we have just learned, we can expect an exponential decay of "rare event probabilities" such as (79). In other words there is reason to hope that the limiting "rate function" f_{G} we introduced in (46) and called *coarse grained multifractal spectrum* will exist:

$$f_{\mathrm{G}}(\alpha) = \lim_{\varepsilon\to 0}\lim_{n\to\infty}\frac{1}{n}\log_2 N_n(\alpha,\varepsilon) \tag{81}$$

with

$$N_n(\alpha) := 2^n \Pr_t\Big[\alpha_{k_n}^n \in (\alpha-\varepsilon, \alpha+\varepsilon)\Big] \tag{82}$$

$$= \#\{\alpha_{k_n}^n \in (\alpha-\varepsilon, \alpha+\varepsilon)\}. \tag{83}$$

[14] To avoid confusion, we will write $\Pr_t$ and $\mathbb{E}_t$ to designate randomness with respect to the position t and $\Pr_\omega$ and $\mathbb{E}_\omega$ to designate randomness with respect to the process Y.

(The factor 2^n is added for convenience.) This rate function f_G is defined (provided the limit exists) for every path of Y and is, hence, random, i.e., a function of ω.

The counting in (83) relates to the notion of dimension: if $f_G(\alpha) = 1$, then all or at least a considerable part of the $\alpha_{k_n}^n$'s are approximatively equal to α. More precisely, $N_n(\alpha) \simeq 2^n$. Such is the case for fBm, with $\alpha = H$ almost surely (see [32]). Furthermore, if a certain constant fraction of $\alpha_{k_n}^n$'s equal α, we have $f_G(\alpha) = 1$ almost surely, as is the case for the concatenation of fBm's described above (see also [70]).

Only if certain values of $\alpha_{k_n}^n$ are considerably more spurious than others will we observe $f_G(\alpha) < 1$. To draw again an analogy, let us assume for a moment that t is a vector in 3-d space. The maximum of f_G in this case will be at the expected α-value with $f_G = 3$. If the points t where $\alpha_{k_n}^n$ is approximately equal to a given α build a surface (spurious in 3-d space), then $f_G(\alpha) = 2 < 3$. If they fill a curve only, then $f_G(\alpha) = 1$. So, there is hope that $f_G(\alpha)$ relates to $\dim(K_\alpha)$. Indeed it can be shown that [29, 32]

$$f_H(\alpha) \le f_G(\alpha) \tag{84}$$

for every path.

A.2.3 Legendre transform

In Large Deviations, the transform that appears on the right side of (79) plays an important rôle. Let $g(a)$ be any function and define its *Legendre transform* g^* by

$$g^*(q) := \inf_{a\in\mathbb{R}} (aq - g(a)). \tag{85}$$

Let us assume first that g is concave at a_0, by which we mean that there is a linear function $s(a) = aq_0 + r$ such that $g(a) \le s(a)$ with equality in a_0 (c.f. Figure 6). This situation is particularly well-suited for the Legendre transform and allows us to compute $g^*(q_0)$. Note that there might be several q_0 meeting the requirements. We claim that $g^*(q_0) = -r$. But this follows from the fact that $aq_0 - g(a) \ge aq_0 - s(a) = -r$ for all a with equality at $a = a_0$. Moreover, we actually found that

$$g^*(q_0) = a_0 q_0 - g(a_0). \tag{86}$$

There is some general wisdom to this: given q, the Legendre transform finds the best linear function s of slope q that lies above the function g. The intercept of s with the ordinate axis is $-g^*(q)$.

If we assume now that g is concave and in addition differentiable at a_0, then there can be only one linear function $s \ge g$ with $s(a_0) = g(a_0)$. We find the value of g^* at $q_0 = g'(a_0)$ to be (c.f. (51)):

$$g^*(g'(a_0)) = a_0\, g'(a_0) - g(a_0). \tag{87}$$

For example, the function $g(a) = -|a-1|+2$ is concave in all points, but it is not differentiable at $a = 1$. Its Legendre transform is easily computed: for $-1 \le q \le 1$ we may choose $a_0 = 1$ and obtain $g^*(q) = q - 2$ by (86). For other q we find $g^*(q) = -\infty$ by applying the definition. Remarkably, the Legendre transform of g^* gives g back. Indeed, $\inf_q(qa - g^*(q)) = \inf_{-1\le q\le 1}(q(a-1)+2) = g(a)$. More generally, we will establish that $g^{**} := (g^*)^*$ equals g, for every concave function g.

To prove this, let us show first that g^* is a concave function provided g is. Indeed, $g^*(q) \le aq - g(a)$ for all q and a by the definition of g^*. Now let us fix a, say, at a_0. Then, $s(q) = a_0 q - g(a_0)$ is a linear function that is larger than g^* and we have, in the notation of (86), $s(q_0) = g^*(q_0)$. We conclude that g^* is concave in q_0. Moreover, we see that $g(a_0) \le a_0 q - g^*(q)$ (still by the definition of g^*), with equality at q_0. But this means nothing more than $g^{**}(a_0) = g(a_0)$. Finally, it is not difficult to see that there is an a_0 (which may lie at $\pm\infty$) as in (86) for every q_0 with $g^*(q_0) \ne -\infty$. Consequently, g^* is concave everywhere.

We continue by noting that g^* is *always* a concave function. The reason is simple: there is a concave function $\overline{g}$ such that its graph is the concave hull of the graph of g. Since $\overline{g}$ and g have the same Legendre transform, i.e. g^*, the claim holds. However, $\overline{g}$ being concave, the above argument shows that applying the Legendre transform to g^* will bring us back to $\overline{g}$, which is in general different from g. In summary:

Lemma 4 *The Legendre transform g^* of any function g is concave. Moreover, $g^{**} = \overline{g}$.*

Since concave functions are necessarily continuous and almost everywhere differentiable, we might wonder what the edges of g^* correspond to. As the example $g(a) = -|a-1|+2$ above shows, points of linearity of g (respectively $\overline{g}$ if g is not concave), correspond to points of non-differentiability of g^* and vice versa. While this situation holds quite generally, it is instructive to verify it assuming that g is C^2 and strictly concave ($g''(a_0) < 0$) at a_0: Using the implicit function theorem, we find indeed that g^* is then differentiable at $q_0 = g'(a_0)$ with $(g^*)'(q_0) = a_0$.

