(Numerical) Investigations into Physical Power-law Models of Internet Traffic Using the Renormalization Group

IFORS Triennial Conference

Honolulu, HI, July 11–15, 2005

Neil Gunther

Performance Dynamics Company^{SM}

Castro Valley, California

`www.perfdynamics.com`

Possibly Perverse Packet History

Bellcore traces kicked off a "paper-mill" industry c.1994

Persistent burstiness: LRD pkt trains $\Rightarrow$ *BIG* queues

FUD: Queueing theory is dead, fractal traffic, power laws, non-Poisson arrivals, size-dependent service, ...
$\Rightarrow$ Internet (I & II) can't be modeled!

Multitudinous math: $M/G/\infty$, chaos thy, large devs, ...

Stop the tram! ...

1. Where is LRD being measured? (Rarely discussed)
   But see e.g., (Downey 2001)
2. Industry network engineers **don't** see it! (cf. SYN flooding)
3. How much should I care? (Never discussed) Only pkt level effects.

1-Pareto, 2-Pareto, . . . ? (Fischer et al. 2004)

With four parameters I can fit an elephant, with five I can make his trunk wiggle. —J. von Neumann

Type	$\bar{F}(x)$	$f(x)$	Domain	Parameters
3-P	$\left(\frac{\beta}{\beta-k+x}\right)^{\alpha}$	$\frac{\alpha}{k}\left(\frac{k}{\beta-k+x}\right)^{1+\alpha}$	$x \geq k$	$\alpha, \beta > 0, k \geq 0$
2-P	$\left(\frac{k}{x}\right)^{\alpha}$	$\frac{\alpha k^{\alpha}}{x^{1+\alpha}}$	$x \geq k$	$\alpha > 0, \beta = k > 0$
1-P	$\frac{1}{(1+x)^{\alpha}}$	$\frac{\alpha}{(1+x)^{1+\alpha}}$	$x \geq 0$	$\alpha > 0, \beta = 1, k = 0$

With α(shape), β(scale), k(location). Conventional choice is 2-P:

$$
\begin{aligned}
F(k, \alpha, x) &= 1 - \left(\frac{k}{x}\right)^{\alpha} \\
f(k, \alpha, x) &= \frac{\alpha k^{\alpha}}{x^{1+\alpha}} \equiv dF(x)/dx \\
\bar{F}(k, \alpha, x) &\equiv 1 - F(x) = \left(\frac{k}{x}\right)^{\alpha}
\end{aligned}
$$

Motivations

Profs. Shortle and Gross are to blame for this talk ☺

Is there a way to decide about n-P power law models?

Multi-parameter (unphysical) models are something I've seen before in computer/networking scalability analysis

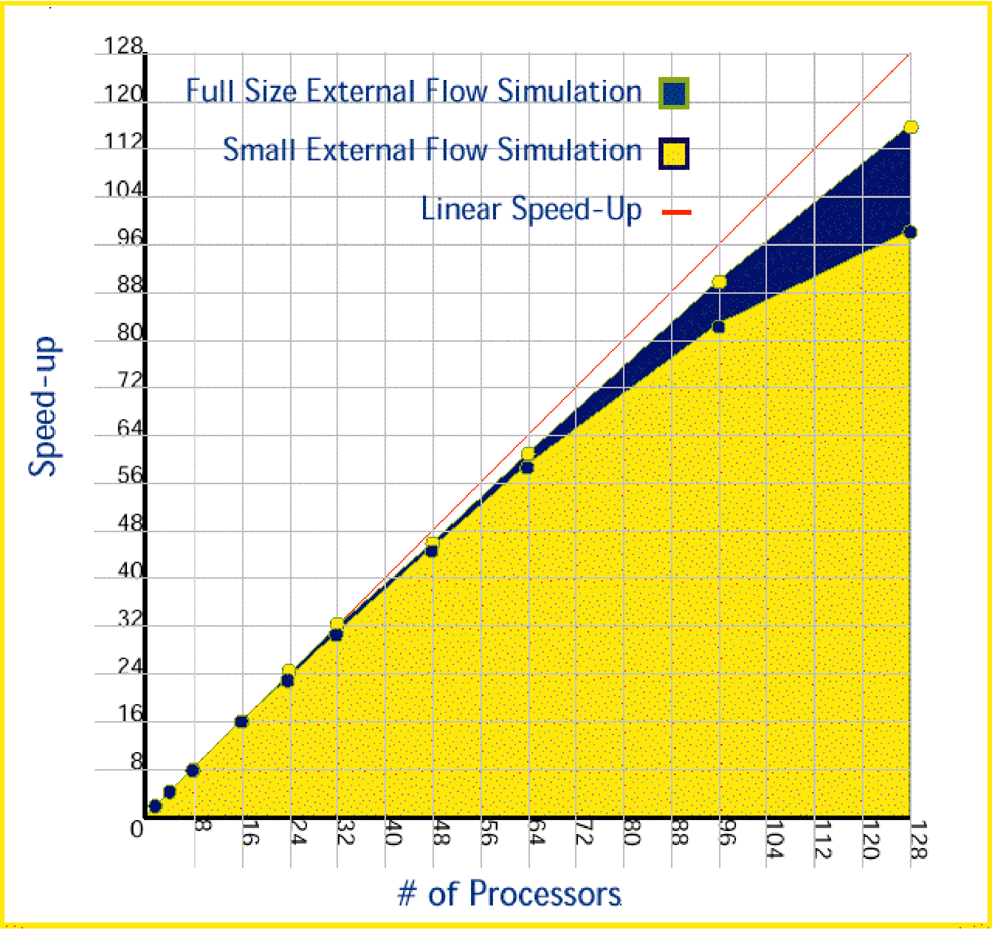
Power laws and scaling arise from
Renormalization (semi-)**group transformations**
$\Rightarrow$ generalized homogeneous functions

If we try to apply RGT as a tool, what happens?
(Not easy $\Rightarrow$ numerical studies)

Computer Scalability

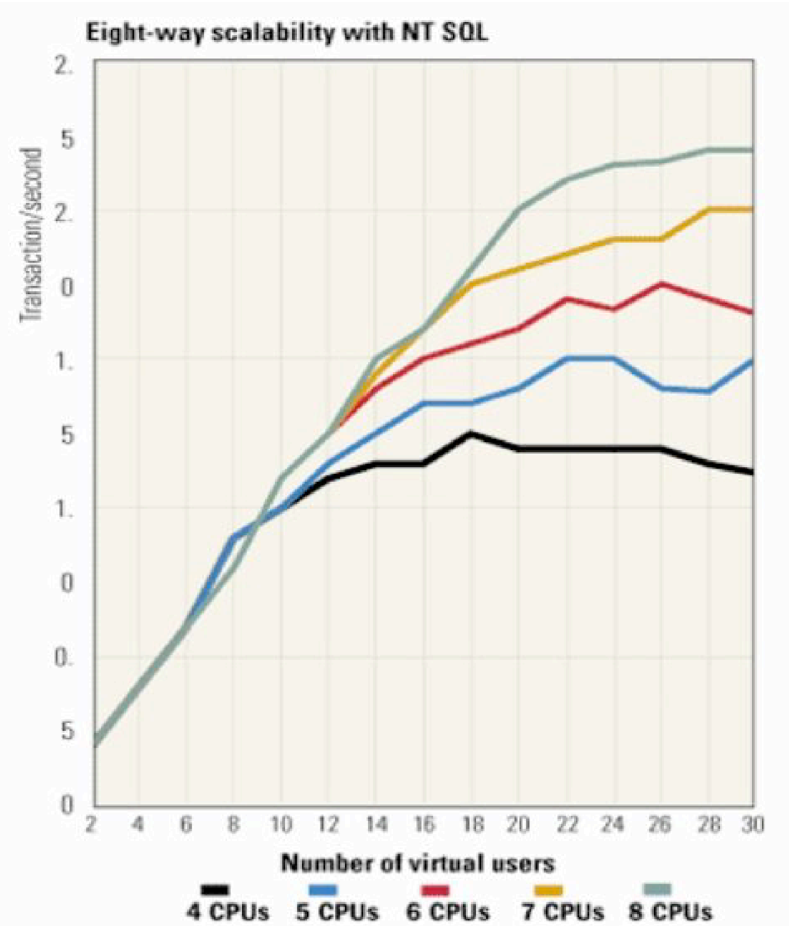
Hardware Scalability:

