Data driven fractal modeling for blackout and malicious threat detection

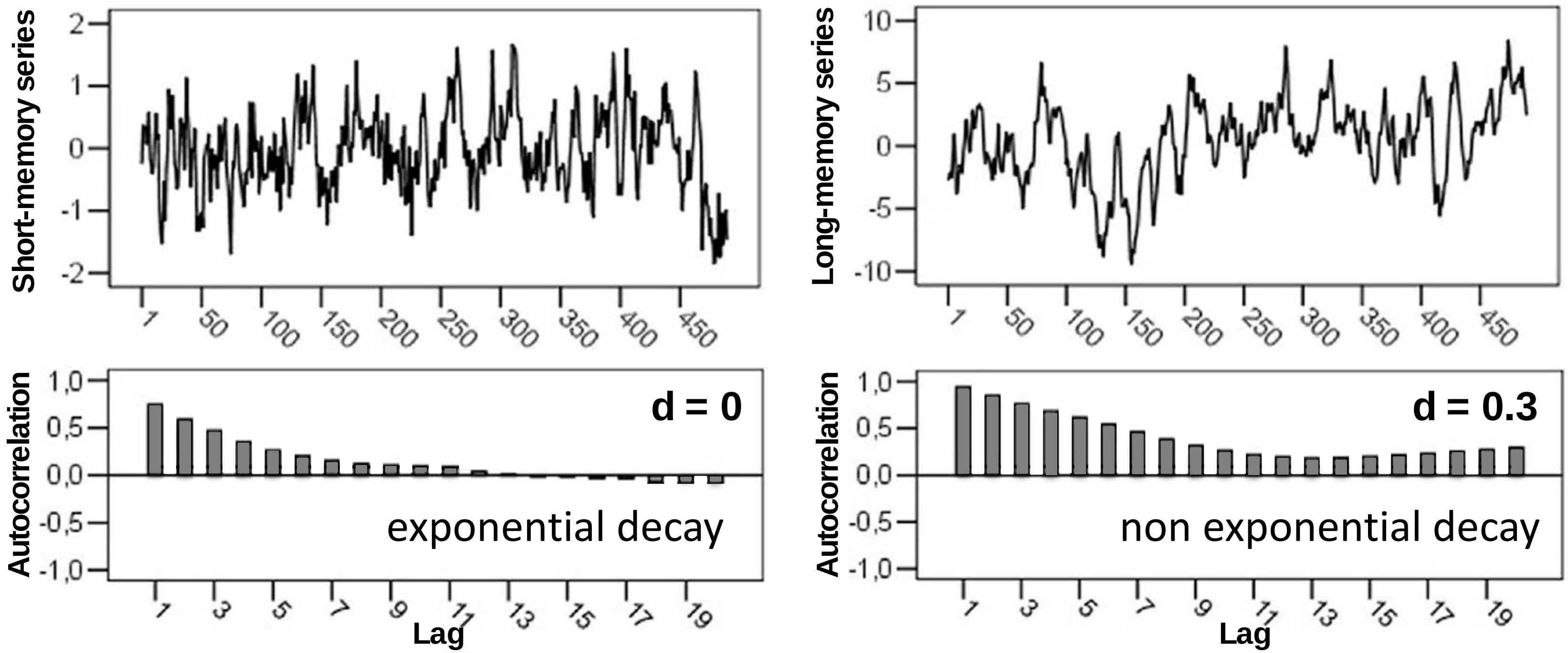
E. Jonckheere
(joint work with L. Shalalfeh and P. Bogdan),
Dept. of Electrical Engineering
University of Southern California
Los Angeles, CA 90089
jonckhee@usc.edu
http://eudoxus2.usc.edu

Plan of Action

- Fractal PMU signal analysis
  - Texas & EPFL (Switzerland) normal PMU data
- Why are the PMU signals fractal???
  - Fractional dynamics load modeling
  - Hidden feedbacks in power grid
  - Strong connectivity of power grid graph, *aggregating all loads*
- Early warning of imminent blackout
  - Indian blackout PMU data
  - Shift in AR(1) coefficient and Hurst exponent.

