Trees, Wavelets and Large Deviations

Rolf Riedi

RICE UNIVERSITY

Dept of Statistics

Stochastic Systems and Modelling in Networking and Finance
Part III
Dependable Adaptive Systems and Mathematical Modeling
Kaiserslautern, August 2006

ON-OFF limits & the small scales

- ON-OFF explains two asymptotic regimes with self-similar limits
  - Beta regime:
  - highly multiplexed slow connections → fBm
  - Alpha regime:
  - Few fast large connections → Levy stable
- However, limits are at large scales, not small.

Highly multiplexed limit is Gaussian

$$\frac{1}{m^{1/2}}\sum_{i=1}^{m}(X_i(t)-\mathsf{E}X_i(t)) \stackrel{m\to\infty}{\longrightarrow} G(t)$$

At large scales self-similar, ie: fBm

$$\frac{1}{T^H}\int_0^{Tt} G(u)du \stackrel{fdd}{\to} \sigma B_H(t)$$

$$H=\frac{3-\min(\alpha_{\text{on}},\alpha_{\text{off}})}{2}$$

Large *scale* limit K is self-similar has indep. increments and heavy tails

$$\frac{1}{T^H}\int_0^{Tt}(X_i(t)-\mathsf{E}X_i(t)) \stackrel{T\to\infty}{\longrightarrow} K_i(t)$$

Highly multiplexed becomes Levy stable motion

$$\frac{1}{m^H}\sum_i K_i(t) \stackrel{m\to\infty}{\sim} L_H(t)$$

$$H=\frac{1}{\min(\alpha_{\text{on}},\alpha_{\text{off}})}$$

Tree based models

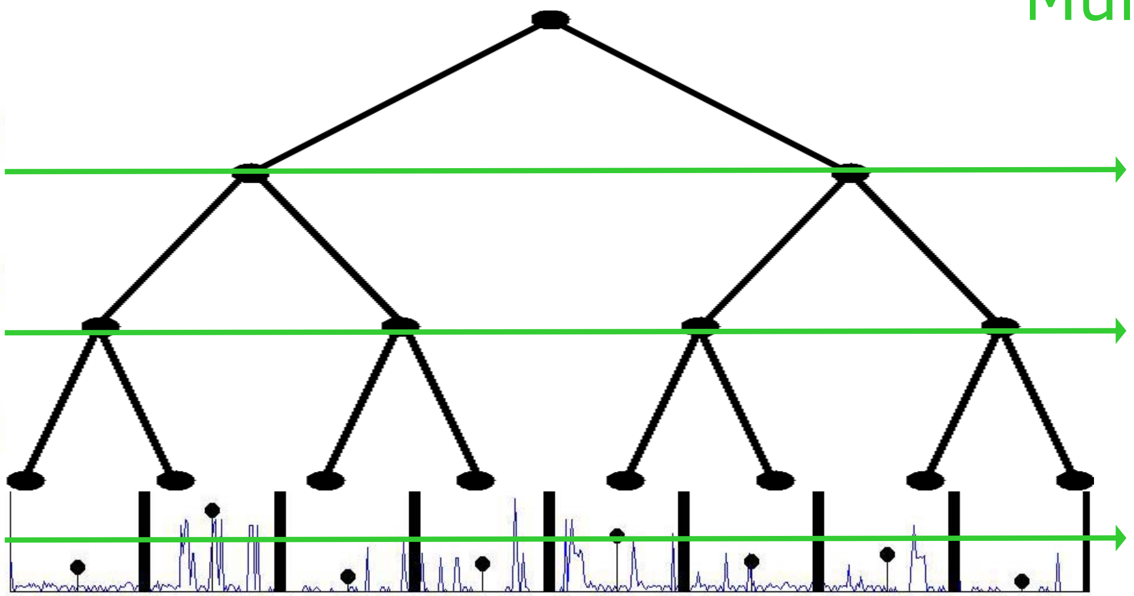
A hierarchical approach

Dyadic Multiscale Analysis

Time $\longrightarrow$

Multiscale statistics

Number of scales $j=\log_2(m)$

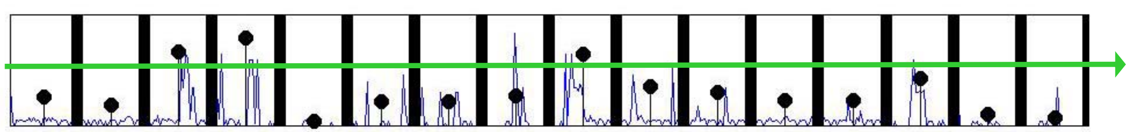


$V_1 = \mathrm{Var}\, X^{(m)}$

$V_2 = \mathrm{Var}\, X^{(m/2)}$

$V_3 = \mathrm{Var}\, X^{(m/4)}$

Flow up : $X_n^{(2m)} = X_{2n}^{(m)} + X_{2n+1}^{(m)}$



$V_j = \mathrm{Var}\, X^{(1)}$

Start at bottom with trace $X_n^{(1)}$

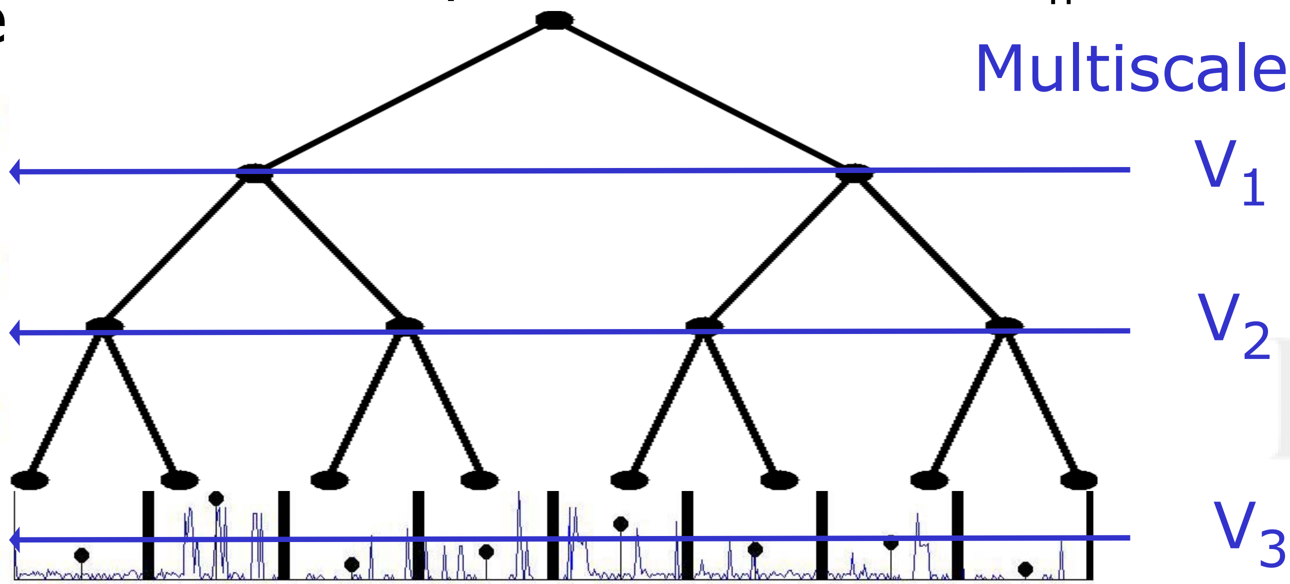
Dyadic Multiscale Synthesis

Time →

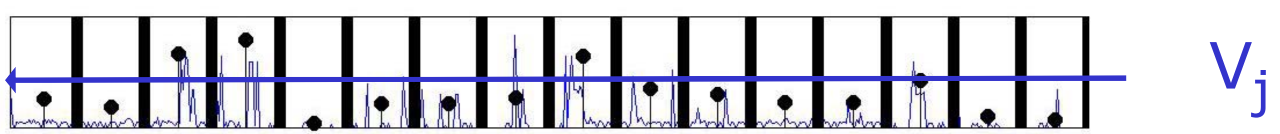
Scale ↕

Start at top with sum of all $X_n^{(1)}$

Multiscale parameters



Flow down : ¿ $X_n^{(2m)} \rightarrow X_n^{(m)}$?