Fix (N) users, vary (p) processors



Software Scalability:

Fix (p) processors, vary (N) users



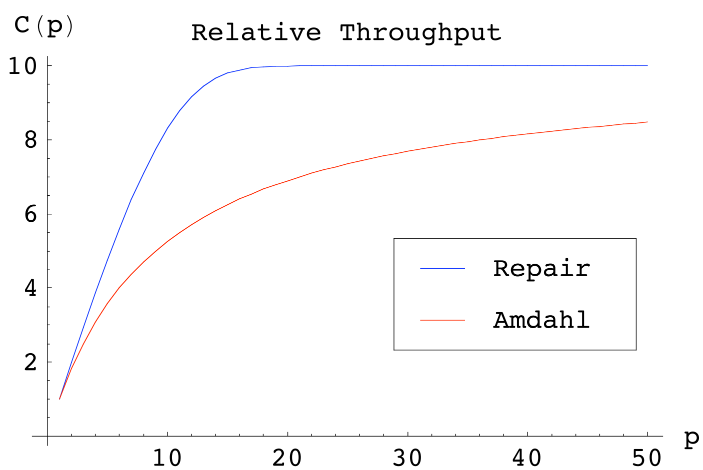
Computer Scaling Models I Have Known

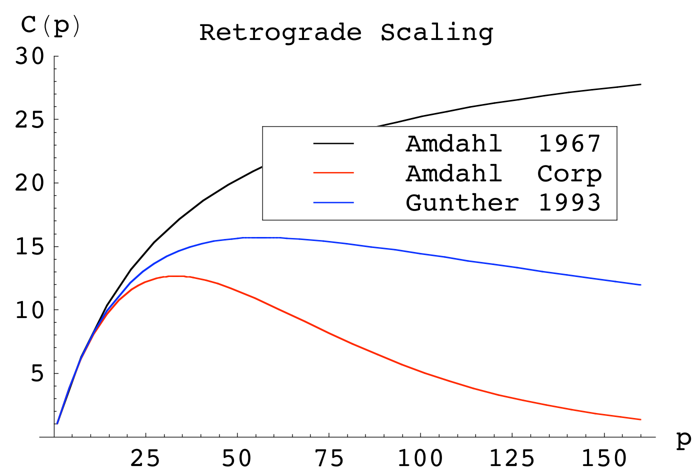
(but not necessarily loved)

Scaling Model	Parameters	Genesis
$S_\sigma(p) = \frac{p}{1+\sigma(p-1)}$	$0 \leq \sigma < 1$	Gene (Amdahl 1967)
$S_\phi(p) = \frac{1-\phi^p}{1-\phi}$	$\phi \lesssim 1$	Unknown
$S_\gamma(p) = p - \gamma p(p-1)$	$0 \leq \gamma < 1$	Gunther 1991
$S_\lambda(p) = \frac{p}{1+\sigma[(p-1)+\lambda p(p-1)]}$	$0 \leq \sigma, \lambda < 1,$	(Gunther 1993)
$S_\alpha(p) = p(1-\alpha)^{(p-1)}$	$0 \leq \alpha < 1$	Amdahl Corp. 1999

- Shown that $S_\phi(p)$ is **unphysical**
  i.e., contradicts (Coxian) queueing theory (Gunther 2002)
- But $S_\sigma(p)$ and $S_\lambda(p)$ **are** physical
  i.e., consistent with queueing theory (Gunther 2004)

Parametric Scalability Models





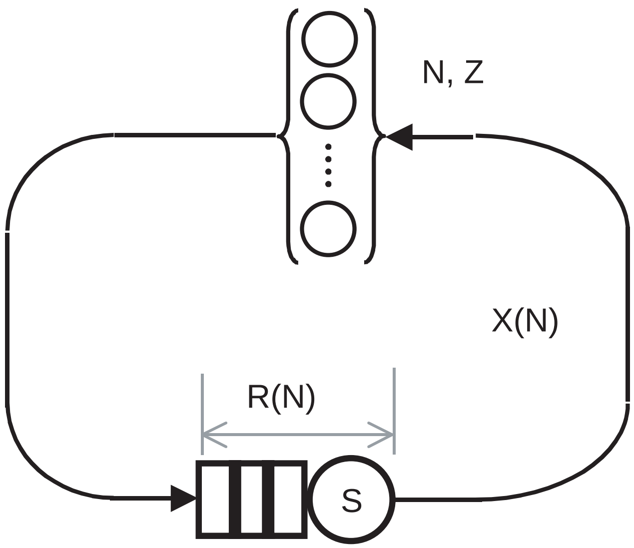
$$X_{mva}(p) = \frac{p}{R(p) + Z}$$

$$S_{\alpha}(p) = p(1-\alpha)^{(p-1)}$$

$$S_{\sigma}(p) = \frac{p}{1 + \sigma(p-1)}$$

$$S_{\lambda}(p) = \frac{p}{1 + \sigma[(p-1) + \lambda p(p-1)]}$$

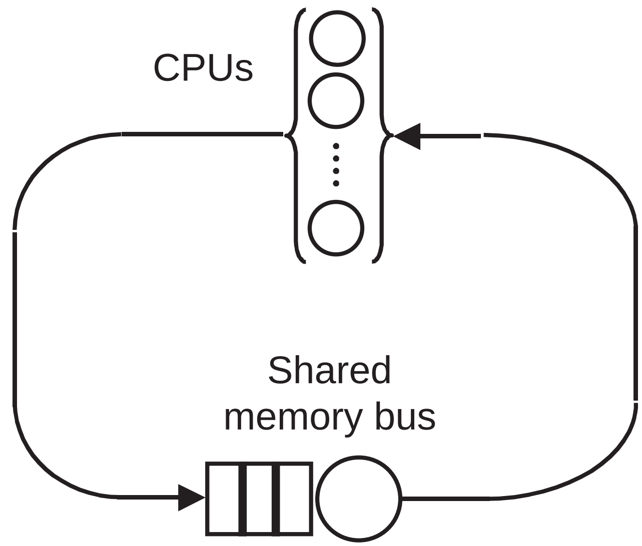
Amdahl Meets the Repairman (Gunther 2004)



Repairman Model

$$S_{mva}(p) = \frac{p(D+Z)}{R(p)+Z}$$

where $S(p) = \frac{X(p)}{X(1)}$



Multiprocessor Model

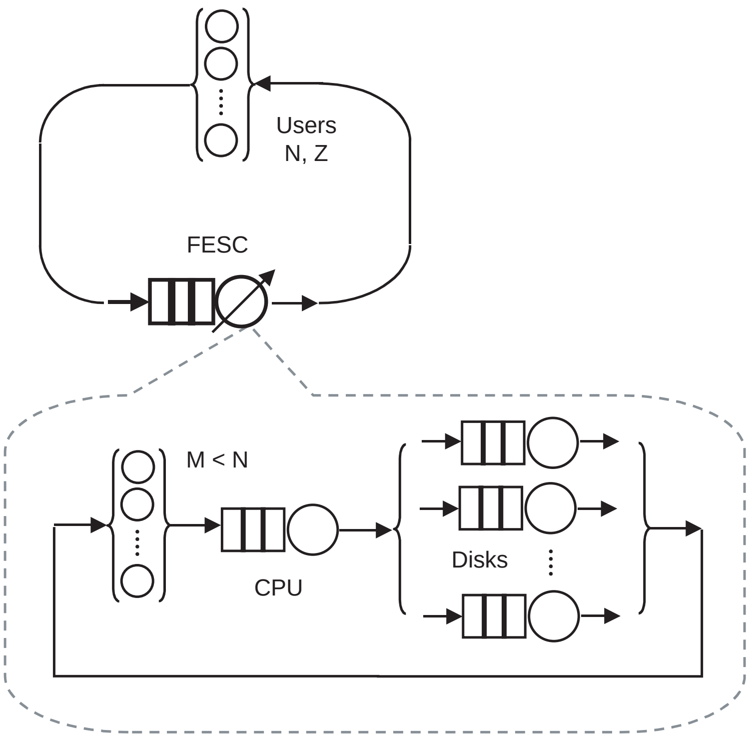
$$S_{sync}(p) = \frac{p(D+Z)}{pD+Z} = \frac{p}{1+\sigma(p-1)}$$