Long-Range Dependence or Memory (in PMU data)

- **Long-range memory is one of the characteristics of fractal patterns. It relates to slow decay of the correlation as the lag between samples increase.**



Long-Range Dependence or Memory

- **There are several parameters that quantify the severity of the fractal behavior in a time series:**
  - **Number of incrementation or differentiation steps (d):**

$$\text{ARFIMA}: \quad \left(1-\sum_{i=1}^{p}\phi_i L^i\right)(1-L)^d X_t = \left(1+\sum_{i=1}^{q}\theta_i L^i\right)\varepsilon_t, \quad \phi_1 = \text{AR}(1)$$

  - Power Spectral Density exponent (β):

$$S(f) \propto \frac{1}{f^{\beta}}$$

  - Hurst exponent (α):
    It relates to the autocorrelation of time series and the rate at which these decrease as the lag increases.

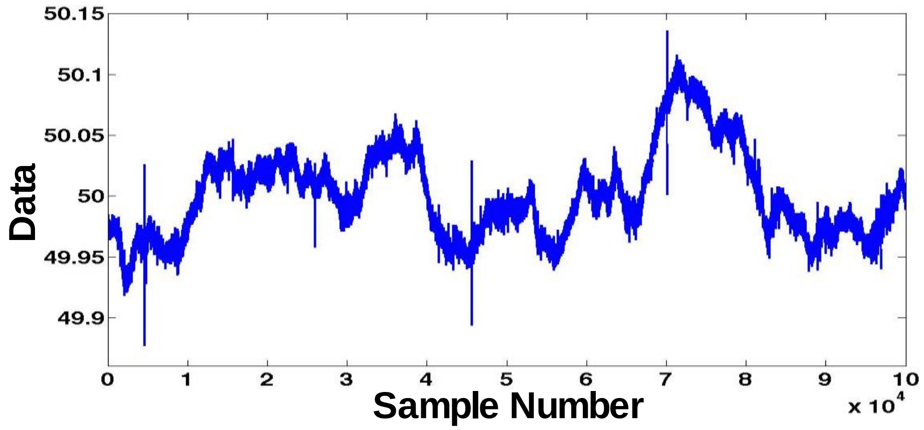
$(2d+1)/2 = \alpha = (\beta+1)/2$

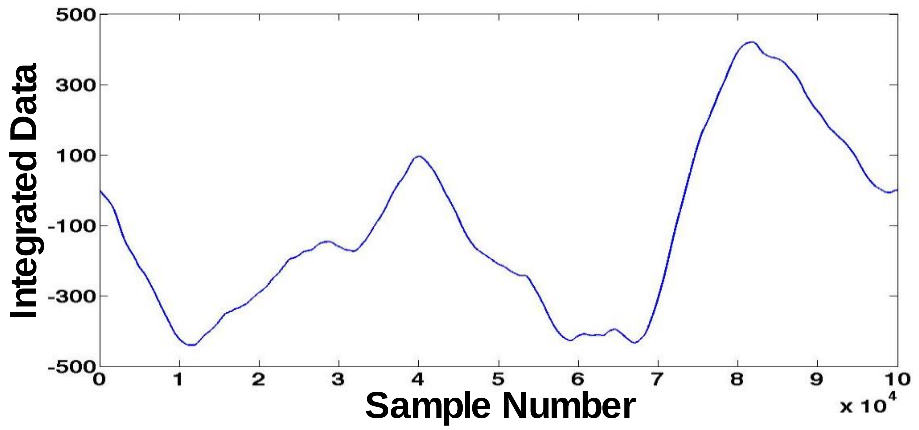
Detrended Fluctuation Analysis (DFA)