A.2.4 Legendre spectrum f_L

The spectrum f_G, though numerically accessible, is hard to estimate directly on real-world data, in particular because of the double limit in (81). Here, the Legendre transform in combination with (79) proves useful. Due to the simple distribution of t as used in the LDP, the moment generating function $\mathbb{E}\left[2^{qV_n}\right]$ reduces to a sample moment. Thus, let us set

$$\tau(q) := -\lim_{n\to\infty} \frac{1}{n} \log_2 S_n(q), \tag{88}$$

where $S_n(q) = 2^n \mathbb{E}_t\left[2^{qV_n}\right]$, i.e.

$$S_n(q) := \sum_{k_n=0}^{2^n-1} (\Delta_{k_n}^n[Y])^q\,. \tag{89}$$

Depending on the context, $\tau(q)$ is called the *partition function* or the *free energy* [15, 78, 79]. Again, we have added a factor 2^n for convenience.

A closer look at (48) reveals that it actually shows that $\tau(q) \le f_G{}^*$. As a matter of fact, it is proven in [32, 55] that

Lemma 5 *For every path of Y*

$$\tau(q) = f_{\mathrm{G}}{}^*(\alpha) := \inf_\alpha (q\alpha - f_{\mathrm{G}}(\alpha)). \tag{90}$$

As an immediate consequence, the function $\tau(q)$ is concave and thus continuous and almost everywhere differentiable.

It is instructive to see how the quick and dirty argument (48) can be strengthened to yield a lower bound on $\tau(q)$. Again, our reasoning can be turned into an actual proof [32]. This time, we will collect the k_n with $\alpha^n_{k_n}$ approximately equal to some given value, say $l\varepsilon$, for varying l. Assuming that the range of $\alpha(t)$ is bounded, we can set $m := \lfloor \sup(\alpha(t))/\varepsilon \rfloor$. Using (81) and observing that q may be positive or negative, we obtain

$$\begin{aligned} S_n(q) &= \sum_{l=0}^m \sum_{|\alpha^n_{k_n} - l\varepsilon| \le \varepsilon/2} (\Delta^n_{k_n}[Y])^q \\ &\le \sum_{l=0}^m \sum_{l\varepsilon \le \alpha^n_{k_n} \le (l+1)\varepsilon} 2^{-n(l\varepsilon q - |q\varepsilon/2|)} \\ &\le \sum_{l=0}^m 2^{n(f_{\mathrm{G}}(l\varepsilon)+\eta)} 2^{-n(l\varepsilon q - |q\varepsilon/2|)} \\ &= 2^{n(\eta + |q\varepsilon/2|)} \sum_{l=0}^m 2^{-n(q(l\varepsilon) - f_{\mathrm{G}}(l\varepsilon))} \\ &\le (m+1)\, 2^{n(\eta+|q\varepsilon/2|)}\, 2^{-n \inf_\alpha (q\alpha - f_{\mathrm{G}}(\alpha))}. \end{aligned} \tag{91}$$

This shows that $\tau(q) \ge f_{\mathrm{G}}{}^*(q) - \eta - |q\varepsilon/2|$, and since $\eta > 0$ and $\varepsilon > 0$ can be made arbitrarily small the argument is complete.[15]

The partition function $\tau(q)$ is clearly easier to estimate than f_{G}, and it depends in a more regular manner on the data since it involves averages. Consequently, we introduce the *Legendre multifractal spectrum*:

$$f_{\mathrm{L}}(\alpha) := \tau^*(\alpha) = \inf_{q \in \mathbb{R}} \left(q\alpha - \tau(q)\right). \tag{92}$$

Recall (87) for the computation of τ^*. Unfortunately, f_{L} may contain less information than f_{G} since the Legendre back-transform yields only

$$f_{\mathrm{G}}(\alpha) \le f_{\mathrm{G}}{}^{**}(\alpha) = f_{\mathrm{L}}(\alpha), \tag{93}$$

where $f_{\mathrm{G}}{}^{**}$ is the concave hull of f_{G}. Strictly speaking, we have to establish that the limit f_{G} actually exists before making such a statement. A simple application of the LDP Theorem of Gärtner-Ellis [59] makes this rigorous under somewhat more restrictive assumptions (see the following theorem which is proven in [32, 55]). Alternatively, we could replace the $\lim_{n\to\infty}$ in the definitions of τ and f_{G} by the mathematically more technical $\limsup_{n\to\infty}$ as it is done in [32, 55].

Theorem 6 *Assume that $\tau(q)$ exists and is differentiable for all real q. Then, the double limit $f_{\mathrm{G}}(\alpha)$ exists for all α, and, moreover*

$$f_{\mathrm{G}}(\alpha) = f_{\mathrm{L}}(\alpha). \tag{94}$$

[15] The argument is not rigorous, since η and ε are entangled, i.e. ε appears in $|\alpha^n_{k_n} - l\varepsilon| \le \varepsilon/2$ twice, once as the approximate location of α and once as the error made in this approximation.

For fBm we obtain the degenerate case of a concave partition function: $\tau(q) = qH - 1$ as we will see in an instant (98). It is consistent with f_{H} taking only one value $f_{\mathrm{H}}(H) = 1$. For the concatenation of fBm's as above we find $\tau(q) = \min_k(qH_k - 1)$, which is again consistent with $f_{\mathrm{H}}(H_k) = 1$ [70]. Truly concave behavior of $\tau(q)$, on the other hand, is found with real data traffic. As a consequence, there is an entire range of α values present, not just a few. In [80] we display estimations of $\tau(q)$ for fBm obtained by numerical simulations. Due to errors, the Legendre transforms cannot perfectly match the predicted spectrum consisting of only the points $(H, 1)$ and $(1+H, 0)$. The accuracy achieved is nevertheless convincing.