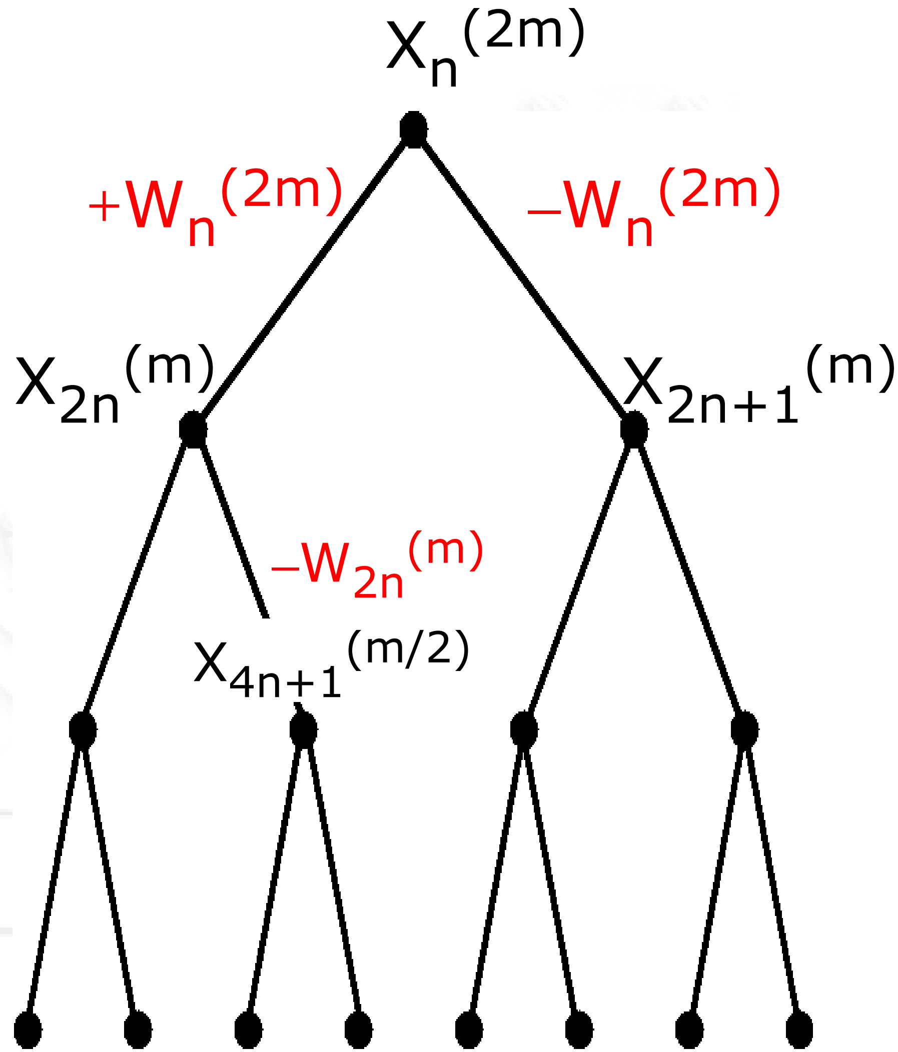
Signal: bottom nodes

Additive Innovations

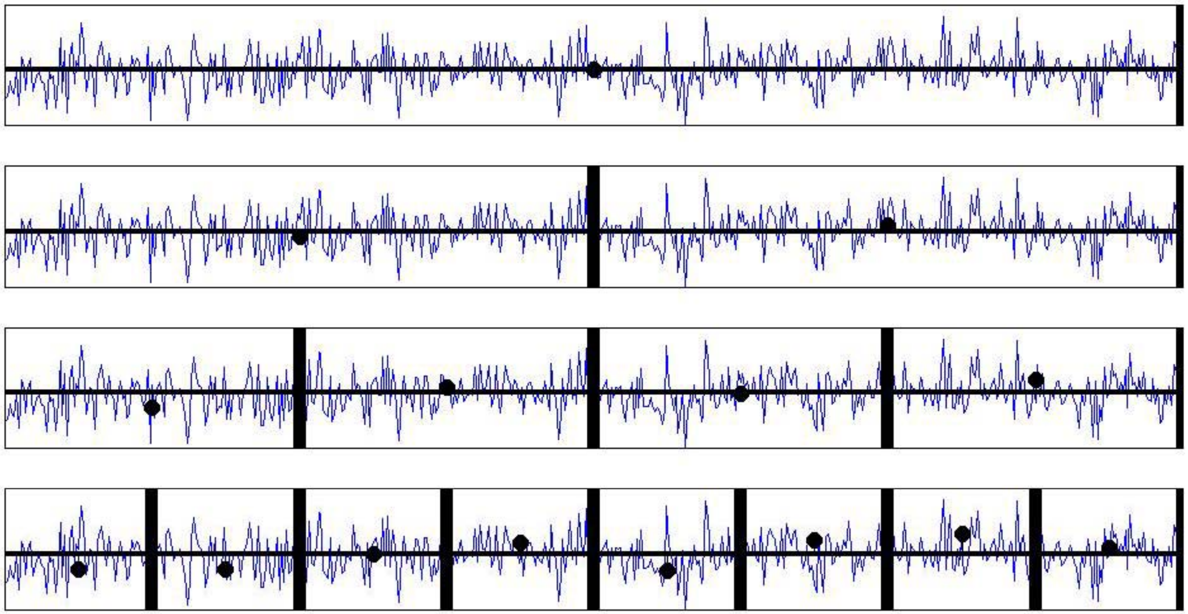
Synthesis:

- Start at root
- Flow down the tree
- Additive, independent innovations $W_n^{(m)}$
- Conservation:

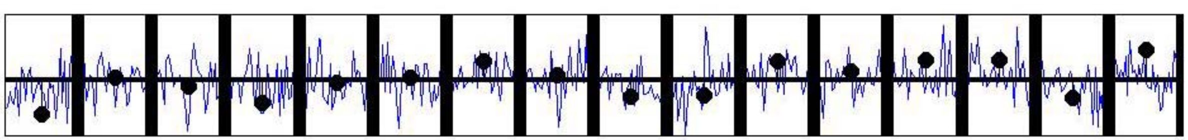
$$X_{2n}^{(m)} = (X_n^{(2m)} + W_n^{(2m)})/2$$
$$X_{2n+1}^{(m)} = (X_n^{(2m)} - W_n^{(2m)})/2$$



Additive Tree: Linear Processes



CLT: asymptotically Gaussian



Additive Innovations $W_n^{(m)} \sim \mathcal{N}(0, \sigma^2 m^{-(2H+1)})$: Model for $B_H(t)$

Multiplicative Innovations

Positive process:

- Add `small' innovation:

  $| W_n^{(m)} | < X_n^{(m)}$

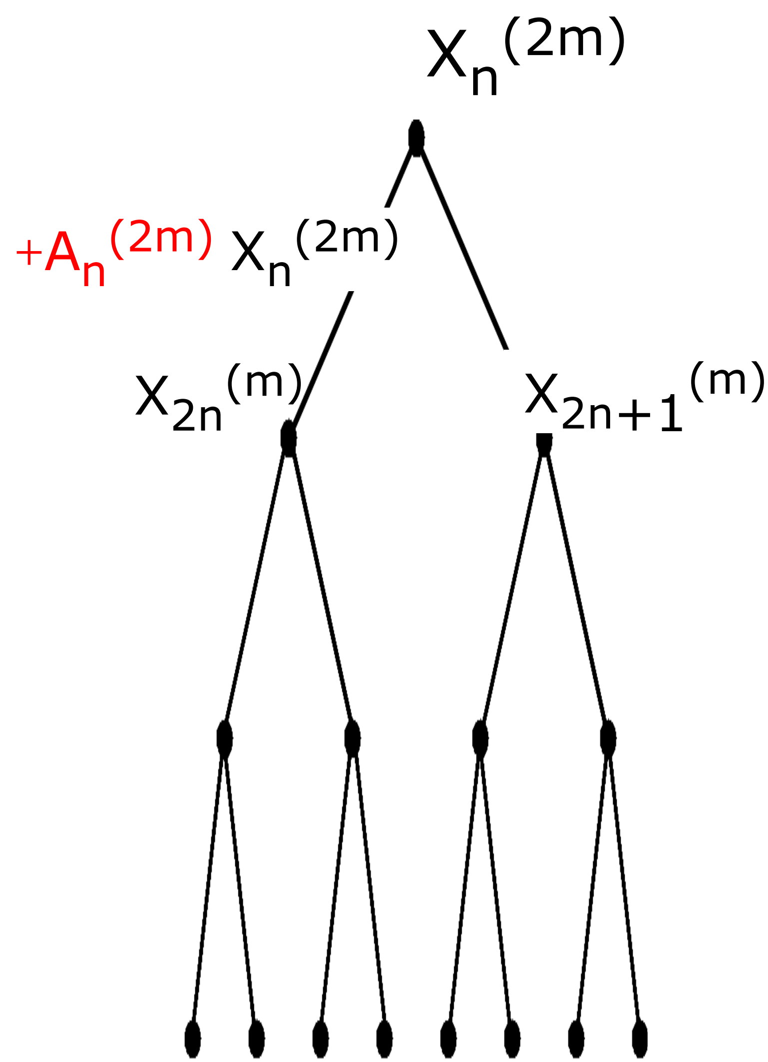
- Introduces dependence X,W
- Model:

  $W_n^{(m)} = A_n^{(m)} \cdot X_n^{(m)}$

  with independent $|A_n^{(m)}|<1$

- Conservation:

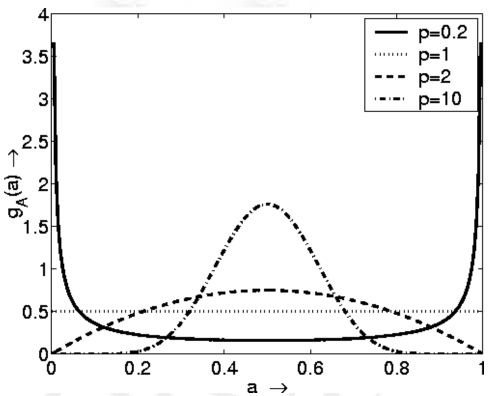
$$X_{2n}^{(m)} = X_n^{(2m)} \cdot (1 + A_n^{(2m)})/2$$
$$X_{2n+1}^{(m)} = X_n^{(2m)} \cdot (1 - A_n^{(2m)})/2$$



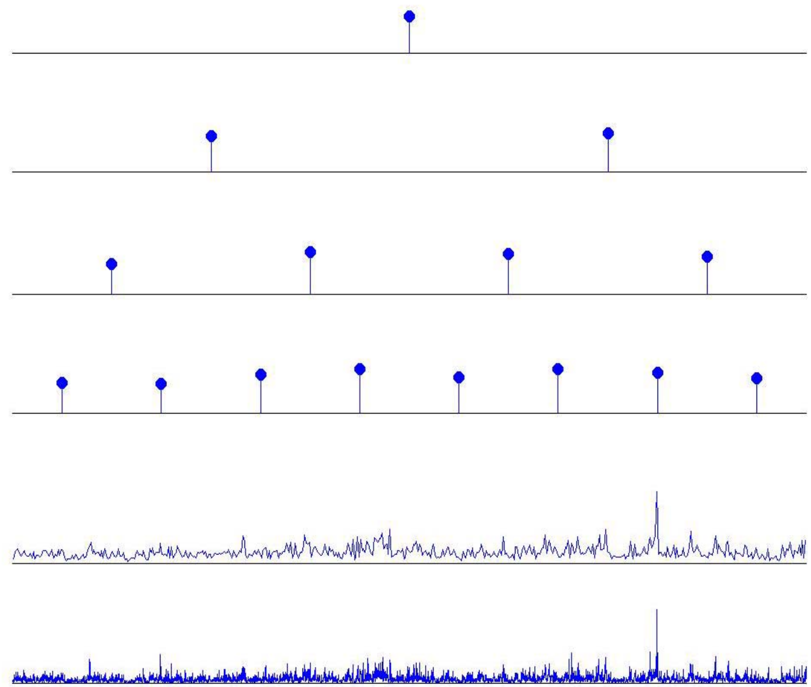
Multiplicative Innovation

Multiplicative Cascade-Model

Multiplicative Innovations



$(1 \pm A_n^{(m)})/2 \sim \mathcal{Beta}\,(\sigma_m)$



positive, spiky (log-normal). LRD

Control variance σ_m to

- Match variance of trace (model fitting)...or...
- match variance progression of LRD with H:

$$\mathrm{Var}X^{(m)} = \mathrm{Var}X^{(2m)} \cdot \frac{1}{4} \cdot \mathrm{Var}(A^{(2m)}) \sim m^{2H} \quad \Rightarrow \mathrm{Var}(A^{(2m)}) = 2^{-2H+2}$$