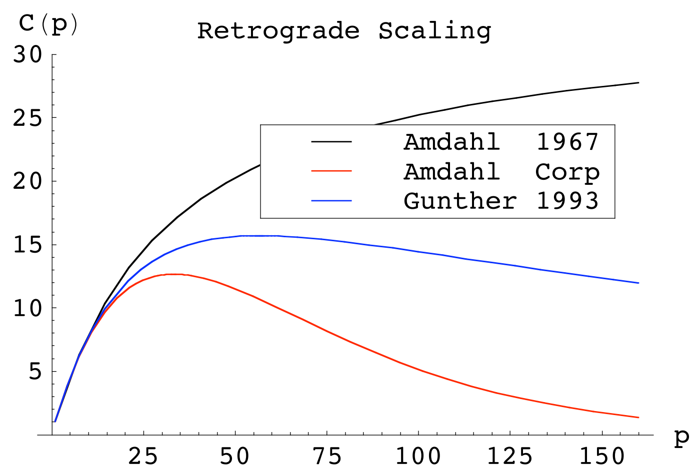
where $\sigma = \frac{D}{D+Z}$

Load-Dependent Queues (Gunther 2005)

$$S_\lambda(p) = \frac{p}{1 + \sigma[(p-1) + \lambda p(p-1)]}$$

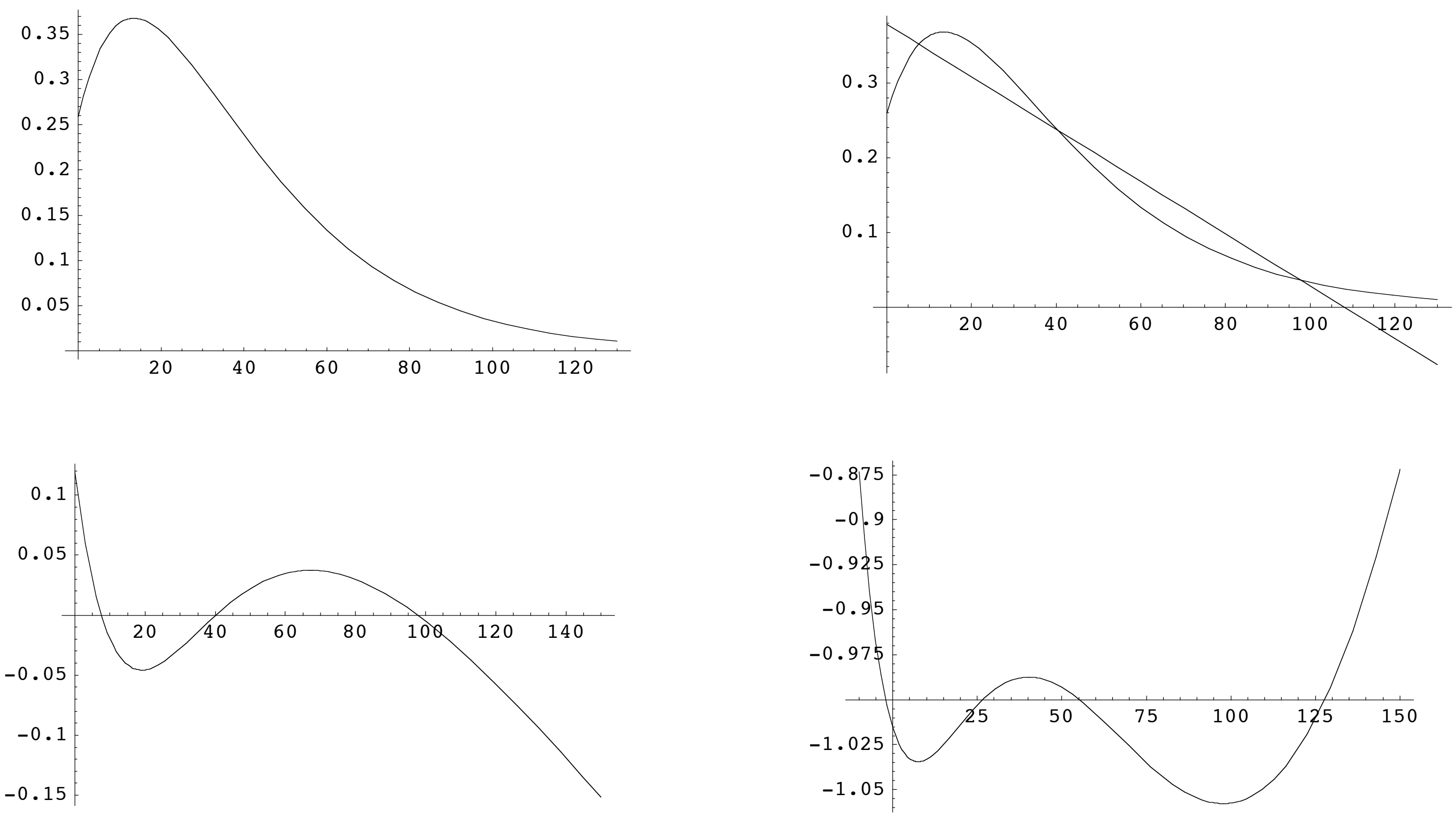
$$S_\alpha(p) \simeq p e^{-\alpha(p-1)} \text{ as } p \to \infty$$





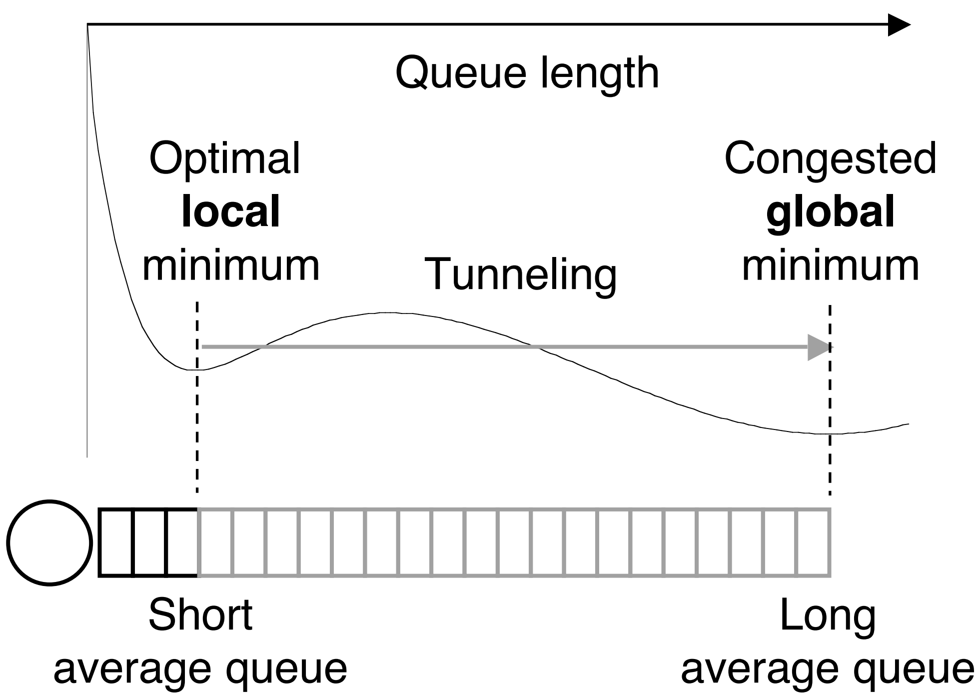
New observation $S_\alpha(p)$ is ALOHA-like