A.2.5 Deterministic envelopes of spectra

Often, we would like to use an analytical approach in order to gain intuition into or an estimate of what f_{G} can be expected to look like on a typical path of Y. To this end, we consider now t as well as Y to be random simultaneously as we apply the LDP. Fubini leads to the "deterministic partition function" (c.f. (47))

$$\begin{aligned} T(q) &:= -1 + \lim_{n\to\infty} \frac{1}{-n} \log_2 \mathbb{E}_{\omega,t}[2^{qV_n}] \qquad (95) \\ &= \lim_{n\to\infty} \frac{1}{-n} \log_2 \mathbb{E}_\omega[S_n(q)]. \qquad (96) \end{aligned}$$

It is not hard to show that

Lemma 7 ([32]) *For any random process we have, with probability one,*

$$\tau(q,\omega) \ge T(q) \qquad \textit{for all } q \textit{ with } T(q) < \infty. \tag{97}$$

This is actually enough to determine $\tau(q)$ for fBm. Indeed, since τ is a concave function with $\tau(0) = -1 = T(0)$, Lemma 7 implies that with probability one

$$\text{fBm:} \qquad \tau(q) = qH - 1 \qquad \text{for all } q > -1. \tag{98}$$

Proof: Let us consider first any q with finite $T(q)$. Given $\varepsilon > 0$ choose N such that $\mathbb{E}_\omega[S_n(q)] \le 2^{-n(T(q)-\varepsilon)}$ for all $n \ge N$. Then, since $\limsup a_n \le \sum_{n>N} a_n$ for positive a_n,

$$\begin{aligned} &\mathbb{E}_\omega\left[\limsup_{n\to\infty} 2^{n(T(q)-2\varepsilon)} S_n(q)\right] \\ &\quad\le \mathbb{E}_\omega\left[\sum_{n\ge N} 2^{n(T(q)-2\varepsilon)} S_n(q)\right] \\ &\quad= \sum_{n\ge N} 2^{n(T(q)-2\varepsilon)} \mathbb{E}_\omega[S_n(q)] \\ &\quad\le 1/(1-2^{-\varepsilon}), \end{aligned} \tag{99}$$

by the definition of T. This allows us to conclude that almost surely $\limsup_{n\to\infty} 2^{n(T(q)-2\varepsilon)} S_n(q,\omega) < \infty$. Hence, $\tau(q) \ge T(q) - 2\varepsilon$. This is trivial if $T(q) = -\infty$. It is clear that this estimate holds with probability one simultaneously for all $\varepsilon = 1/m$ ($m \in \mathbb{N}$) and some countable, dense set of q values with $T(q) < \infty$. The fact that $\tau(q)$ is always continuous completes the argument. ⋄

Corollary 8 *With probability one, for all α*

$$f_{\mathrm{H}}(\alpha) \le f_{\mathrm{G}}(\alpha) \le f_{\mathrm{L}}(\alpha) \le T^*(\alpha). \tag{100}$$

Equality holds for cascades at certain values of α as the following version of Theorem 11 states:

Theorem 9 *(***Multifractal formalism for the MWM***) Consider the MWM as given in (25) or (38), with multipliers $A_{j,k}$ identically distributed within scale and independent along lines of descendants (c.f. (40)). Assume furthermore that the $A_{(j)}$ (or equivalently the $M_i^{(j)}$, $i = 0, 1$) converge in distribution as $j \to \infty$. Then, with probability one we have that*

$$f_{\mathrm{H}}(\alpha) = f_{\mathrm{G}}(\alpha) = f_{\mathrm{L}}(\alpha) = T^*(\alpha) \tag{101}$$

for any countable set of α's with $T^*(\alpha) > 0$. Moreover, since f_{L} and T^* are continuous, they must be equal on the entire interval $\{\alpha : T^*(\alpha) > 0\}$.

Remark: It can be shown that all spectra remain unchanged if $\alpha_{k_n}^n$ is replaced by $H_{k_n}^n$ [32].

A.2.6 Multifractals and Besov spaces

Besov spaces are also useful for analyzing the regularity of functions, especially since an elegant description of these regularity spaces in terms of wavelet coefficients has become available. In [81] it is shown that the norm of the Besov space $B_v^s(L^u)$ of a process with wavelet coefficients $W_{j,k}$ is equivalent to

$$|U_{0,0}| \;+\; \left(\sum_j \left(\sum_k 2^{jsu} 2^{-j} \left|2^{j/2} W_{j,k}\right|^u\right)^{v/u}\right)^{1/v}. \tag{102}$$

Roughly speaking, this norm measures the smoothness of order s in L^u, where v is an additional parameter for making finer distinctions in smoothness.

Multifractal analysis (using wavelet coefficients) can be viewed as determining in which Besov spaces the analyzed process lies. Using a convenient wavelet, define $\widetilde{\tau}(q)$ as in $\tau(q)$ but with $\widetilde{S}_n(q)$ (see (55)) replacing $S_n(q)$. Then, we find easily that the $B_v^s(L^u)$ norm of a path of the process is finite if $su < \widetilde{\tau}(u) + 1$ and infinite if $su > \widetilde{\tau}(u) + 1$.

For (102) to hold, s must be smaller than the regularity r of the wavelet, i.e. we need r vanishing moments as well as r continuous derivatives. Given this, Besov norms do not depend on the choice of the wavelet basis. Since the multifractal analysis using wavelets determines the Besov spaces that contain the signal, we conclude that $\widetilde{\tau}(u)$ will not depend on the choice of the wavelet, provided the above regularity conditions are met.

For an MWM signal $C(t)$ with identically distributed multipliers, we can say more. It can be shown [32] that the wavelet coefficients W_{n,k_n} of μ for any $[0,1]$–supported mother wavelet are distributed as $\widetilde{A}_{n,k_n} 2^{n/2} M_{k_n}^n \cdots M_{k_1}^1$ with $\widetilde{A}_{n,k_n}$ independent of $M_{k_i}^i$ and distributed as $W_{0,0}$. So, it follows that (56) holds also in this setting with T given by (65) (see [51] for a similar result on deterministic cascades). Choosing a compactly supported wavelet with enough regularity we find, using Lemma 7, that an MWM signal $C(t)$ with identically distributed multipliers is in $B_v^s(L^u)$ for all $s < (T(u) - u + 1)/u$ almost surely.

A.3 Interpretation of multifractal spectra

We collect here as a summary a few basic properties of multifractal spectra that follow directly from the above definitions and theorems. Here, Y is an arbitrary increasing process.