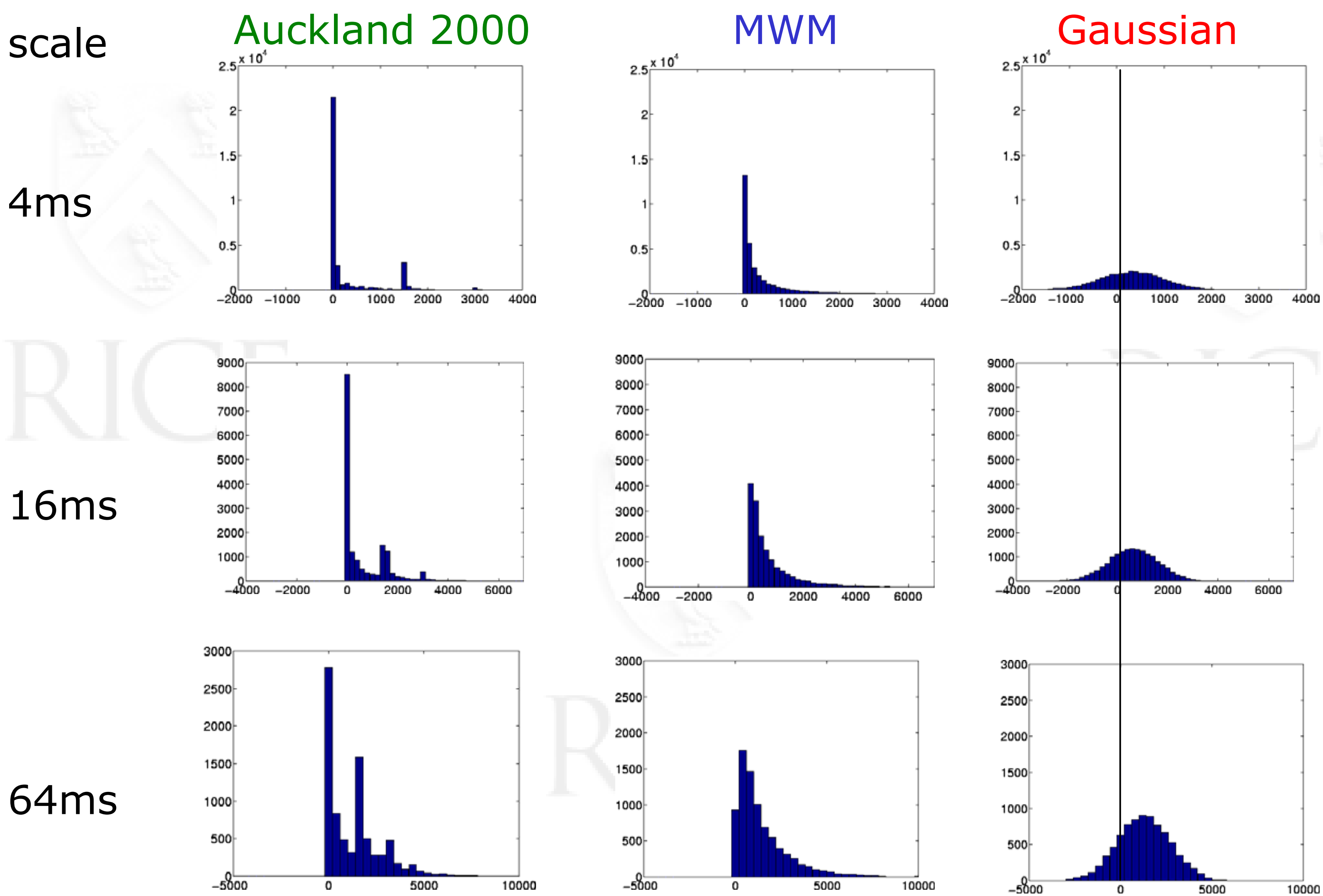
Network relevance

Simulation

Performance (Queuing)

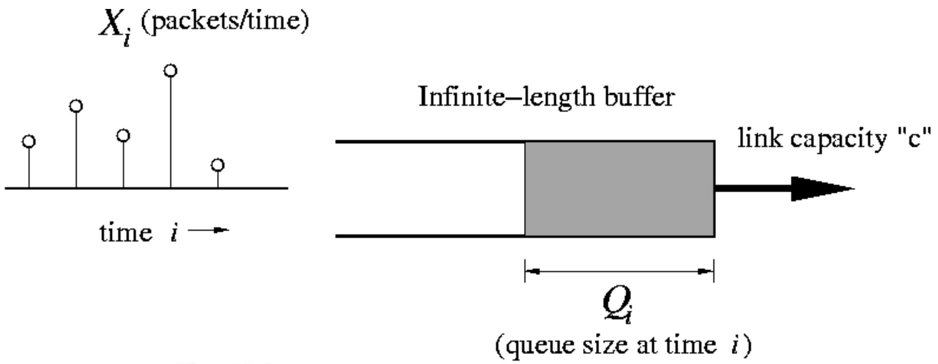
Inference (bandwidth estimation) → later

Multiscale Marginals



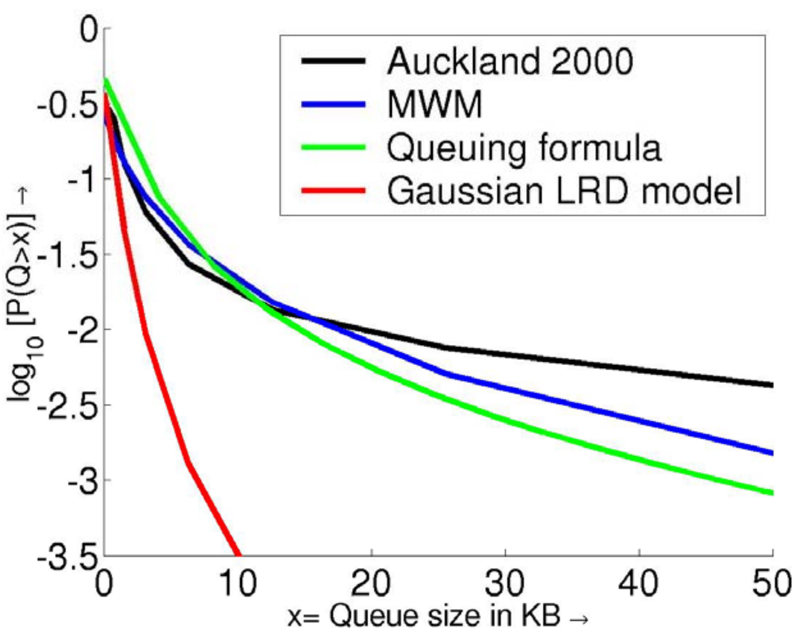
Equal variance on all scales

Queuing analysis



Q-tail: P[Q>b]

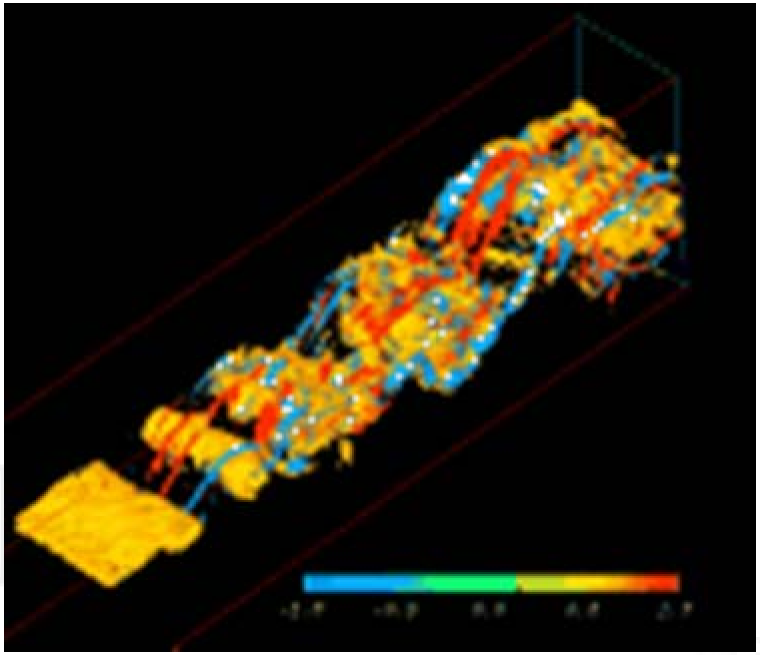
- Tree structure allows for analytical queuing formula
- Multiplicative model superior to additive
- Importance of multiscale marginal distributions

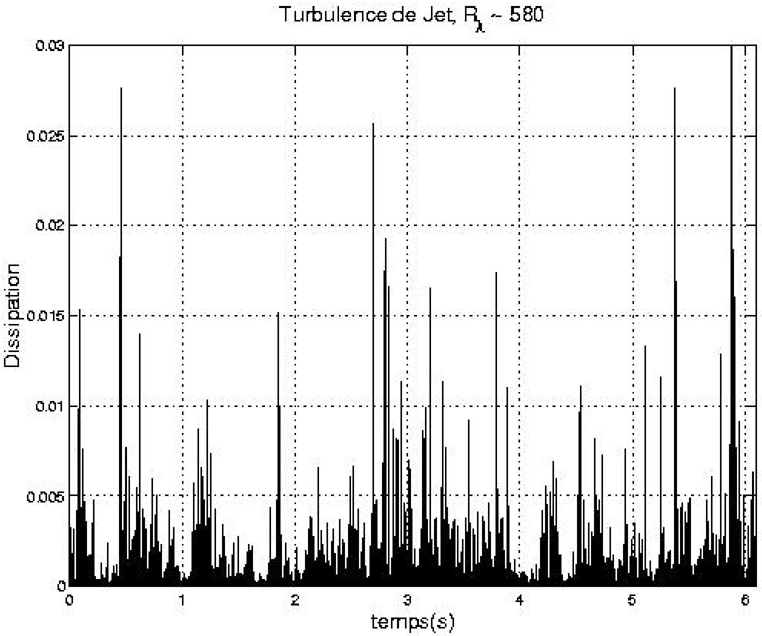


Binomial Cascade

Multifractal Toy

Why Cascades





Courtesy P. Chainais

- Turbulence:

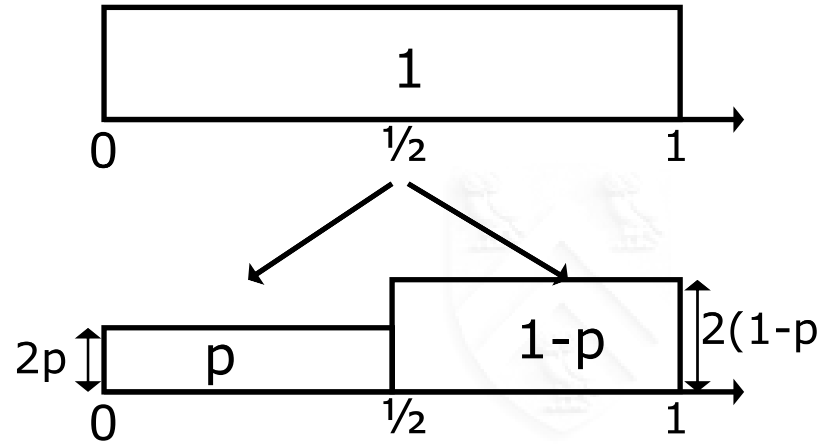
Kolmogorov 41: $\mathbb{E}[|v(t+\delta) - v(t)|^q] \simeq \delta^{q/3} \Rightarrow$ fBm $H = 1/3$

Kolmogorov 62: $\mathbb{E}[|v(t+\delta) - v(t)|^q] \simeq \delta^{\tau(q)}$?????