ALOHA Network Stability



Kleinrock, Metcalfe, Weiss, Nelson, Gunther, (Ganesh et al. 2004)

ALOHA as a Queue (Gunther 1990)



Internet routers did congest c.1986 → TCP slow-start

Critical behavior → VM, ALOHA *universality*

Bellcore Ethernet Measurements

This case seems to have been forgotten! (since c.1989)

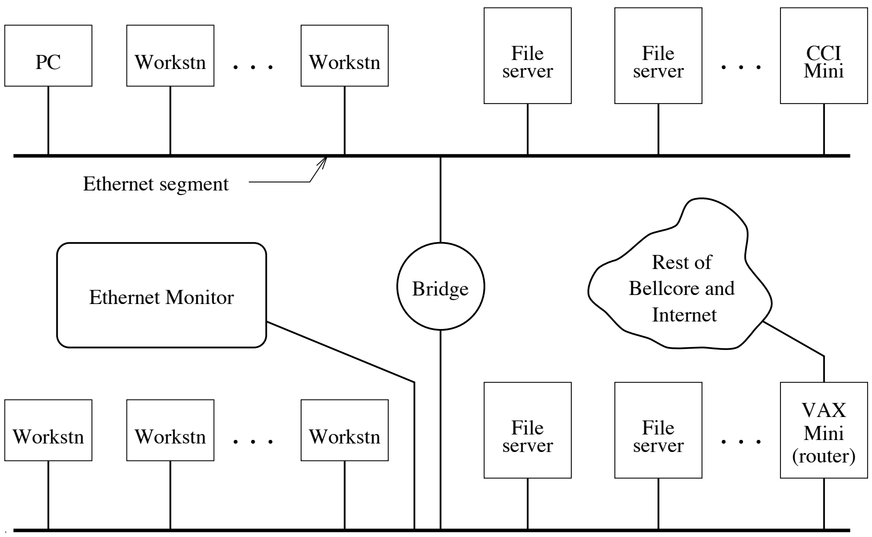


Figure 2.2.1. Network from which the August and October 1989 measurements were taken.

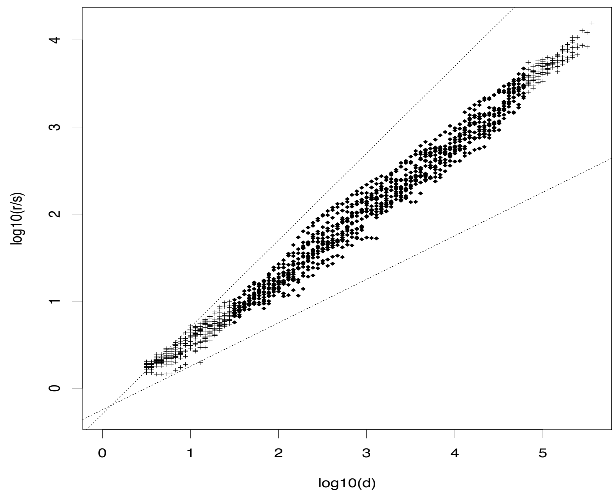


Figure 4.2.1 (b). Pox plot of R/S for sequence AUG89.MB. The plot tightly clusters around a straight line whose asymptotic slope clearly lies between the slopes 0.5 (lower dotted line) and 1.0 (upper dotted line) and is readily estimated (using the "brushed" points) to be about 0.79.

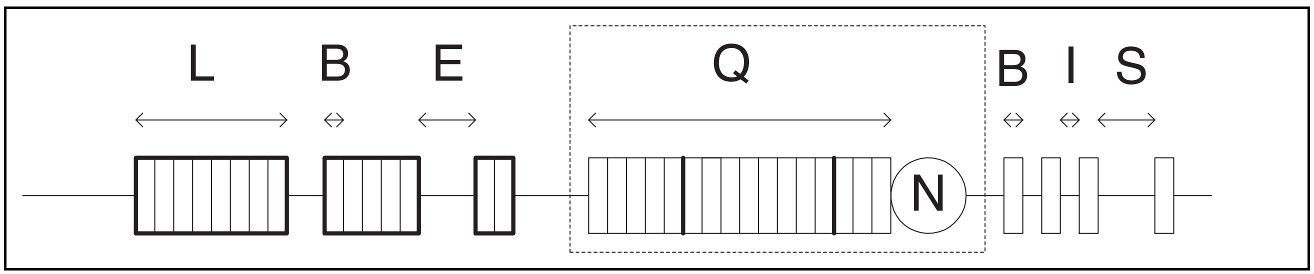
FUD has been how to model the Internet (Park & Willinger 2000)

Fancy Internet models had better include Ethernet

Fast Switched Ethernet

Ethernet 802.3 (bus type) is not unstable in the sense of ALOHA

Nor is switched ethernet 802.3u (100Base-T) & 802.3a,b (1000Base-T)

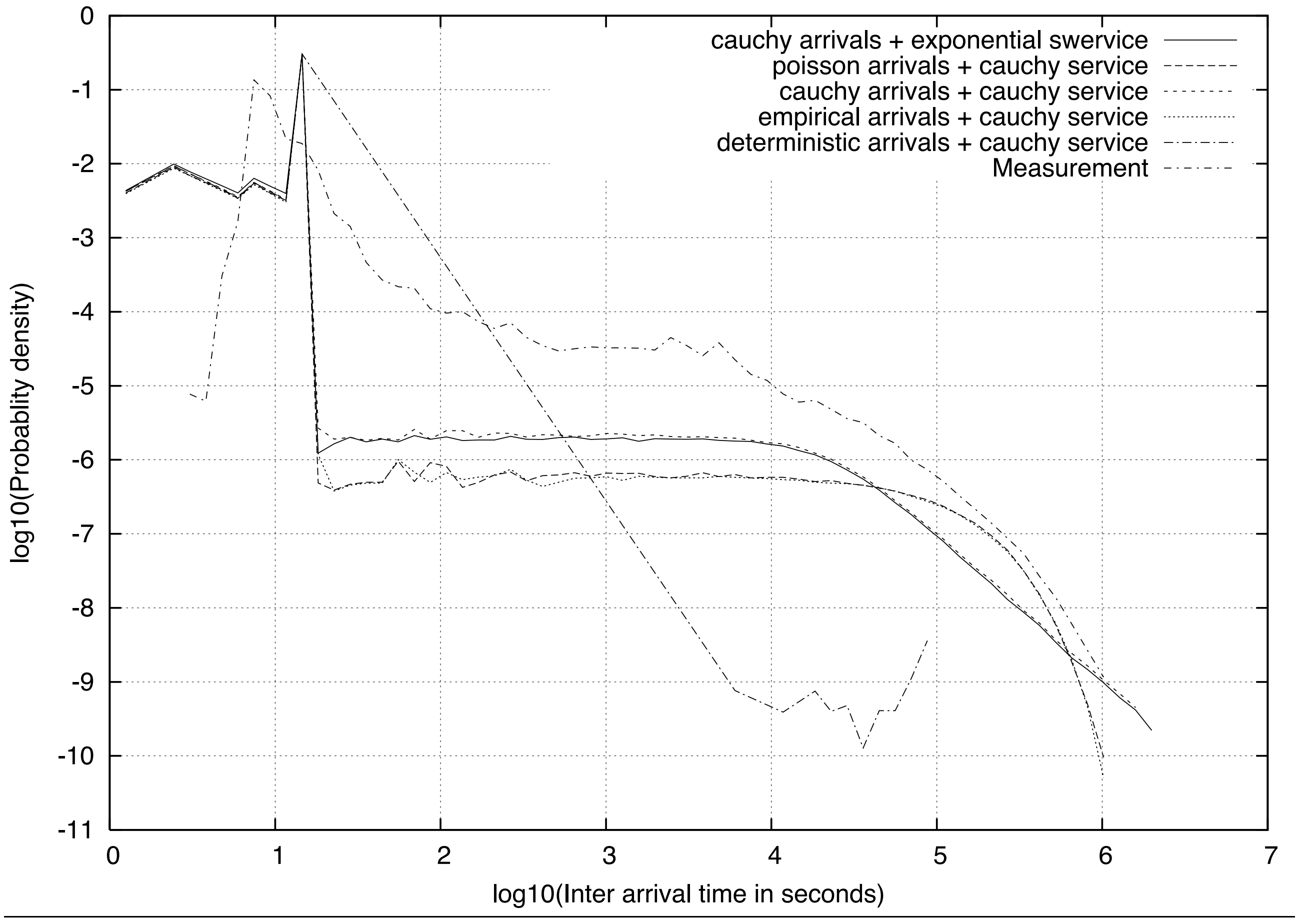


(Harrison et al. 2004) measure pkts on 1000Base-T network
Trace incoming/outgoing pkts near NIC of web server

1. Arrivals are **not Pareto** distributed
2. **Packetization** process is key (Cauchy-$M/G/1$)
3. LRD time correlations have **1/f power spectrum**
4. Even with **Poisson** arrivals!