- **$\alpha(t) > 1$:** Y differentiable at t with derivative 0. In the case of a cascade, the plot Figure 5 is a graph of the approximative derivative of D, i.e. $C^{(n)}[k]/2^{-n} = 2^{-n(\alpha_k^n - 1)} \sim 2^{-n(\alpha-1)} \to 0$, at resolution 2^{-n} near t.
- **$\alpha(t) < 1$:** These are points where Y is singular and has "instant growth": The plot Figure 5 will show height $C^{(n)}[k]/2^{-n} = 2^{-n(\alpha-1)} \to \infty$ at resolution 2^{-n} near t.
- **$f_{\mathrm{H}}(\overline{\alpha}) = 1$:** This means that at almost all points $\alpha(t) = \overline{\alpha}$. Recall that $\overline{\alpha} > 1$ for increasing processes such as the binomial distribution function D.
- **$f_{\mathrm{G}}(\overline{\alpha}) = 1$:** This says that for a significant number of $k = 0, \ldots, 2^n - 1$ we see increments of the size $\Delta_k^n[Y] = 2^{-n\alpha_k^n} \simeq 2^{-n\overline{\alpha}}$.
- **$f_{\mathrm{G}}(a) < f_{\mathrm{G}}(b)$:** The chance to encounter an interval $[k2^{-n}, (k+1)2^{-n}]$ with $\alpha_k^n \simeq a$ is significantly smaller than finding $\alpha_k^n \simeq b$. These chances are $2^{n(f_{\mathrm{G}}(a)-1)}$ and $2^{n(f_{\mathrm{G}}(b)-1)}$, respectively. Both are very small regardless, unless $b = \overline{\alpha}$.
- **$\cap$-shape of f_{G}:** If this is the case, then the multifractal formalism holds, i.e. $f_{\mathrm{G}} = \tau^*$. This is true for the MWM and binomial measures. It may fail, however, e.g. for superpositions of MWM's with different spectra [55].

B Proof of the Multifractal Formalism for MWM

Here, we outline the proof of the *multifractal formalism* (Theorem 9) for the MWM model. We will consider a slightly more general setting, i.e. we assume only that there are random variables $M_0^{(n)}$ and $M_1^{(n)}$ such that $M_0^{(n)} + M_1^{(n)} = 1$ almost surely, and that $M_{k_n}^n$ is identically distributed to $M_{k'_{n-1}}^{(n)}$ for all n and k_n. This corresponds to choosing $A_{(n)}$ identically distributed as $M_0^{(n)} - M_1^{(n)}$. With this, we leave the original setting where $A_{(n)}$ must be symmetric. We do so in order to first study the deterministic case and acquaint ourselves with the methods. In the deterministic case, the requirement of symmetry would force all $A_{(n)}$ to be zero.

A closer look at (64) yields immediately:

Lemma 10 *Consider an MWM as given in (25) or (38), with multipliers $A_{j,k}$ identically distributed within scale and* *independent along lines of descendants (c.f. (40) and (63)), but not necessarily symmetric. Then,*

$$T(q) = -\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n} \log_2 \mathbb{E}_\omega\Big[(M_0^{(i)})^q + (M_1^{(i)})^q\Big] \quad (103)$$

provided the limit exists. We set $T(q) = -\infty$ *if* $\mathbb{E}_\omega\Big[(M_0^{(i)})^q + (M_1^{(i)})^q\Big] = \infty$ *for large* i.

We aim to establish the following:

Theorem 11 *Let the assumptions of Lemma 10 be in force. In addition, assume that the multipliers* $M^{(n)}$ *converge in a very weak sense: we require the limit (103) to exist for all* q*. Then, for any* α *with* $T^*(\alpha) > 0$

$$\dim(K_\alpha) = f_G(\alpha) = \tau^*(\alpha) = T^*(\alpha) \quad (104)$$

almost surely. For any other α*, the set* K_α *is empty and* $f_G(\alpha) = -\infty$ *almost surely.*

With (100), we only need to show that $T^*(\alpha) \le \dim(K_\alpha)$. We will start by giving the basic argument for a deterministic binomial cascade and show first how to generalize this result to a cascade with multipliers whose distributions vary with scale, but converge as $j \to \infty$. Then, we will outline the method of Falconer [24] that generalizes the basic argument to the random case and explain how to adapt it to the case of variable multipliers. As will be apparent, we only need convergence of the multipliers in a mean sense, as in (103). However, our generalization applies to arbitrary "statistically self-similar" measures as introduced in [24], provided we have convergence in distribution.

B.1 Deterministic cascade

In this section, we will assume that the binomial measure μ (recall Section 5.1) was constructed via a deterministic cascade, i.e. there are two positive numbers m_0 and m_1 with $m_0 + m_1 = 1$ and $M_0^0 = 1$, $M_{k_n}^n = m_{k'_{n-1}}$ for all n almost surely.

Consider a more careful look into the Large Deviation result for this case. The LLN, as we have seen in (74), tells us that $\alpha(t) = \overline{\alpha} := -(1/2)\log_2(m_0 m_1)$ for Lebesgue-almost all t. In other words, $K_{\overline{\alpha}}$ is a set of positive length. Therefore,

$$\dim(K_{\overline{\alpha}}) = 1. \quad (105)$$

This implies with (100) that the peaks of the histograms (82) will be close to $\overline{\alpha}$. To obtain information about other $\dim(K_\alpha)$ and other parts of the histograms, we need to have a way of choosing intervals (or points t) where the "unusual" happens, i.e. where $\alpha_{k_n}^n$ is "far" from $\overline{\alpha}$ (c.f. (43) and (74)).

This we will achieve through a "change of probability", meaning that the points t are chosen randomly according to a law μ_q that insures the convergence of $\alpha_{k_n}^n$ towards some value a_q. This distribution μ_q is defined in the same way as μ but with probabilities $\overline{m}_0 := m_0^q 2^T$ and $\overline{m}_1 := m_1^q 2^T$. Note that $\overline{m}_0 + \overline{m}_1 = 1$ due to (103), i.e., $T(q) = -\log_2(m_0^q + m_1^q)$.