- Datatraffic: Cascades provide better match

The Toy: Binomial Cascade

- Start with unit mass
- Redistribute uniformly
  portion p <½ to the left
  portion 1-p to the right



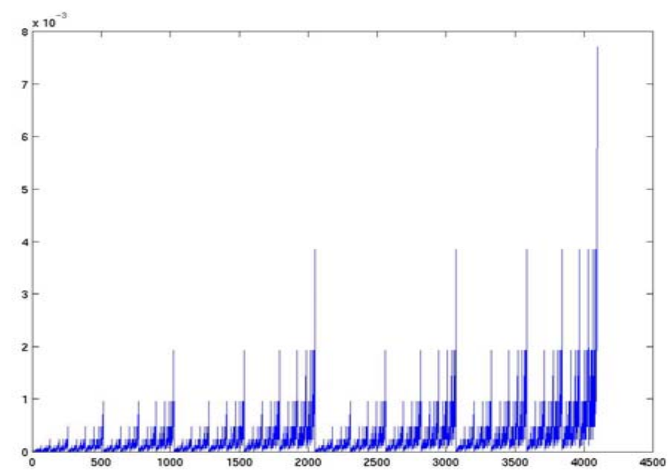
Mass re-distribution

- Iterate

$$t = \sum_{k=1}^{\infty} \epsilon_k / 2^k \quad \text{with } \epsilon_k = 0, 1$$

$$I(\epsilon_1 \ldots \epsilon_n) := [t_n, +1/2^n)) \quad t_n := \sum_{k=1}^{n} \epsilon_k / 2^k$$

$$l_n(t) := \#\{k \leq n : \epsilon_k = 1\} = \sum_{k=1}^{n} \epsilon_k$$



Increment: $X_{2^n t_n}^{(-n)} = p^{n - l_n(t)} (1-p)^{l_n(t)}$

Multifractal Spectrum

- Oscillate ~ $|t|^{\alpha}$ → local strength α

$$\alpha(t) := \liminf_{n} \alpha_n(t)$$

$$\alpha_n(t) = \frac{\log \Delta I_n(t)}{\log |I_n(t)|}$$

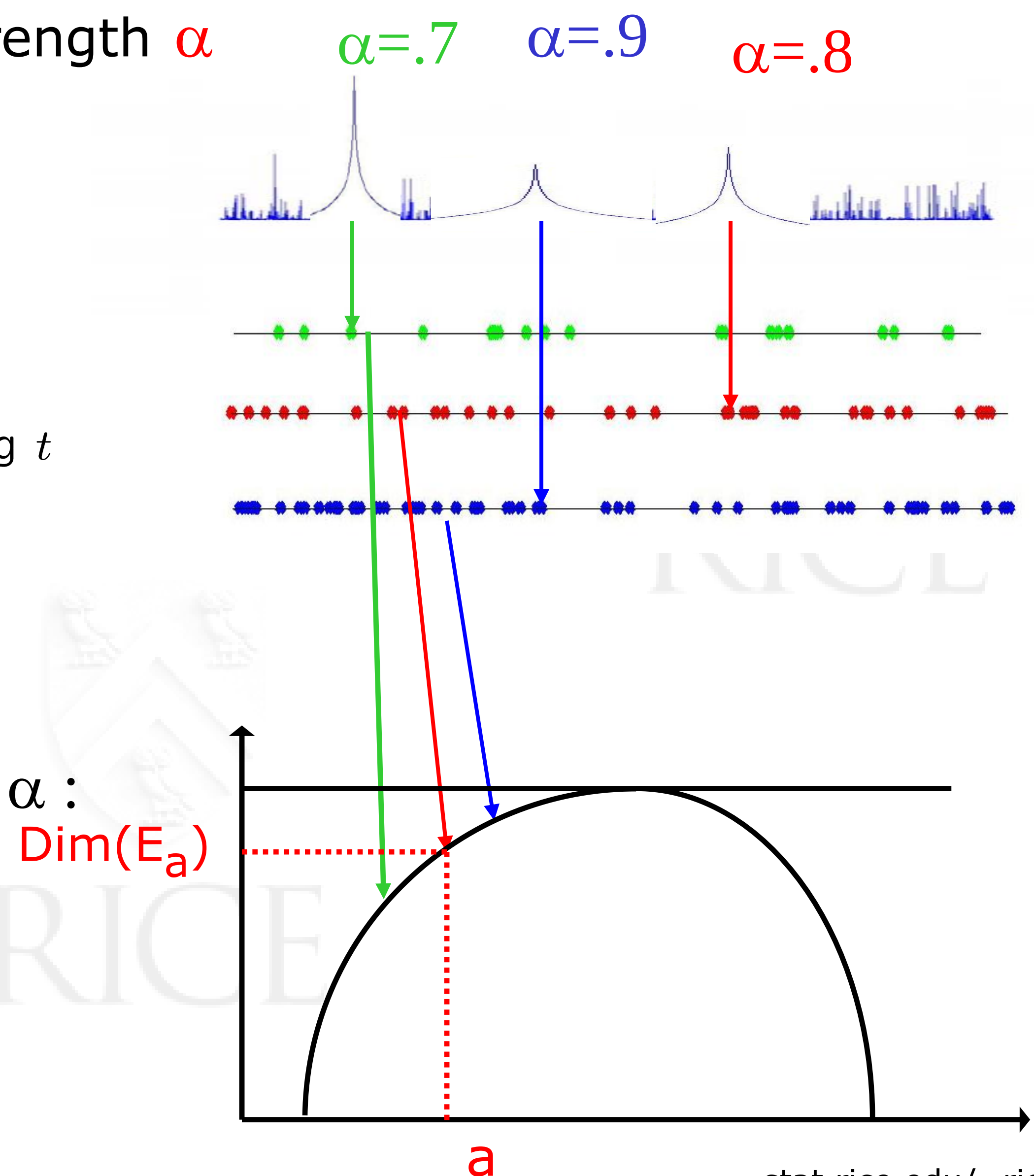
$I_n(t)$: dyadic interval containing t

$\Delta I_n(t)$: oscillation indicator
total increment over I_n,
max increment in I_n,
wavelet coefficients,...

- Collect points t with same α :

$$E_a := \{t : \alpha(t) = a\}$$

- Dim(E_a): Spectrum
→prelevance of α



Binomial

Recall

$$\alpha(t) := \liminf_n \alpha_n(t)$$

$$\alpha_n(t) = \frac{\log \Delta I_n(t)}{\log |I_n(t)|}$$

$$l_n(t) = \#\{k \le n : \epsilon_k = 1\}$$

We take dyadic partition:

$$I_n(t) = I(\epsilon_1 \ldots \epsilon_n) := [t_n, t_n + 1/2^n))$$

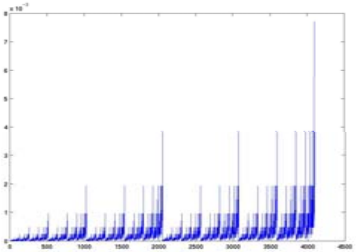
$$\Delta I_n(t) \;=\; X_{2^n t_n}^{(-n)} = p^{l_n(t)}(1-p)^{n-l_n(t)}$$

$$\alpha_n(t) = -\frac{n - l_n(t)}{n}\log_2(p) - \frac{l_n(t)}{n}\log_2(1-p)$$

Range of exponents:

$t = 0$: $l_n = 0$, $\alpha_n \to -\log_2(p)$ > 1: Smooth

$t = 1$: $l_n = n$, $\alpha_n \to -\log_2(1-p) < 1$: Bursty



"Typical" exponents

Recall

$$\alpha(t) := \liminf_n \alpha_n(t)$$

$$\alpha_n(t) = \frac{\log \Delta I_n(t)}{\log |I_n(t)|}$$

$$l_n(t) = \#\{k \leq n : \epsilon_k = 1\}$$

t=0, t=1 seem "atypical".

Intuition: for a "typical" t:

$$l_n(t) \simeq n/2$$

Rigorously: Law of Large Numbers

- Binary digits ϵ_k are independent, $P[\epsilon_k=0]= P[\epsilon_k=1]= ½$:
- t is uniformly distributed (i.e., with Lebesgue measure $\mathcal{L}$)
- $$\frac{l_n(t)}{n} = \frac{1}{n}\sum_{k=1}^{n} \epsilon_k \to \mathbb{E}_{\mathcal{L}}[\epsilon] = 1/2$$

- "Typical" exponent:

$$\boxed{\begin{aligned} \alpha_n(t) &= -\frac{n-l_n(t)}{n}\log_2(p) - \frac{l_n(t)}{n}\log_2(1-p) \\ &\to a_0 := -\frac{1}{2}\log_2(p) - \frac{1}{2}\log_2(1-p) \quad > 1 \end{aligned}}$$

A first point on the Spectrum

Conclusion:

- At almost all locations we have a_0, so:

$$\boxed{\dim E_{a_0} = 1}$$

- “Where” or “how many” are the other exponents?