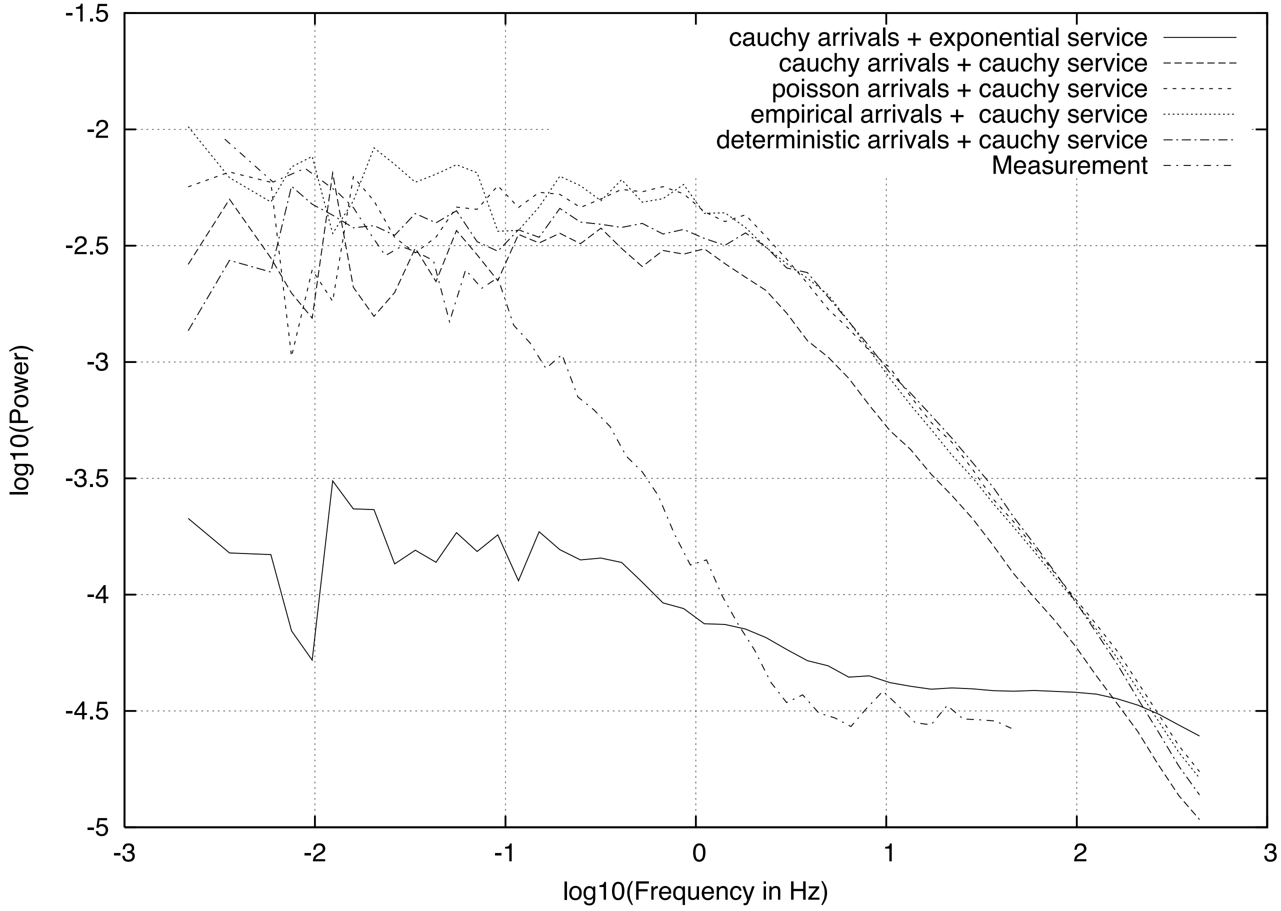
Ethernet Arrivals (Harrison et al. 2004)

Arrivals are **not** power law!



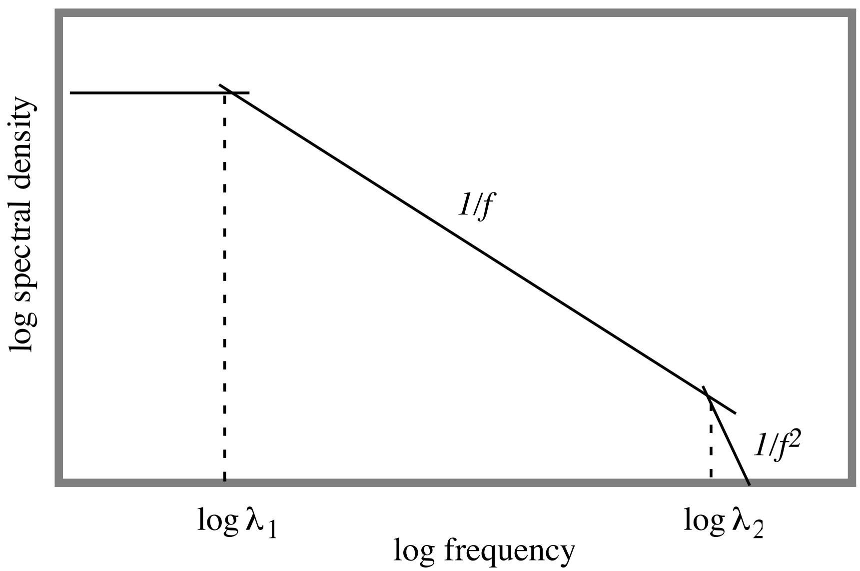
Ethernet Throughput (Harrison et al. 2004)

Departures **are** power law with $\simeq 1/f$ spectral density



Flicker Noise

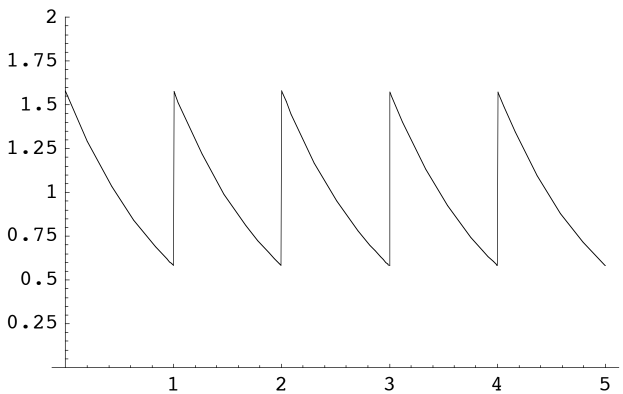
Studied by Schottky in **1918** for thermionic tubes, circuits, etc.