The key observation is that $\mu_q(I_{k_n}^n) = \overline{m}_{k'_{n-1}}\,\overline{m}_{k'_{n-2}} \cdots \overline{m}_{k'_0}$ is the μ_q-probability that a μ_q-random point t lies in the interval $I_{k_n}^n = [k_n 2^{-n}, (k_n+1)2^{-n})$. (Recall that $k_n 2^{-n} = \sum_{i=0}^{n-1} k'_i 2^{-1-i}$ from (22).) In other words, for any i the μ_q-probability to observe the dyadic digit $k'_i = j$ is $\overline{m}_j$. Applying now the LLN to μ_q yields

$$\begin{aligned} \alpha_{k_n}^n = -\frac{1}{n}\log_2 \mu(I_{k_n}^n) &= -\frac{1}{n}\log_2\left(m_{k'_{n-1}}\, m_{k'_{n-2}} \cdots m_{k'_0}\right) \\ \to \mathbb{E}_{\mu_q}[-\log_2(m_{k'_0})] &= -\sum_{i=0}^{1} \overline{m}_i \log_2 m_i = T'(q). \quad (106)\end{aligned}$$

In other words, for the points picked randomly with distribution μ_q, the $\alpha_{k_n}^n$ converge (almost surely) to $a_q := T'(q)$. Thus these points all lie in

$$K_{a_q} := \{t : \alpha(t) := \lim_n \alpha_{k_n}^n = a_q\}. \quad (107)$$

To determine the dimension of K_α let us note that for the same points t in K_α we have

$$\begin{aligned} -\frac{1}{n}\log_2 \mu_q(I_{k_n}^n) &= -\frac{1}{n}\log_2\left(\overline{m}_{k'_{n-1}} \cdot \overline{m}_{k'_{n-2}} \cdots \overline{m}_{k'_0}\right) \\ &\to q a_q - T(q) = T^*(a_q), \quad (108)\end{aligned}$$

using $\overline{m}_i := m_i^q 2^T$. This result is helpful in two ways. First, it gives an intuitive proof of the theorem, or at least one for $f_G(\alpha) = T^*(\alpha)$. Indeed, the following very rough estimation (which can be made precise along the lines of [29, p. 137]) yields the number of intervals that have $\alpha(I_{k_n}^n) \simeq a_q$. These intervals are the ones contributing the bulk probability to μ_q. Using (108),

$$\begin{aligned} 1 &\simeq \sum_{\alpha(I_k^{(n)})\simeq a_q} \mu_q\left(I_k^{(n)}\right) \\ &\simeq \#\left\{k : \alpha\left(I_k^{(n)}\right) \simeq a_q\right\} 2^{-nT^*(a_q)}. \quad (109)\end{aligned}$$

Thus, the number of such intervals is approximately $2^{nT^*(a_q)}$; in other words $f_G(\alpha) = T^*(\alpha)$.

Second, (108) allows us the estimate $\dim(K_\alpha) \ge T^*(\alpha)$ using [35, Prop. 4.9]. Intuitively, we can think of μ_q as generalizing d-dimensional volume, since it scales in the right way: if a subset E of K_α is shrunk by a factor r then its μ_q-measure multiplies by r^{T^*}. If T^* was an integer d this would be exactly the definition of d-dimensional volume. Now a planar object in space has infinite 1-d volume (length), zero 3-d volume, but finite, positive 2-d volume (area); its dimension is 2, after all. Generalizing, we say that K_α has at least dimension $T^*(\alpha)$ since $\mu_q(K_\alpha) = 1$ is positive, i.e. $\dim(K_\alpha) \ge T^*(\alpha)$. A complete argument is given in [23, 55].

B.2 Deterministic cascade with variable multipliers

Let us now generalize μ slightly by allowing the almost sure multipliers m_i to depend on scale: $M_i^{(n)} = m_i^{(n)}$ for all n almost surely, where $m_0^{(n)} + m_1^{(n)} = 1$. Let us assume, however, that the $m_i^{(n)}$ converge, say to m_i. Then, using (103)

$$\begin{aligned} T(q) &= \lim_{n\to\infty} \frac{-1}{n} \sum_{i=1}^{n} \log_2\Big((m_0^{(i)})^q + (m_1^{(i)})^q\Big) \\ &= -\log_2\left((m_0)^q + (m_1)^q\right), \end{aligned} \tag{110}$$

and we obtain the same formula as in the previous section.

Applying now the Strong LLN to the same auxiliary measures μ_q as before we find

$$\frac{\sum_{i=1}^n \log_2 m_{k'_{i-1}}^{(i)} - \mathbb{E}_{\mu_q} \sum_{i=1}^n \log_2 m_{k'_{i-1}}^{(i)}}{n} \to 0 \tag{111}$$

for μ_q almost all points. But

$$\begin{aligned} &\frac{\mathbb{E}_{\mu_q} \sum_{i=1}^n \log_2 m_{k'_{i-1}}^{(i)}}{n} \\ &= \frac{\sum_{i=1}^n \left(\overline{m}_0 \log_2 m_0^{(i)} + \overline{m}_1 \log_2 m_1^{(i)}\right)}{n} \\ &\to \overline{m}_0 \log_2 m_0 + \overline{m}_1 \log_2 m_1, \end{aligned} \tag{112}$$

whence we obtain $\alpha_{k_n}^n \to a_q$ μ_q-almost surely, exactly as before. In summary, we have again $T^*(\alpha) \le \dim(K_\alpha)$.

B.3 Random cascades

Let us turn finally to the case of random multipliers. For a start, we assume the same distribution on all scales, i.e. all $M_i^{(n)}$ ($n \in \mathbb{N}$) are distributed as some M_i, where $M_0 + M_1 = 1$ almost surely.

Such cascades have been termed "conservative" by Mandelbrot [53] due to the conservation of mass in every step. Subsequent mathematical studies on cascades considered the case of independent M_i with $\mathbb{E}[M_0 + M_1] = 1$ [22]. These results have been generalized to conservative cascades [54] and [52], and to more general invariant measures [24–26].

Here, we present the argument of Falconer [24]. Essentially, there are two difficulties to deal with. First, the auxiliary measures μ_q are now random, and we have to ensure their existence. Second, as the multipliers for each realization will have different values from scale to scale (though drawn randomly with equal distribution), not even the strong LLN helps here and we have to prove $\mu_q(K_\alpha) = 1$ directly.