Large Deviations

and the
Multifractal Formalism

Counting via Large Deviations

Recall

$$I_n(t) = I(\epsilon_1 \dots \epsilon_n)$$

$$\alpha_n(t) = \frac{\log \Delta I_n(t)}{\log |I_n(t)|}$$

$$E_a := \{t : \alpha(t) = a\}$$

- Notation:
  - Number of dyadic intervals with exponent ~ a:

$$N_{n,\delta}(a) := \#\{(\epsilon_1 \dots \epsilon_n) : a - \delta \le \alpha_n(\epsilon_1 \dots \epsilon_n) < a + \delta\}.$$

  - Partition sum: a microscope inspired by LDP

$$S_n(q) := \sum_{\epsilon_1 \dots \epsilon_n} |\Delta I_n(\epsilon_1 \dots \epsilon_n)|^q = \sum_{\epsilon_1 \dots \epsilon_n} |2^n|^{q\alpha_n(\epsilon_1 \dots \epsilon_n)}.$$

  - Assume powerlaws:

$$N_{n,\delta}(a) \sim 2^{nf(a)} \qquad S_n(q) \sim 2^{-n\tau(q)}$$

  - Typically (LDP) $\boxed{f(a) = \inf_q (qa - \tau(q))}$

LDP and the Legendre transform

Recall

$$I_n(t) = I(\epsilon_1 \dots \epsilon_n)$$

$$\alpha_n(t) = \frac{\log \Delta I_n(t)}{\log |I_n(t)|}$$

$$E_a := \{t : \alpha(t) = a\}$$

- Finding the dominating terms in S(q):

$$\begin{aligned}
2^{-n\tau(q)} \sim S_n(q) &= \sum_{(\epsilon_1 \dots \epsilon_n)} |\Delta I_n(\epsilon_1 \dots \epsilon_n)|^q \\
&= \sum_{l=1}^{m} \sum_{\alpha_n(\epsilon_1 \dots \epsilon_n) \in [l\delta - \delta/2, l\delta + \delta/2]} |\Delta I_n(\epsilon_1 \dots \epsilon_n)|^q \\
&\sim \sum_{l=1}^{m} N_{n,\delta/2}(l\delta) \cdot 2^{-nql\delta} \\
&\sim \sum_{l=1}^{m} 2^{-n(ql\delta - f(l\delta))} \\
&\sim 2^{-n(\inf_a (qa - f(a)))}
\end{aligned}$$

- …shows that τ and f are Legendre pairs

$$\tau(q) = \inf_a (qa - f(a)) \qquad f(a) = \inf_q (qa - \tau(q))$$

Legendre spectrum

- Thm: provided $\alpha_n(t)$ are bounded we have

$$\boxed{f(a) = \tau^*(a) \qquad \text{for } a = \tau'(q).}$$

- …in other words

$$\#\{(\epsilon_1 \ldots \epsilon_n) : a - \delta \leq \alpha_n(\epsilon_1 \ldots \epsilon_n) < a + \delta\}$$
$$\sim 2^{n \inf_q (qa - \tau(q))}$$

- …and the multifractal spectrum is the Legendre transform of the partition scaling exponent

Legendre transform 101

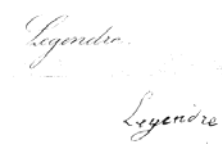
- Elementary calculus:

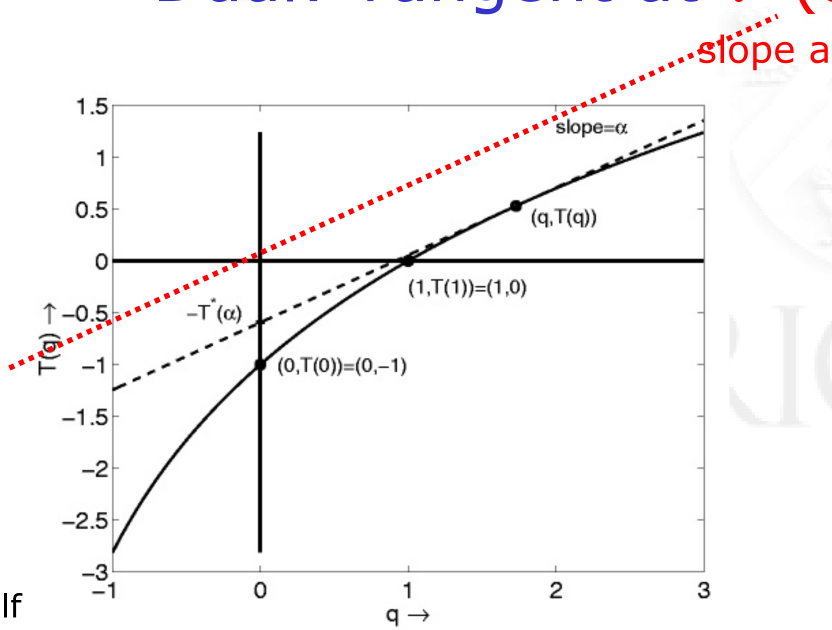
$$\tau^*(a) := \inf_q (qa - \tau(q)) = \overline{q}a - \tau(\overline{q})$$

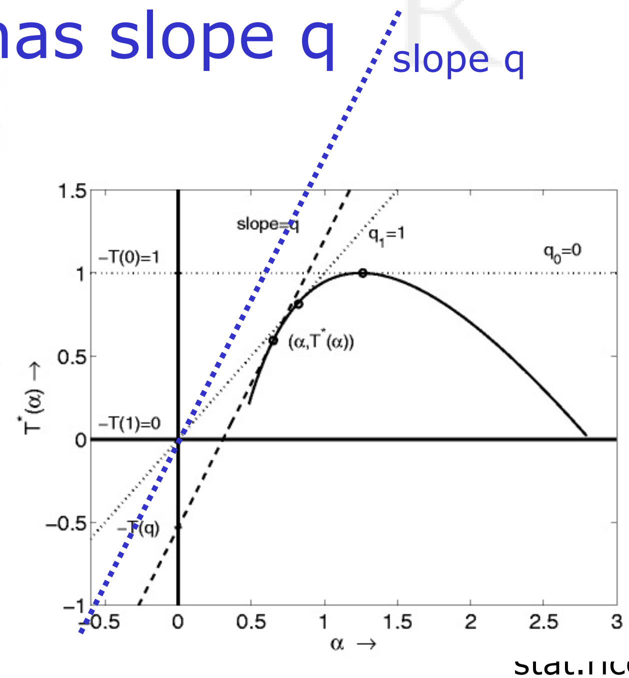
where $\overline{q}$ is defined by $a = \tau'(\overline{q})$

  - Draw tangent of slope a to τ(q).
  - The intersection with y-axis yields -τ*(a)
  - Dual: Tangent at τ*(a) has slope q

Legendre







Binomial Spectrum

continued

Multifractal analysis of the Binomial

$$\begin{aligned} S_n(q) &= \sum_{\epsilon_1 \ldots \epsilon_n} |\Delta I_n(\epsilon_1 \ldots \epsilon_n)|^q \\ &= \sum_{\epsilon_1 \ldots \epsilon_n} [p^{n-l_n(\epsilon_1 \ldots \epsilon_n)}(1-p)^{l_n(\epsilon_1 \ldots \epsilon_n)}]^q \\ &= \sum_{l=0}^{n} \begin{pmatrix} n \\ k \end{pmatrix} [p^{n-l}(1-p)^l]^q \\ &= [p^q + (1-p)^q]^n. \end{aligned}$$

- Partition function

$$\tau(q) = \lim_{n\to\infty} -\frac{1}{n} \log_2 S_n(q) = -\log_2[p^q + (1-p)^q]$$

- Via Legendre: Most often we see exponent a_0 such that $f(a_0)$ is maximal. This happens where the tangent is horizontal, thus where q=0. So, as before:

$$a_0 = \tau'(0) = -\frac{1}{2}\log_2(p) - \frac{1}{2}\log_2(1-p) \quad > 1$$

Insight from Large Deviations

- From steepest ascent:

$$
\begin{aligned}
S_n(q) &= \sum_{\epsilon_1 \ldots \epsilon_n} |\Delta I_n(\epsilon_1 \ldots \epsilon_n)|^q \simeq 2^{-n(\inf_a(qa - f(a))} \\
&= 2^{-n(q\bar{a} - f(\bar{a}))} \simeq \sum_{\alpha_n(\epsilon_1 \ldots \epsilon_n) \simeq a} |\Delta I_n(\epsilon_1 \ldots \epsilon_n)|^q
\end{aligned}
$$

  - Dominant terms in $S_n(q)$, for fixed q, are the ones with

$$
\alpha_n(\epsilon_1 \ldots \epsilon_n) = \frac{\log \Delta I_n}{\log |I_n|} \simeq \bar{a} = \tau'(q)
$$

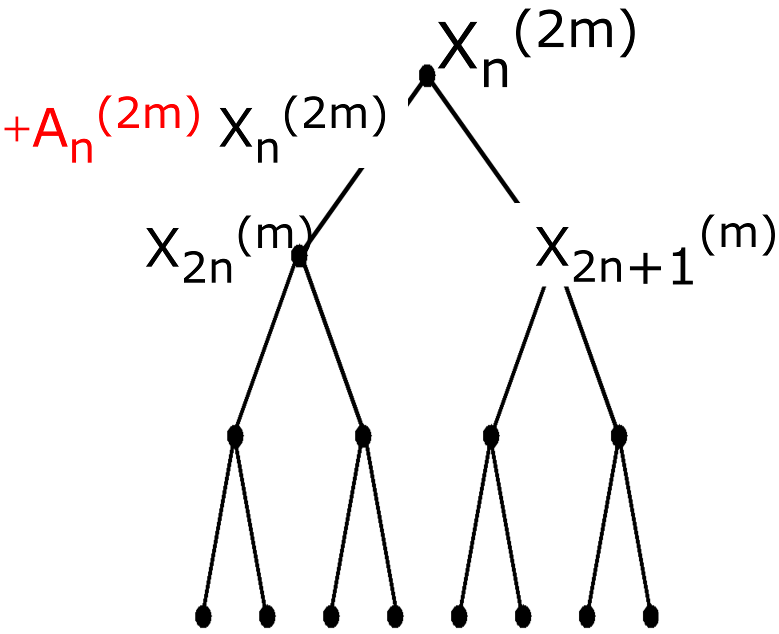
    - …and vice versa: these terms contribute such that

$$
S_n(q) \simeq 2^{-n\tau(q)} = (p^q + (1-p)^q)^n
$$

    - For the Binomial these correspond to choosing digits in the ratio p^q to $(1-p)^q$

Spectrum of the MWM

Multifractal Wavelet Model



Choose *independent* r.v. $A_{(j)}$ *symmetrically* distributed in $[-1,1]$.