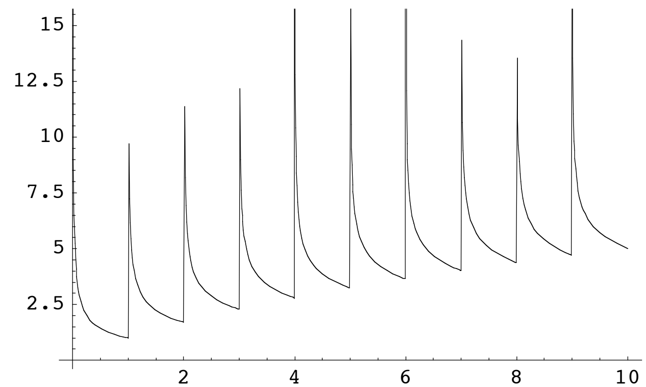
Power spectrum $S(f)$ is Fourier transform of correlation function $C(\tau)$.
If $C(\tau) \sim \tau^{\alpha-1}$ then $S(f) \sim 1/f^{\alpha}$ where
$\alpha \to 0$: white noise
$\alpha \to 1$: flicker noise
$\alpha \to 2$: Brownian process

- Still no single (universal) model of 1/f noise (yet)
- Validity of "sandpile model" (Bak et al. 1987) now questioned

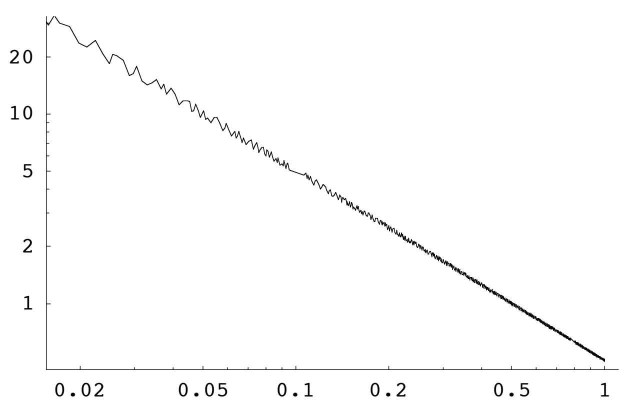
Power Law Impulse Filter



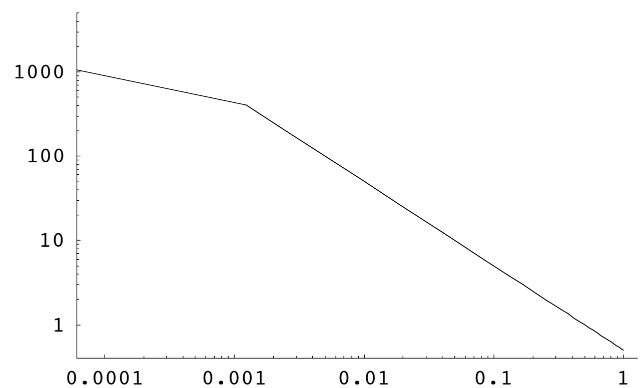
Simple sawtooth waveform



Impulse function $h(t) = 1/t^{\beta}$

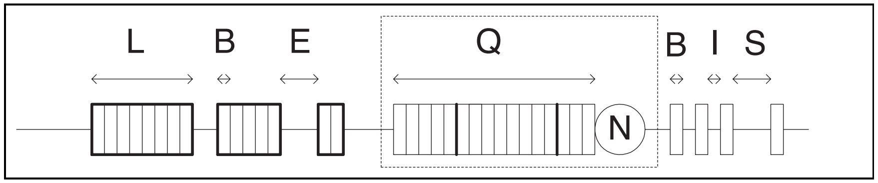


$S(f) = 1/f^{\alpha}; \alpha = 2(1 - \beta)$



$S(f)$ output with cutoff

Power Law Queue



- Poisson arrivals with rate λ
- Service distribution $\sim t^{-\frac{1}{2}}$ (file size-dept.)
  ((Harrison et al. 2004) use Cauchy size dsn. $\sim 1/x^2$)
- Output (network link) distribution $\sim 1/f^{\alpha}$ with ($\alpha = 1$)
- Seems to account for ethernet (Needs more investigation)
- That still leaves the Internet ...

Renormalization Group Therapy

(for Recovering Physicists)

- RGT is mathematical group with **no inverse**

  Rescaling $\Rightarrow$ averaging or filtering. Lose information

  Physics from group **invariants** e.g., rotational symmetry group $\rightarrow$ conservation of angular momentum

- RGT developed by physicists to study connection between **local** and **global** system dynamics. (cf. ALOHA)

  Connection occurs when system **correlations** go "critical"

  (Physics-speak for *fluctuations occur on all scales—length, time)*

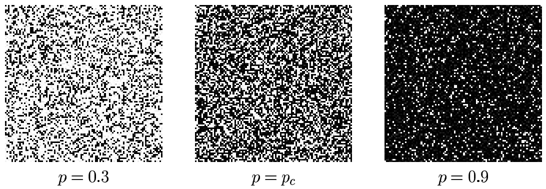
  System characterization goes **singular** at critical point

  Singular behavior characterized by a **power law** (Stanley 1999)

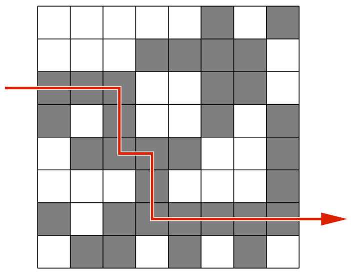
- RGT produces power law solutions with the added capability of characterizing the singularity at **fixed point** of RGT (sometimes)

Site Percolation Models

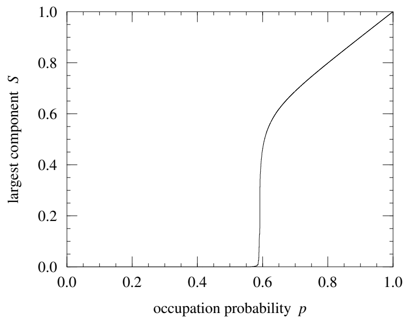
Purely geometric arguments (Almost physics-free!) ☺



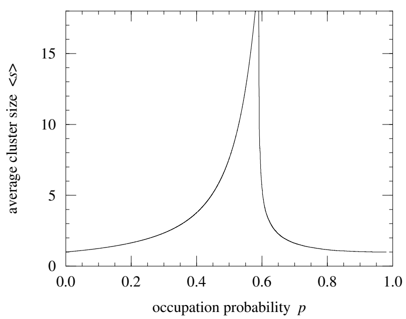
Critical concentration p_c?



First conducting (porous) path

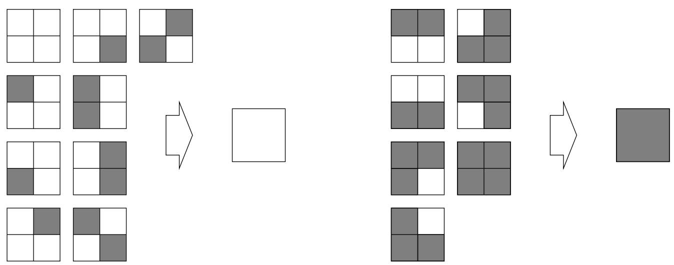


Order parameter $|p - p_c|$

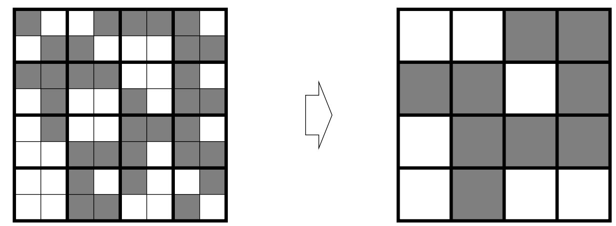


Correlation length $\xi \sim \frac{1}{|p-p_c|^{\nu}}$

Spatial Renormalization



Local transformation rules



Block re-scaling transformations

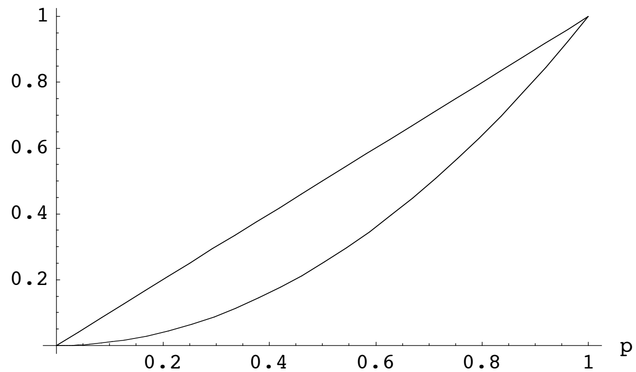
Local fluctuations averaged out under RGT (low-pass filter)