To guarantee the convergence of the construction of μ_q, we use a martingale argument. Let

$$\overline{M}_{k_n}^n := (M_{k_n}^n)^q 2^{\gamma_n}. \tag{113}$$

Since the $M_{k_n}^n$ are distributed as $M_{k'_{n-1}}^{(n)}$ we have $\overline{M}_{k_n}^n \overset{d}{=} (M_{k'_{n-1}}^{(n)})^q 2^{\gamma_n}$, which we abbreviate by $\overline{M}_{k'_{n-1}}^{(n)}$. Thereby, $\gamma_n(q)$ is chosen such that $\mathbb{E}\left[\overline{M}_0^{(n)} + \overline{M}_1^{(n)}\right] = 1$. We define μ_q^n as

$$\mu_q^n(I_{k_n}^n) := \overline{M}_{k_n}^n\, \overline{M}_{k_{n-1}}^{n-1} \cdots \overline{M}_{k_1}^1. \tag{114}$$

Now keeping k_n fixed, we write $I_{k_n}^n$ as a union of smaller dyadic intervals $I_{k_{m+1}}^{m+1}$, where $m > n$ and where k_{m+1} runs over $2^{m+1-n}k_n, \ldots, 2^{m+1-n}(k_n+1) - 1$, we obtain

$$\begin{aligned} &\mathbb{E}_\omega\left[\mu_q^{m+1}(I_{k_n}^n) | \mu_q^m\right] \\ &= \sum_{i=n}^{m} \sum_{k'_i=0,1} \mathbb{E}_\omega\left[\mu_q^{m+1}(I_{k_{m+1}}^{m+1}) | \mu_q^m\right] \\ &= \sum_{i=n}^{m} \sum_{k'_i=0,1} \mathbb{E}_\omega\left[\overline{M}_{k_{m+1}}^{m+1}\, \overline{M}_{k_m}^m \cdots \overline{M}_{k_1}^1 | \mu_q^m\right] \\ &= \sum_{i=n}^{m} \sum_{k'_i=0,1} \mathbb{E}_\omega\left[\overline{M}_{k_{m+1}}^{m+1}\right] \overline{M}_{k_m}^m \cdots \overline{M}_{k_1}^1 \\ &= \sum_{i=n}^{m-1} \sum_{k'_i=0,1} \mathbb{E}_\omega\left[\overline{M}_0^{(m+1)} + \overline{M}_1^{(m+1)}\right] \overline{M}_{k_m}^m \cdots \overline{M}_{k_1}^1 \\ &= \sum_{i=n}^{m-1} \sum_{k'_i=0,1} \overline{M}_{k_m}^m \cdots \overline{M}_{k_1}^1 \\ &= \mu_q^m(I_{k_n}^n). \end{aligned} \tag{115}$$

This shows that $\mu_q^m(I_{k_n}^n)$, $m \in \mathbb{N}$, forms a martingale and, thus converges. The limit is denoted by $\mu_q(I_{k_n}^n)$ and defines a true measure as we let n and k_n vary.

In our situation, all $\gamma_n(q)$ are equal to $T(q)$ since the distributions of the multipliers do not depend on scale. However, as presented here, it becomes clear that the martingale construction holds also for variable multipliers. Furthermore, it is indeed easy to see that under the assumption of Theorem 11, the γ_n converge to T. This knowledge is enough to generalize the proof of [24] to our case.

Falconer's proof applies to general random measures that are statistically self-similar [24], i.e. where the multipliers of "mass" as well as "geometry" are random. It is notable that the generalization indicated above works also in this case, i.e. when the distributions are allowed to depend on scale. However, a slightly stronger assumption has to be imposed: we require that these multipliers converge in distribution. In the case of a binomial cascade, the geometry is deterministic by definition. This is why the weaker condition (103) is enough here.

Finally, for simplicity we have not bothered with the fact that [24] assumes that the multipliers are bounded away from zero. In order to make the proof complete for arbitrary MWM processes, where the multipliers may be arbitrarily small, the more involved approach of [56] needs to be taken. This is, however, certainly beyond the scope of this paper.

Acknowledgments

The authors give special thanks to Walter Willinger (AT&T) and Edward Knightly (Rice) for enlightening discussions on issues related to Internet traffic. Rudolf Riedi would like to extend his gratitude to Fabrice Clerot and CNET 95 8B 069, France for financial support that enabled him to make first contact with the exciting field of networking. Finally, thanks to Ramesh Neelamani and Justin Romberg (Rice) for their detailed reading of the final manuscript.

References

[1] J. Feder, *Fractals*. New York, Plenum Press, 1989.

[2] D. Cox, "Long-range dependence: A review," *Statistics: An Appraisal*, pp. 55–74, 1984.

[3] A. E. Jacquin, "Fractal image coding: A review," *Proc. of IEEE*, vol. 81, pp. 1451–1465, Oct. 1993.

[4] T. Lundahl, W. Ohley, S. Kay, and R. Siffert, "Fractional Brownian motion: A maximum likelihood estimator and its application to image texture," *IEEE Trans. on Medical Imaging*, vol. 5, pp. 152–161, Sep. 1986.

[5] W. Leland, M. Taqqu, W. Willinger, and D. Wilson, "On the self-similar nature of Ethernet traffic (extended version)," *IEEE/ACM Trans. on Networking*, vol. 2, pp. 1–15, 1994.

[6] I. Norros, "A storage model with self-similar input," *Queueing Systems*, vol. 16, pp. 387–396, 1994.

[7] G. Samorodnitsky and M. Taqqu, *Stable non-Gaussian random processes*. Chapman and Hall, New York ISBN 0-412-05171-0, 1994.

[8] M. Taqqu, V. Teverovsky, and W. Willinger, "Estimators for long-range dependence: An empirical study," *Fractals.*, vol. 3, pp. 785–798, 1995.

[9] I. Daubechies, *Ten Lectures on Wavelets*. New York: SIAM, 1992.

[10] M. Vetterli and J. Kovačević, *Wavelets and Subband Coding*. Englewood Cliffs, NJ: Prentice-Hall, 1995.

[11] G. W. Wornell, "A Karhunen-Loève like expansion for $1/f$ processes via wavelets," *IEEE Trans. on Info. Theory*, vol. 36, pp. 859–861, Mar. 1990.

[12] P. Flandrin, "Wavelet analysis and synthesis of fractional Brownian motion," *IEEE Trans. on Info. Theory*, vol. 38, pp. 910–916, Mar. 1992.

[13] L. M. Kaplan and C.-C. J. Kuo, "Fractal estimation from noisy data via discrete fractional Gaussian noise (DFGN) and the Haar basis," *IEEE Trans. on Signal Proc.*, vol. 41, pp. 3554–3562, Dec. 1993.