Define recursively

$$W_{j,k} = A_{(j)} \cdot U_{j,k}$$

Resulting *stationary* (1st order) series

$$X_k \stackrel{d}{=} U_{J_0,0} \cdot \prod_{j=J_1}^{J_0} (1 + A_{(j)})/2$$

$$X_{2n}^{(m)} = X_n^{(2m)} \cdot (1 + A_n^{(2m)})/2$$

$$X_{2n+1}^{(m)} = X_n^{(2m)} \cdot (1 - A_n^{(2m)})/2$$

Multifractal analysis of the MWM

- Partition function

$$\frac{-1}{n}\log_2 \underbrace{\sum_{i=1}^{n}\sum_{k_i=0,1}}_{\text{sum over the } 2^n \text{ dyadic points of order } n} \mathbb{E}\left(\prod_{i=1}^{n}(1+(-1)^{k_i'}A_{(-i,k_i)})\right)^{q}$$

Binomial formula

$$\doteq \frac{-1}{n}\log_2 \prod_{i=1}^{n}\left(\mathbb{E}(1+A_{(-i)})^q+\mathbb{E}(1-A_{(-i)})^q\right)$$

Symmetry of $A_{(-i)}$

$$\doteq -1-\frac{1}{n}\sum_{i=1}^{n}\log_2 \mathbb{E}[(1+A_{(-i)})^q]$$

$$\to -1-\log_2 \mathbb{E}[(1+A)^q] \qquad \text{provided } A_{(j)} \stackrel{distr}{\to} A.$$

- Special case of Beta-variables A

$A_{(j)} \simeq \beta(p_{(j)},p_{(j)})$ with $p_{(j)} \to p$ as $j \to -\infty$

$$\tau(q) = -1-\log_2 \frac{\Gamma(p+q)\Gamma(2p)}{\Gamma(2p+q)\Gamma(p)}.$$

Spectrum of self-similar processes

Mono-fractals

MFA of Self-similar processes

Assume Y is H-sssi with increments

$$X_k^{(n)} = Y(k2^{-n}) - Y((k-1)2^{-n})$$

$$\mathbb{E} \sum_{k=1}^{2^n} |X_k^{(n)}|^q = 2^n \mathbb{E}|X_1^{(n)}|^q = 2^{n-nqH} \mathbb{E}|X_1^{(1)}|^q,$$

Assume $\mathbb{E}|X_1^1|^q < \infty$ for $q_{\text{bot}} < q < q_{\text{top}}$.

fBm: $q_{\text{bot}} = -1$, $q_{\text{top}} = \infty$,

α-stable process: $q_{\text{bot}} = -1$, $q_{\text{top}} = \alpha$.

Then

$$\tau(q) = \begin{cases} qH - 1 & \text{for } q_{\text{bot}} < q < q_{\text{top}}, \\ -\infty & \text{else.} \end{cases}$$

Linear Spectrum!

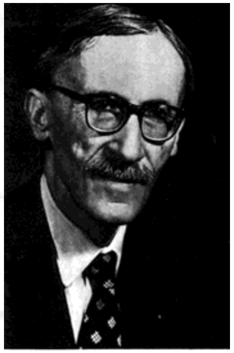
$$f(a) = \begin{cases} 1 + q_{\text{top}}(\alpha - H) & \text{for } \alpha < H \\ 1 + q_{\text{bot}}(\alpha - H) & \text{for } \alpha \geq H. \end{cases}$$

Wavelets

A powerful multiscale tool

History of wavelets

- Fourier series (1807)
- Levy (1930): Haar basis superior to Fourier for Brownian motion
- Weiss-Coifman (‘60-’80):
  - decompose functions into atoms
- Grossman-Morlet ’80: defined wavelets
- Mallat ’85: pyramidal algorithm, o.n. basis
- →Meyer: continuously diff wavelets
- →Daubechies: compactly supported wavelets



PAUL LÉVY 1886-1971

Ortho-normal Wavelets

- Multi-resolution analysis (Mallat, Daubechies):
  - There are compactly supported ψ and ϕ s.t.

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

$$\phi_{j,k}(t) \;:=\; 2^{-j/2}\,\phi\Big(2^{-j}t-k\Big)$$

  form orthonormal bases of $\mathcal{L}^2$.

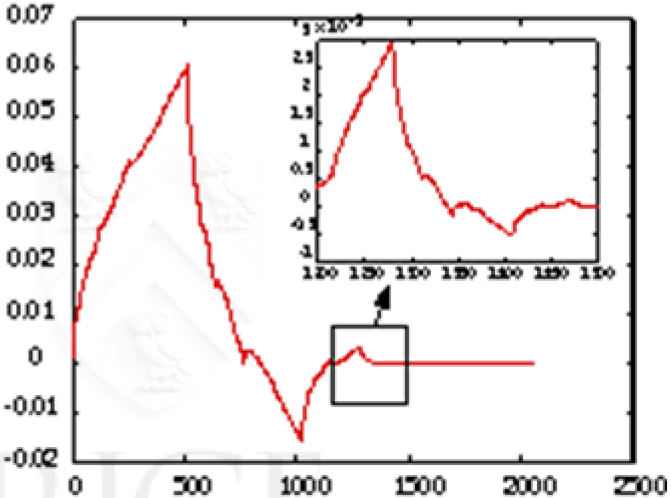
  - For X supported on $[0,2^{J_0}]$

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

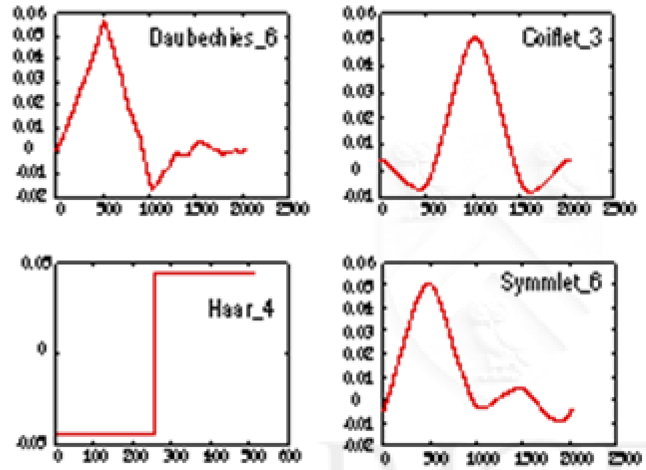
  with

$$W_{j,k} := \int X(t)\,\psi^*_{j,k}(t)\,dt \quad \text{and} \quad U_{j,k} := \int X(t)\,\phi^*_{j,k}(t)\,dt.$$

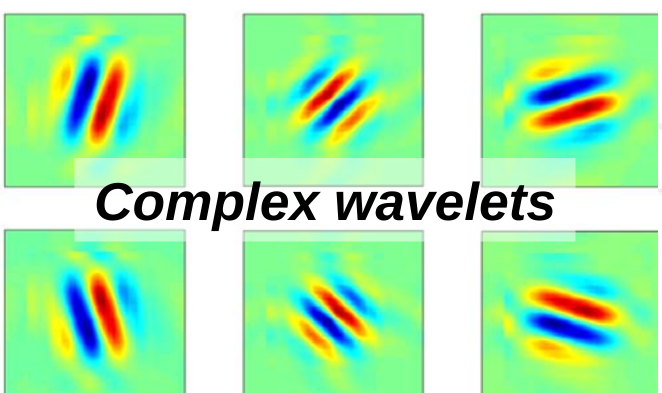
Wavelets: what they look like



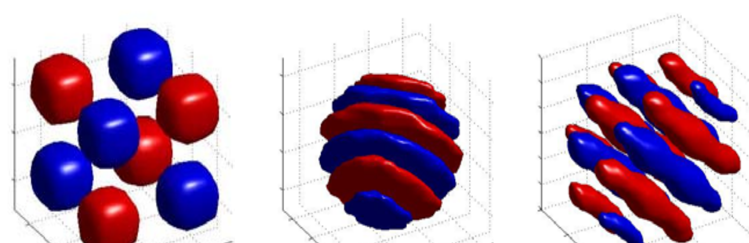
Daubechies 4 Mother wavelet



Daubechies 6 Coiflet 3
Haar 4 Symmlet 6



Complex wavelets



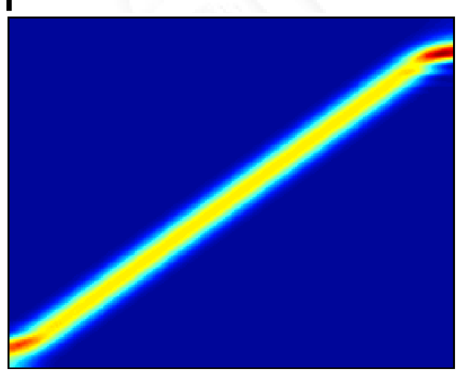
Multidimensional wavelets

Continuous wavelets

- Continuous rescaling of mother wavelet
- Continuous (redundant) set of coefficients
- Often used: Mexican hat $(\exp(-x^2))''$

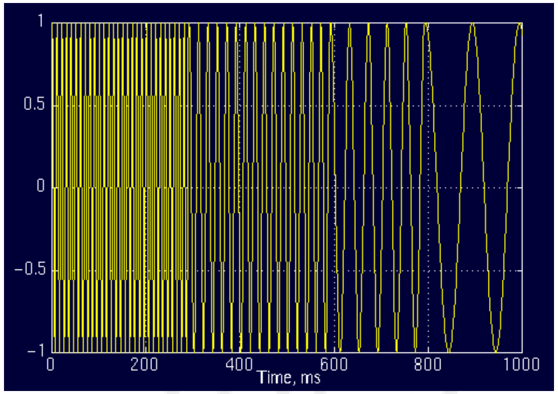
$$T(a,t) = \frac{1}{a}\int X(s)\psi\left(\frac{s-t}{a}\right)ds$$

- Form of a convolution → Fourier, Parseval



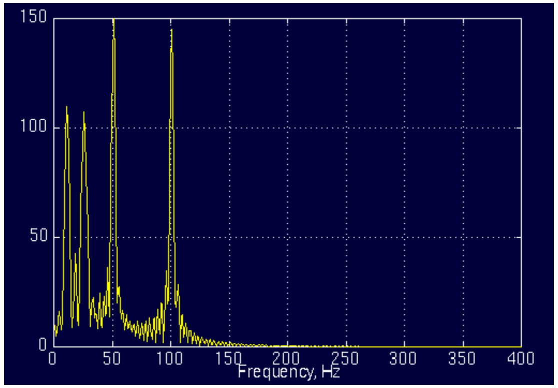
Wavelet vs Fourier

- Fourier
  - timing information is hidden in the phase
  - sin(t) and cos(t) are not localized in time



Signal

- Power spectral density
  - Identifies frequency content only, but not their location
  - Relation to Auto-correlation