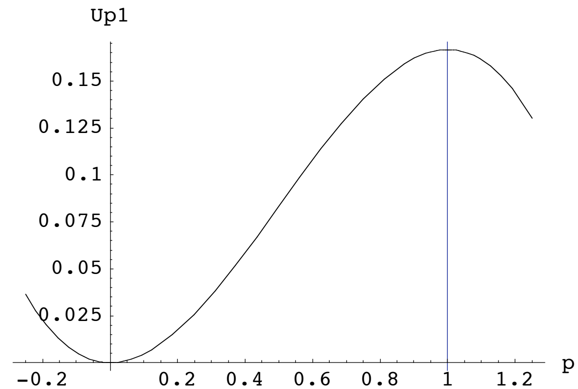
Want fixed points of the re-scaling transformation

Lattice geometry invariant at fixed point p^* $\Rightarrow$ LRD!

Compute power law exponents near p^*

Example 1: Linear Geometry $(d = 1)$





$R(p) = p^2$ (taken pair-wise)

$U(p) = \frac{p^2}{2} - \frac{p^3}{3}$

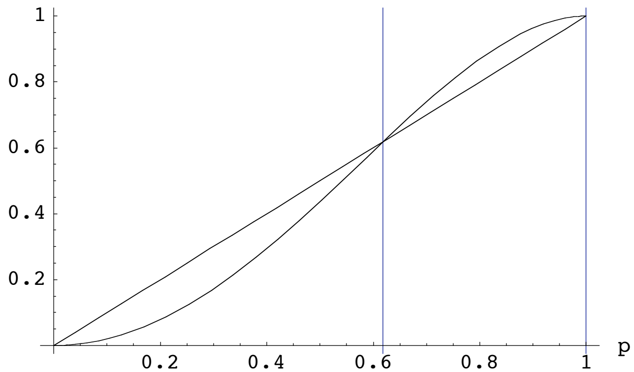
$R(p)$ is the RG transformation (from geometry)

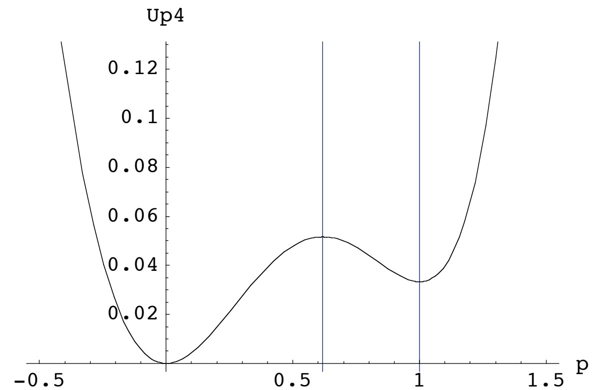
Fixed points p^* occur where $R(p)$ intersects p

In this case, they're trivial: $p^* = 1$ (unstable), $p^* = 0$ (stable)

$U(p)$ is the control function (maxima and minima)

Example 2: $\square$ Lattice $(d = 2)$





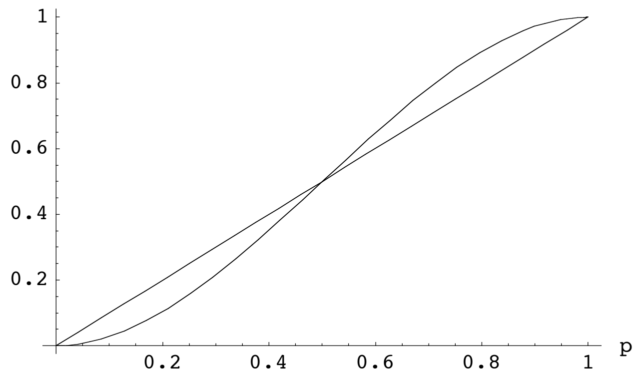
$R(p) = 2p^2 - p^4$

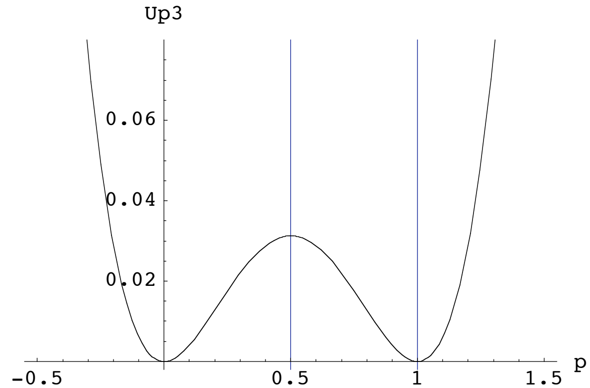
$U(p) = \frac{p^2}{2} - \frac{2p^3}{3} + \frac{p^5}{5}$

Fixed points: 0, 0.618034, and 1 (numerically)

cf. ALOHA stability

Example 3: Δ Lattice $(d = 2)$





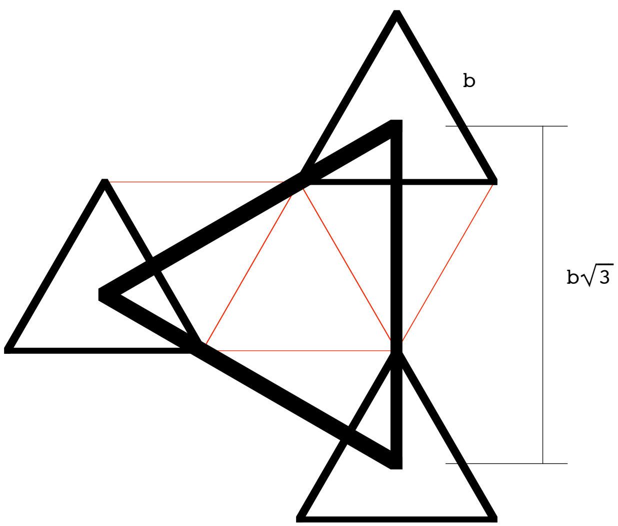
$$R(p) = p^3 - 3p^2(1-p)$$

$$U(p) = \frac{p^2}{2} - p^3 + \frac{p^4}{2}$$

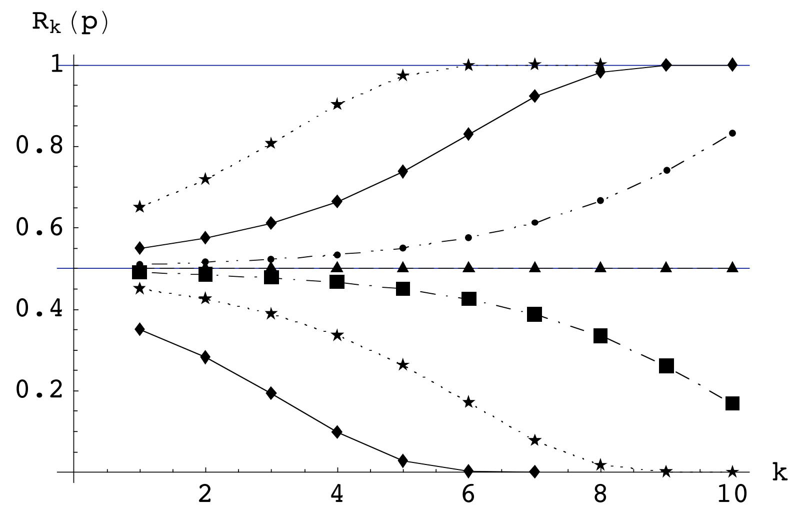
Fixed points p^*: 0, **0.5**, and 1 (algebraically)