[14] P. Abry and D. Veitch, "Wavelet analysis of long range dependent traffic," *IEEE Trans. on Info. Theory*, vol. 4, no. 1, pp. 2–15, 1998.

[15] B. B. Mandelbrot, "Intermittent turbulence in self similar cascades: Divergence of high moments and dimension of the carrier," *J. Fluid. Mech.*, vol. 62, p. 331, 1974.

[16] U. Frisch and G. Parisi, "Fully developed turbulence and intermittency," *Proc. Int. Summer School on Turbulence and Predictability in Geophysical Fluid Dynamics and Climate Dynamics*, pp. 84–88, 1985.

[17] B. B. Mandelbrot and C. J. G. Evertsz, "Multifractality of the harmonic measure on fractal aggregates and extended self-similarity," *Physica*, vol. A 177, pp. 386–393, 1991.

[18] B. B. Mandelbrot, *Fractals and Scaling in Finance*. Springer New York, 1997.

[19] R. Riedi and J. L. Véhel, "Multifractal properties of TCP traffic: a numerical study," *Technical Report No 3129, INRIA Rocquencourt, France*, Feb, 1997. Available at `www.dsp.rice.edu/~riedi`.

[20] A. Feldmann, A. C. Gilbert, and W. Willinger, "Data networks as cascades: Investigating the multifractal nature of Internet WAN traffic," *Proc. ACM/Sigcomm 98*, vol. 28, pp. 42–55, 1998.

[21] A. C. Gilbert, W. Willinger, and A. Feldmann, "Scaling analysis of random cascades, with applications to network traffic," *IEEE Trans. on Info. Theory, this issue*, 1999.

[22] J.-P. Kahane and J. Peyrière, "Sur certaines martingales de Benoit Mandelbrot," *Adv. Math.*, vol. 22, pp. 131–145, 1976.

[23] R. Cawley and R. D. Mauldin, "Multifractal decompositions of Moran fractals," *Advances Math.*, vol. 92, pp. 196–236, 1992.

[24] K. J. Falconer, "The multifractal spectrum of statistically self-similar measures," *J. Theor. Prob.*, vol. 7, pp. 681–702, 1994.

[25] M. Arbeiter and N. Patzschke, "Self-similar random multifractals," *Math. Nachr.*, vol. 181, pp. 5–42, 1996.

[26] L. Olsen, "Random geometrically graph directed self-similar multifractals," *Pitman Research Notes Math. Ser.*, vol. 307, 1994.

[27] G. Brown, G. Michon, and J. Peyriere, "On the multifractal analysis of measures," *J. Stat. Phys.*, vol. 66, pp. 775–790, 1992.

[28] R. H. Riedi and B. B. Mandelbrot, "Multifractal formalism for infinite multinomial measures," *Adv. Appl. Math.*, vol. 16, pp. 132–150, 1995.

[29] R. H. Riedi and B. B. Mandelbrot, "Exceptions to the multifractal formalism for discontinuous measures," *Math. Proc. Cambr. Phil. Soc.*, vol. 123, pp. 133–157, 1998.

[30] K.-S. Lau and S.-M. Ngai, "Multifractal measures and a weak separation condition," *Adv. Math.*, p. to appear, 1999.

[31] Y. Pesin and H. Weiss, "A multifractal analysis of equilibrium measures for conformal expanding maps and Moran-like geometric constructions," *J. Stat. Phys.*, vol. 86, pp. 233–275, 1997.

[32] R. H. Riedi, "Multifractal processes," *Stoch. Proc. Appl.*, preprint, to be submitted 1999.

[33] R. H. Riedi and W. Willinger, "Toward an improved understanding of network traffic dynamics," in *Self-similar Network Traffic and Performance Evaluation*, Wiley, Int. Edition, June 1999.

[34] B. B. Mandelbrot, *The Fractal Geometry of Nature*. New York: Freeman, 1983.

[35] K. J. Falconer, *Fractal Geometry: Mathematical Foundations and Applications*. John Wiley and Sons, New York, 1990.

[36] B. B. Mandelbrot and J. W. V. Ness, "Fractional Brownian motion, fractional noises and applications," *SIAM Reviews*, vol. 10, pp. 422–437, 1968.

[37] P. Abry, P. Gonçalvés, and P. Flandrin, "Wavelets, spectrum analysis and $1/f$ processes," in *Lecture Notes in Statistics: Wavelets and Statistics* (A. Antoniadis and G. Oppenheim, eds.), vol. 103, pp. 15–29, 1995.

[38] A. Feldmann, A. Gilbert, W. Willnger, and T. Kurtz, "Looking behind and beyond self-similarity: Scaling phenomena in measured WAN traffic," in *Proc. of 35th Annual Allerton Conf. on Comm., Control, and Computing*, pp. 269–280, June 1997.

[39] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*. New York: McGraw-Hill, 1991.

[40] M. S. Crouse, R. D. Nowak, and R. G. Baraniuk, "Wavelet-based statistical signal processing using hidden Markov models," *IEEE Trans. on Signal Proc.*, vol. 46, pp. 886–902, April 1998.

[41] M. T. Orchard and K. Ramchandran, "An investigation of wavelet-based image coding using an entropy-constrained quantization framework," in *Data Compression Conference '94*, (Snowbird, Utah), pp. 341–350, 1994.

[42] M. Pasquier, "Positive function synthesis using wavelets," *ELEC 696 Semester Project Report, Dept. of Electrical and Computer Engineering, Rice University*, April 1997.

[43] M. Crouse and J. Lewis, "Wavelets and hidden Markov models for network traffic modeling," *ELEC 537 Course Project Report, Dept. of Electrical and Computer Engineering, Rice University*, Dec. 1996.

[44] K. E. Timmerman and R. D. Nowak, "Multiscale Bayesian estimation of Poisson intensities," in *Proc. 31st Asilomar Conf.*, (Pacific Grove, CA), Nov. 1997.

[45] K. E. Timmerman and R. D. Nowak, "Multiscale modeling and estimation of Poisson processes with application to medical imaging," *IEEE Trans. on Info. this issue*, 1999.

[46] M. S. Crouse and R. G. Baraniuk, "Simplified wavelet-domain hidden Markov models using contexts," in *Proc. 31st Asilomar Conf.*, (Pacific Grove, CA), Nov. 1997.