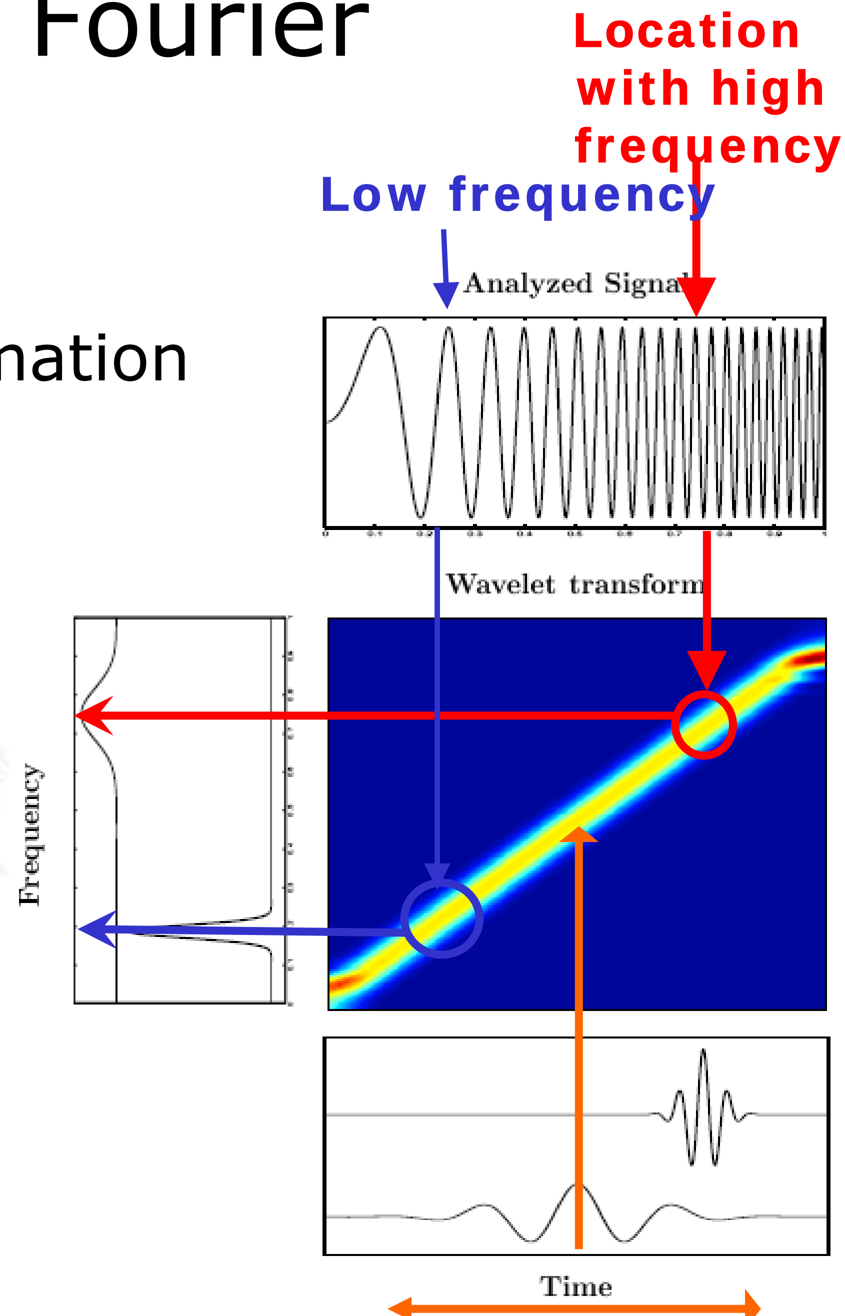


Fast Fourier Transform

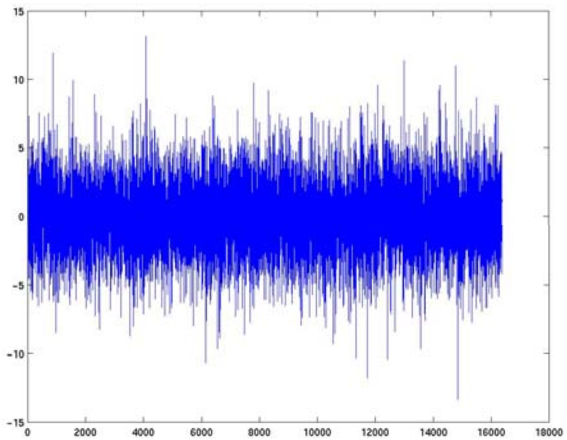
Wavelet vs Fourier

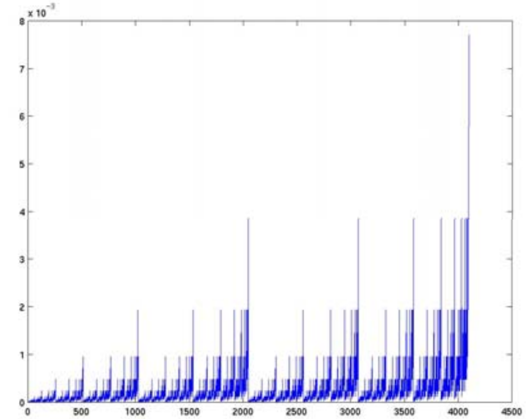
- Power spectrum
  - provides no timing information
  - sin(t) and cos(t) are not localized in time

- Wavelets are localized
  - both in time
  - and in frequency



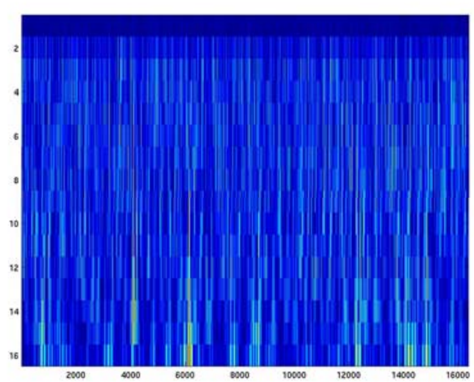
Toy examples

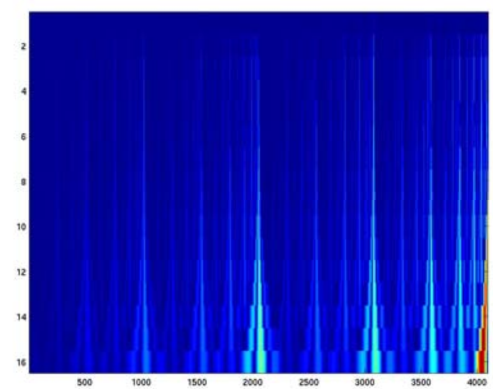




White noise

Cascade



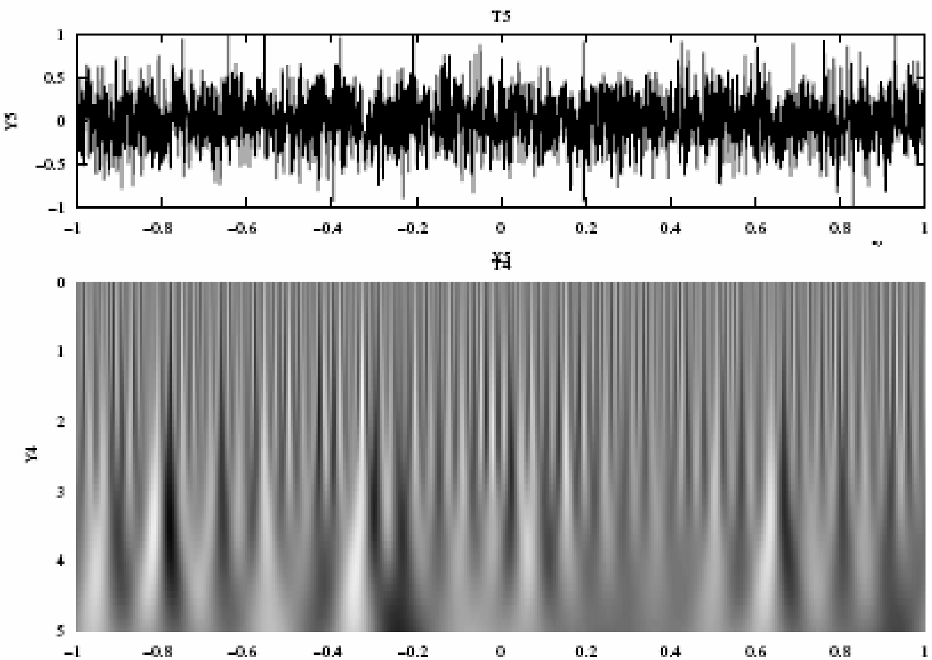


Wavelet trafo indicates:
Mono-fractal

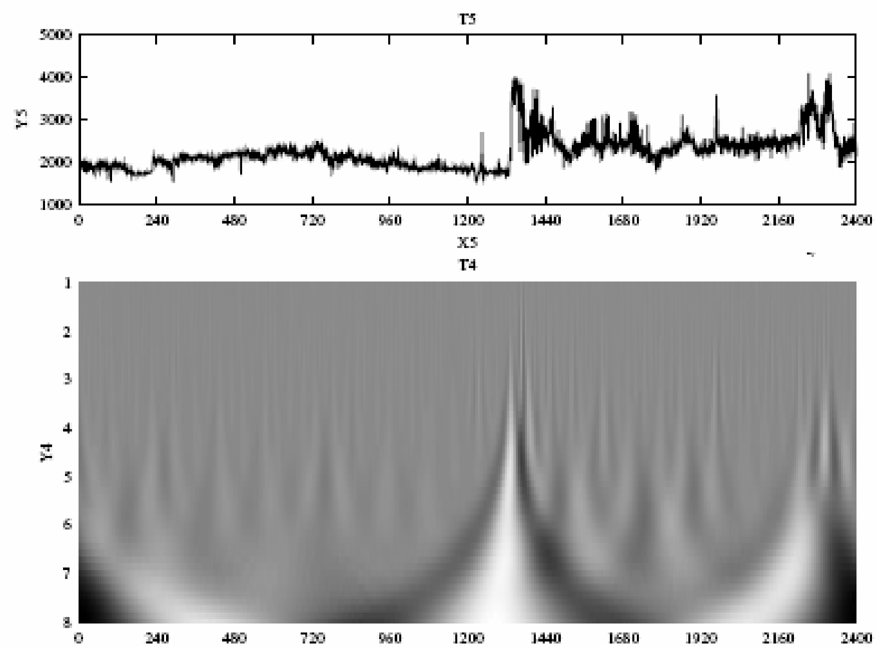
Wavelet trafo indicates:
Multi-fractal

Continuous wavelet at work

- Wavelets are an excellent tool to identify local frequency content



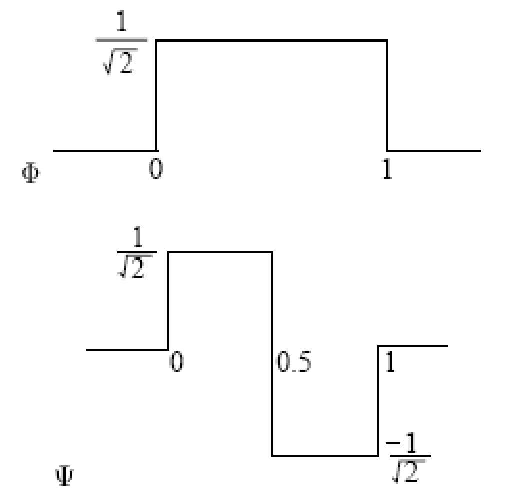
Fractional Gaussian Noise



Geological Well data

Haar wavelet

- Haar wavelet (See plot)



Haar scaling Φ and wavelet Ψ functions.