Use this ("simple") example to see how power law exponents for correlation length can be computed from RGT

Critical Exponents for Δ Lattice



Geometry of re-scaling to $b/\sqrt{3}$



RGT flows near $p^* = \frac{1}{2} \equiv p_c$

$R(p)$ is *linear* in $(p - p_c)$ near critical pt. Taylor expand:

$$p' = p_c - 6p_c(p_c - 1)(p - p_c) + O(p - p_c)^2$$

where p' is the concentration on the rescaled lattice

Critical Exponent for Correlation Length

2nd term in p' contains Taylor series derivative:

$$\Lambda = \partial_p R(p) = 6(1-p)p\Big|_{p=\frac{1}{2}} = \frac{3}{2}$$

Rescaled correlation length: $\xi' = \xi(p') = \xi(p)/b$

ξ is a **homogeneous function** of $(p - p_c)$:

$$\xi(\Lambda \times |p - p_c|) = \xi(|p - p_c|)/b$$

satisfied by **power law** $\xi = |p - p_c|^{-\nu}$. Hence:

$$\Lambda^{-\nu}\ |p - p_c|^{-\nu} = |p - p_c|^{-\nu}/b$$

and solving for the **exponent** ν with $b = \sqrt{3}$:

$$\nu = \frac{\ln(\sqrt{3})}{\ln(3/2)} = 1.35476$$

Empirical value is $\nu = \frac{4}{3} = 1.3333$

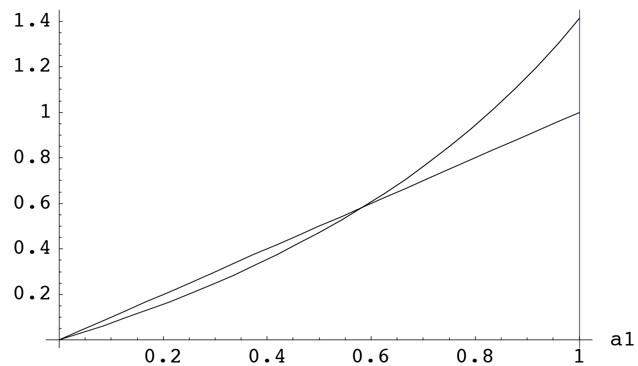
RG Theory of FARIMA Systems

Conventional ARIMA(0, 1, 0) $\rightarrow$ random walk process

ARIMA(0, d, 0) with **fractional difference** $d \in (0, \frac{1}{2}]$

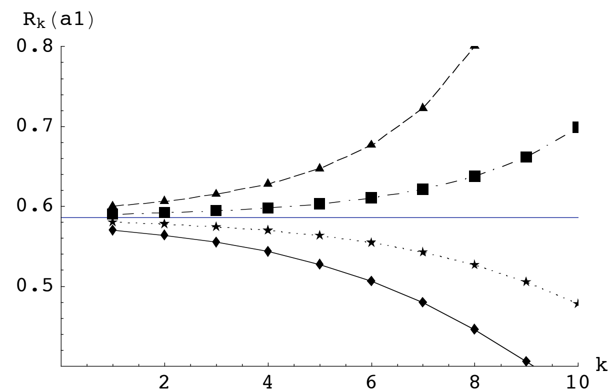
Fractional ARIMA $\rightarrow S(f) \sim f^{-2d}$ (Hosking 1981)

RGT generator



$R(a_1, d) = \frac{4a_1}{2-a_1} 2^{-(d+1)}$

RGT flows



Stable fixed point a_1^*

Fixed point: $a_1^* = 2(2 - 2^{-d})|_{d=\frac{1}{2}} = 0.585786$

RGT and Pareto

(Back to the future)

```
From: Dr. Neil Gunther
Sent: Monday, May 24, 2004 11:08 PM
To: Fischer, Martin J.
Subject: FW: CORS follow up to MC25.2

Dear Dr. Fischer,

I enjoyed your presentation MC25.2 at CORS ...  I may have something to offer
regarding your question about the reduction of variance in the 2-Pareto model.
```

Fischer et al.	**NJG Conjecture**
If X is 1-P distributed r.v., then	$X \to Z$ is an **affine** transformation
$Z = \beta X + k$ is a 3-P distributed r.v.	Affine group > RG
"Stretched" by β	β: **dilatation** transf. (RG **invariant**)
"Shifted" by k	k: is a **translation** (**not** RG invariant)
Stretching does *not* change the CoV	RGT preserves ratio CoV $= \sigma^2/\mu^2$
Given α, can get any μ and σ^2	Not from RGT. Only 1-P satisfies RGT

Conclusions (my unfunded research)

- Has Internet LRD phenomenon been oversold? (Ethernet?)
- RTG $\rightarrow$ physical power laws (from critical dynamics)
- Unstable ALOHA from load-dept. queueing **fixed points**
- Poisson input Ethernet **packetization** $\rightarrow 1/f$ LRD
- RGT for FARIMA models of $1/f$ LRD $\rightarrow$ fixed points
  May explain lack of observable Internet impact (critical dynamics)
  Very localized on Internet (not backbones or WANs)
- Which Pareto is **physical**?
  cf. unphysical multi-parametric scalability models
  1-P ($\bar{F}(x) \sim 1/x^{\alpha}$) is RGT invariant (cf. $S(f) \sim 1/f^{\alpha}$)
- Plenty of cleanup to do ...

References

[Amdahl 1967] G. Amdahl, "Validity of the Single Processor Approach to Achieving Large Scale Computing Capabilities", Proc. AFIPS Conf., 30, 483-485, Apr. 18–20, Atlantic City, New Jersey, 1967

[Bak et al. 1987] P. Bak, C. Tang and K. Wiesenfeld, Phys. Rev. Lett. **59** 381 (1987)

[Downey 2001] A. B. Downey. "Evidence for Long-tailed Distributions in the Internet", ACM SIGCOM Internet Measurement Workshop, Nov. 2001

[Harrison et al. 2004] T. Field, U. Harder and P. Harrison, "Measurement and Modeling of Self-Similar Traffic in Computer Networks", Imperial College London, 2004

[Fischer et al. 2004] M. J. Fischer, D. M. Bevilacqua Masi, D. Gross and J. Shortle "One-Parameter, Two-Parameter, Three-Parameter Pareto: Is There a Modeling Difference?", CORS/INFORMS Joint Conference, Session MC-25, Banff, Canada, 2004

[Ganesh et al. 2004] A. Ganesh, N. O'Connell and D. Wischik, *Big Queues*, Springer LNM #1838, 2004

[Gunther 1990] N. J. Gunther and J. G. Shaw, "Path Integral Evaluation of ALOHA Network Transients", Information Processing Letters, **33(6)**, 289–295, 1990

[Gunther 1993] N. J. Gunther, "A Simple Capacity Model for Massively Parallel Transaction Systems", Proc. CMG Conf., San Diego, California, pp. 1035–1044, 1993

[Gunther 2002] N. J. Gunther, "A New Interpretation of Amdahl's Law and Geometric Scalability", `xxx.lanl.gov/abs/cs.DC/0210017`, Oct 17, 2002

[Gunther 2004] N. J. Gunther, "On the Connection Between Scaling Laws in Parallel Computers and Manufacturing Systems", CORS/INFORMS Joint Conference, Session SC-16, Banff, Canada, 2004

[Gunther 2005] N. J. Gunther, *Analyzing Computer System Performance Using Perl::PDQ*, Springer-Verlag, 2005

[Hosking 1981] J. Hosking, "Fractional Differencing", Biometrika, **68(1)**, 165–176, 1981

[Park & Willinger 2000] K. Park and W. Willinger, *Self-Similar Network Traffic and Performance Evaluation*, Wiley, 2000

[Stanley 1999] H. E. Stanley, "Scaling, Universality, and Renormalization: Three Pillars of Modern Critical Phenomena", Reviews of Modern Physics, **71**, No. 2, S358–S366, 1999