[47] L. M. Kaplan and C.-C. J. Kuo, "Extending self-similarity for fractional Brownian motion," *IEEE Trans. on Signal Proc.*, vol. 42, pp. 3526–3530, Dec. 1994.

[48] N. Johnson, S. Kotz, and N. Balakrishnan, *Continuous Univariate Distributions*, vol. 1-2. New York: John Wiley & Sons, 1994.

[49] R. Benzi, G. Paladin, G. Parisi, and A. Vulpiani, "On the multifractal nature of fully developed turbulence and chaotic systems," *J. Phys. A: Math. Gen.*, vol. 17, pp. 3521–3531, 1984.

[50] V. R. Chechetkin, "Multifractal structure of fully developed hydrodynamic turbulence II," *Stat. Phys.*, vol. 61, p. 589, 1990.

[51] E. Bacry, J. Muzy, and A. Arneodo, "Singularity spectrum of fractal signals from wavelet analysis: Exact results," *J. Stat. Phys.*, vol. 70, pp. 635–674, 1993.

[52] F. Ben Nasr, "Mandelbrot random measures associated with substitution," *C. R. Acad. Sc. Paris*, vol. 304, no. 10, pp. 255–258, 1987.

[53] B. B. Mandelbrot, "Limit lognormal multifractal measures," *Physica A*, vol. 163, pp. 306–315, 1990.

[54] R. Holley and E. Waymire, "Multifractal dimensions and scaling exponents for strongly bounded random cascades," *Ann. Appl. Prob.*, vol. 2, pp. 819–845, 1992.

[55] R. H. Riedi, "An improved multifractal formalism and self-similar measures," *J. Math. Anal. Appl.*, vol. 189, pp. 462–490, 1995.

[56] J. Barral, "Continuity, moments of negative order, and multifractal analysis of Mandelbrot's multiplicative cascades," *Université Paris Sud, Ph.D thesis no. 4704*, 1997.

[57] R. H. Riedi and I. Scheuring, "Conditional and relative multifractal spectra," *Fractals. An Interdisciplinary Journal*, vol. 5, no. 1, pp. 153–168, 1997.

[58] J.-D. Deuschel and D. W. Stroock, *Large Deviations.* Academic Press, 1984.

[59] R. Ellis, "Large deviations for a general class of random vectors," *Ann. Prob.*, vol. 12, pp. 1–12, 1984.

[60] S. Jaffard, "Local behavior of Riemann's function," *Contemporary Mathematics*, vol. 189, pp. 287–307, 1995.

[61] S. Jaffard, "Pointwise smoothness, two-microlocalization and wavelet coefficients," *Publicacions Matematiques*, vol. 35, pp. 155–168, 1991.

[62] R. Adler, *The Geometry of Random Fields.* John Wiley & Sons, New York, 1981.

[63] N. Duffield and N. O'Connell, "Large deviations and overflow probabilities for the general single-server queue, with applications," *Math. Proc. Cambr. Phil. Soc.*, vol. 118, pp. 363–374, 1995.

[64] N. Likhanov, B. Tsybakov, and N. Georganas, "Analysis of an ATM buffer with self-similar input traffic," *Proc. IEEE, Info com '95 (Boston 1995)*, pp. 985–992, 1995.

[65] A. Erramilli, O. Narayan, and W. Willinger, "Experimental queueing analysis with long-range dependent traffic," *IEEE/ACM Trans. on Networking*, pp. 209–223, April 1996.

[66] I. Norros, "Four approaches to the fractional Brownian storage," *Fractals in Engineering*, pp. 154–169, 1997.

[67] N. Duffield, "Economies of scale for long-range dependent traffic in short buffers," *Telecommunication Systems,* to appear, 1998.

[68] D. P. Heyman and T. V. Lakshman, "What are the implications of long-range dependence for VBR-video traffic engineering?," *IEEE/ACM Transactions on Networking*, vol. 4, pp. 301–317, June 1996.

[69] M. Taqqu and V. Teverosky, "Is network traffic self-similar or multifractal?," *Fractals*, vol. 5, no. 1, pp. 63–73, 1997.

[70] J. L. Véhel and R. Riedi, "Fractional Brownian motion and data traffic modeling: The other end of the spectrum," *Fractals in Engineering*, pp. 185–202, Springer 1997.

[71] E. Knightly and J. Qiu, "Measurement-based admission control with aggregate traffic envelopes," *Proceedings of IEEE ITWDC '98*, Sep. 1998.

[72] R. Cruz, "A calculus for network delay, part I: Network elements in isolation," *IEEE Trans. on Info. Theory*, vol. 37, pp. 114–121, January 1991.

[73] O. Narayan, "Exact asymptotic queue length distribution for fractional Brownian traffic," *Advances in Performance Analysis*, vol. 1, no. 1, p. 39, 1998.

[74] I. Norros, "On the use of fractional Brownian motion in the theory of connectionless networks," *COST*, vol. 242, 1994.

[75] M. Taqqu and J. Levy, *Using renewal processes to generate LRD and high variability.* In: Progress in probability and statistics, E. Eberlein and M. Taqqu eds., vol. 11. Boston: Birkhaeuser, 1986. pp 73–89.

[76] S. Jaffard, "Multifractal formalism for functions," *CRAS*, vol. 317, pp. 745–750, 1993.

[77] P. Billingsley, *Probability and Measure.* New York: Wiley, 1979.

[78] T. Halsey, M. Jensen, L. Kadanoff, I. Procaccia, and B. Shraiman, "Fractal measures and their singularities: The characterization of strange sets," *Phys. Rev. A*, vol. 33, pp. 1141–1151, 1986.

[79] P. Grassberger and I. Procaccia, "Measuring the strangeness of strange attractors," *Physica D*, vol. 9, pp. 189–208, 1983.

[80] P. Gonçalvès, R. Riedi, and R. Baraniuk, "Simple statistical analysis of wavelet-based multifractal spectrum estimation," in *Proc. 32nd Conf. on Signals, Systems and Computers*, (Asilomar, Pacific Grove, CA), Nov. 1998.

[81] Y. Meyer, "Principe d'incertitude, bases Hilbertiennes et algèbres d'opérateurs," *Sémanaire Bourbaki*, vol. 662, 1985–1986.