- Clearly

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

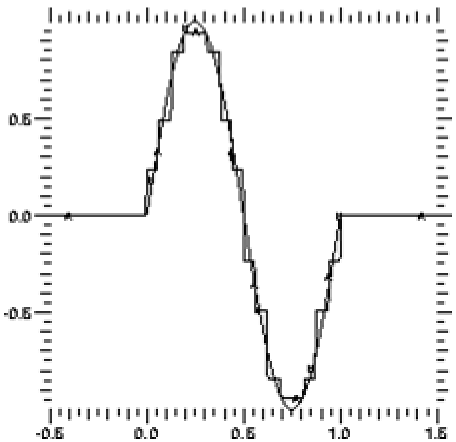
$$\phi_{j,k}(t) \;:=\; 2^{-j/2}\,\phi\big(2^{-j}t-k\big)$$

form orthonormal bases of $\mathcal{L}^2$.

- Coefficients
  - $$W_{j,k} = 2^{-j/2}\left[\int_{2^jk}^{2^j(k+\frac{1}{2})} f(x)dx - \int_{2^j(k+\frac{1}{2})}^{2^j(k+1)} f(x)dx\right]$$
  - For $Z(t)=\int_0^t f dx,\ X_k^{(m)} = Z(km)-Z(km-m)$

$$W_{j+1,k} = X_{2k}^{(2^j)} - X_{2k+1}^{(2^j)}$$

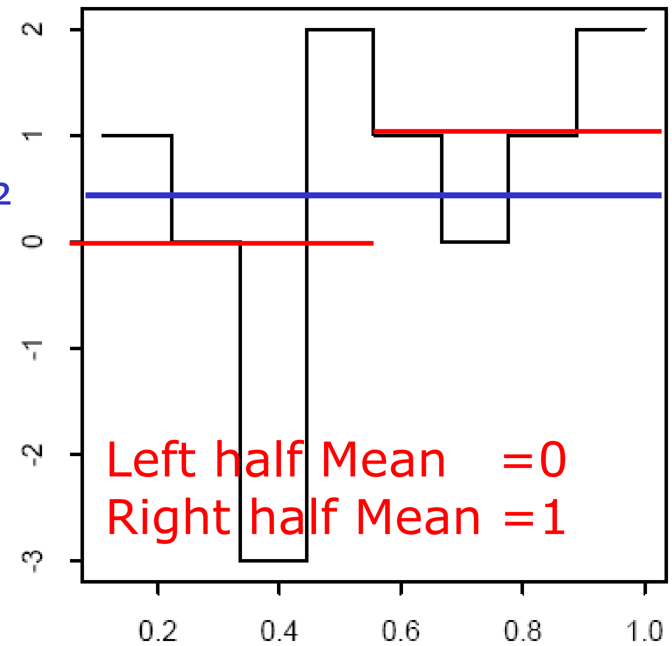
Haar wavelet at work



Haar approximation of sin(x)

Haar approximation of some f(x)

Overall Mean= ½



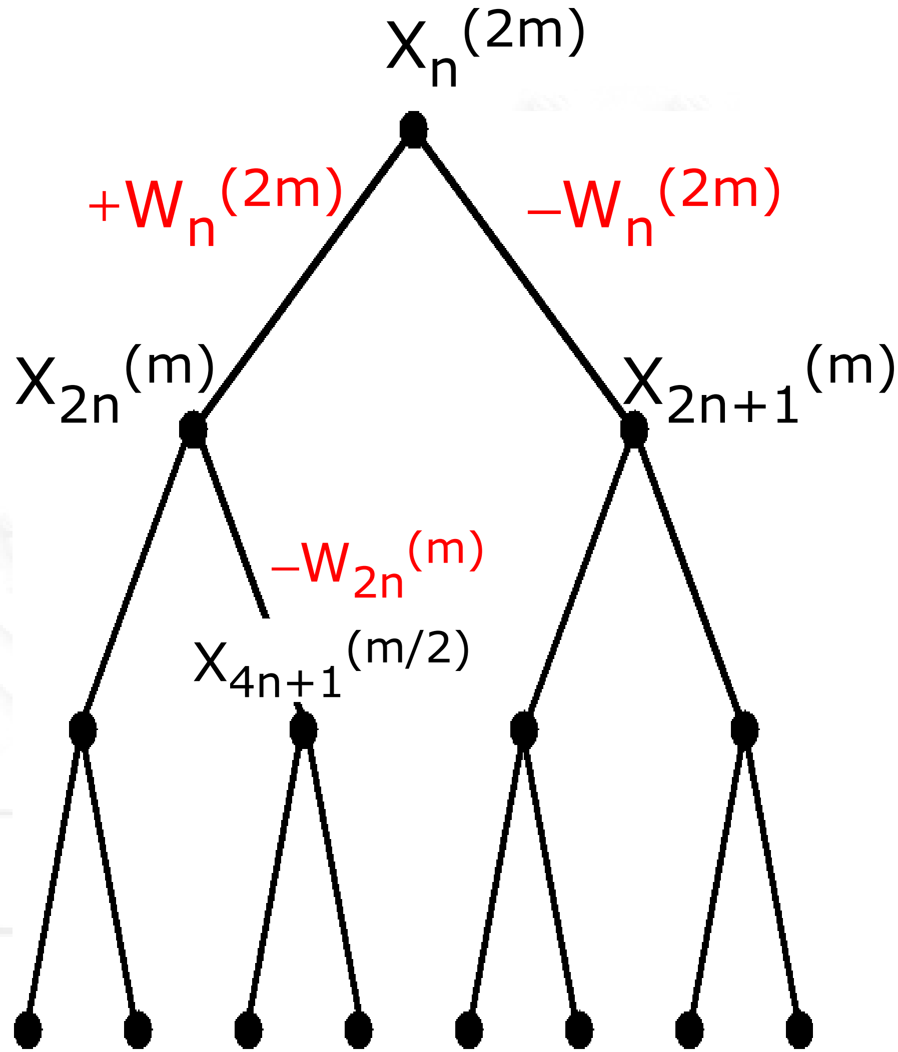
$$f = \frac{1}{2}\phi - \frac{1}{2}\psi_{00} + \frac{1}{2\sqrt{2}}\psi_{10} - \frac{1}{2\sqrt{2}}\psi_{11} + \frac{1}{4}\psi_{20} - \frac{5}{4}\psi_{21} + \frac{1}{4}\psi_{22} - \frac{1}{4}\psi_{23}$$

Additive Tree is Haar Model

Synthesis:

- Start at root
- Flow down the tree
- Additive, independent innovations $W_n^{(m)}$
- these are essentially the Haar wavelet coefficients

- Idea: use any wavelet
  - The better the frequency response, the better the spectral approximation to fBm



LRD vs. Large Deviations

Large vs Small scales

Doubly stochastic modeling

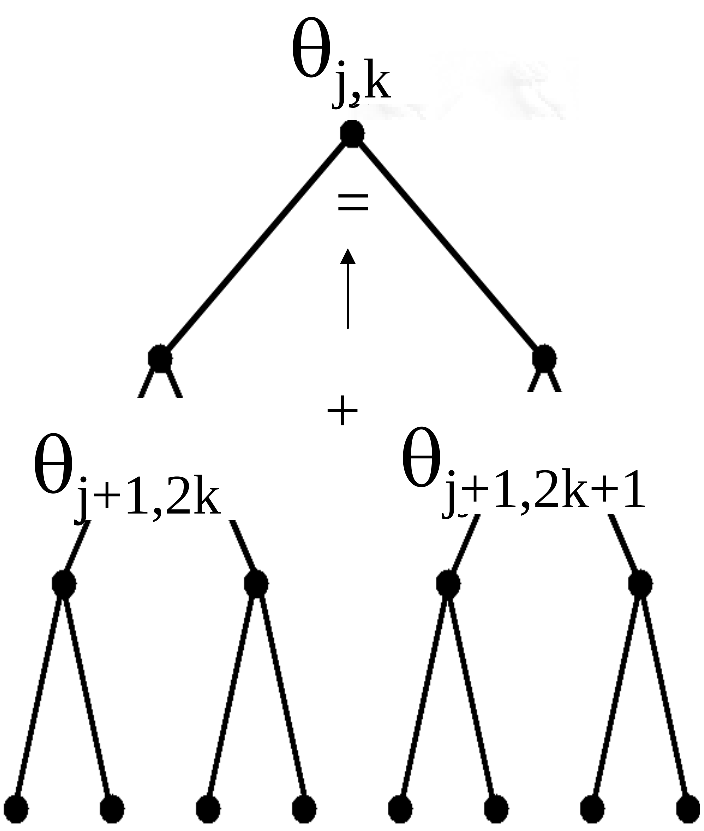
Setting:

- Background process $\theta_{j,k}$
- Given $\theta_{j,k}$ the observed multiscale loads $U_{j,k}$ are *independent* of *mean* $\theta_{j,k}$

- $E[U_{j,k} \mid \theta] = \theta_{j,k}$

- $\theta_{j,k}$ fill a multiscale tree

$$\theta_{j+1,2k} + \theta_{j+1,2k+1} = \theta_{j,k}$$



Gaussian versus Poisson

- Gaussian

$U_{j,k} = \mathcal{N}(\theta_{j,k}, \sigma^2_{j,k})$

- Poisson

$U_{j,k} = \mathcal{P}(\theta_{j,k})$

Iteration scheme for synthesis

$U_{j+1,2k} \mid U_{j,k} =$

$\mathcal{N}(U_{j,k}/2 + (\theta_{j+1,2k} - \theta_{j+1,2k+1})/2 \,,\, \sigma^2_{j,k}/2)$

additive innovation

$\mathcal{B}inom(U_{j,k} \,,\, \theta_{j+1,2k} / \theta_{j,k})$

multiplicative innovation

Scaling from a modeling perspective

- ON-OFF limits and LRD
  - User driven: heavy tail file sizes
  - Additive, Gaussian
  - Large scales
- Cascades and multifractal scaling
  - Network driven: heterogeneity of RTT
  - Multiplicative, log-Normal
  - Small scales