- **Steps:**

  1. **Subtract average and integrate the data set:**

$$y_{int}(k) = \sum_{i=1}^{k} (y(i) - y_{avg})$$

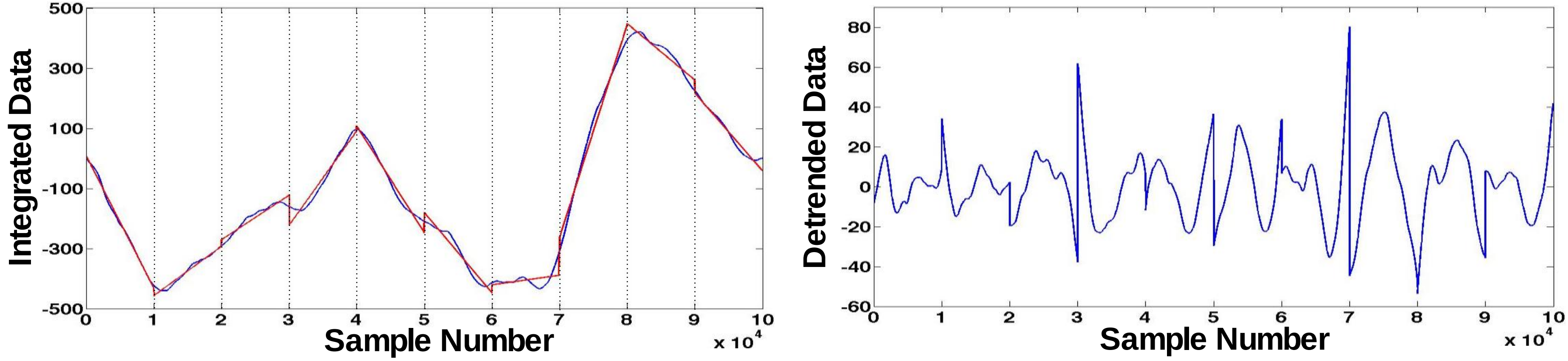




Detrended Fluctuation Analysis (DFA)

2. Divide the data into n equal-sized boxes and find the **Linear Least Squares (LLS) line** inside each box.
3. Subtract the **LLS fitting** from the **integrated data** to generate the **detrended data:**

$$y_{int}(k) - y_n(k) = y_d(k)$$



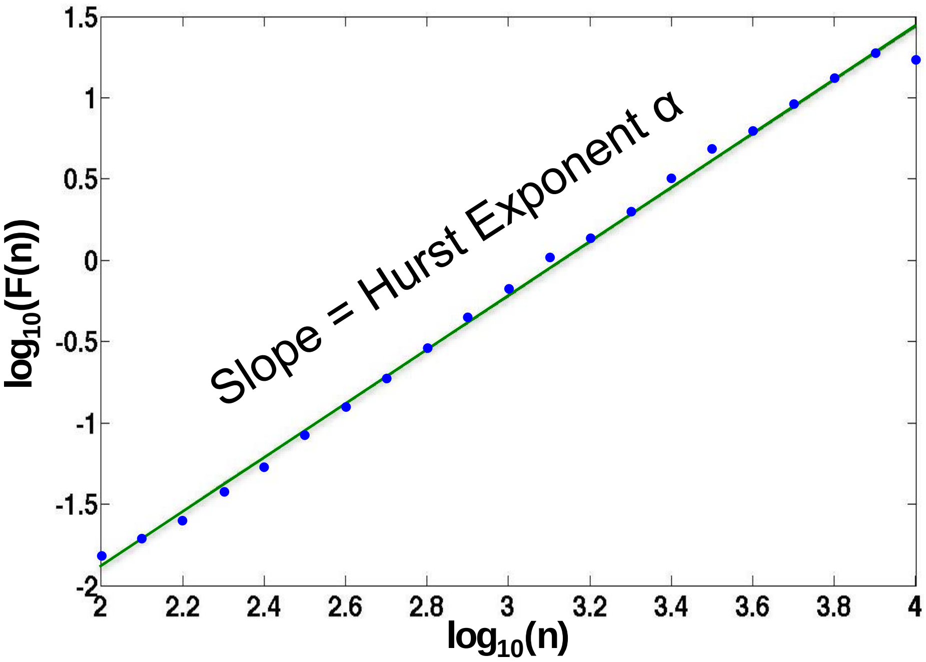
Detrended Fluctuation Analysis (DFA)

4. **Find the Root Mean Square (RMS) fluctuation of the detrended data:**

$$F(n) = \sqrt{\frac{1}{n}\sum_{k=1}^{n}\left(y_d(k)\right)^2}$$

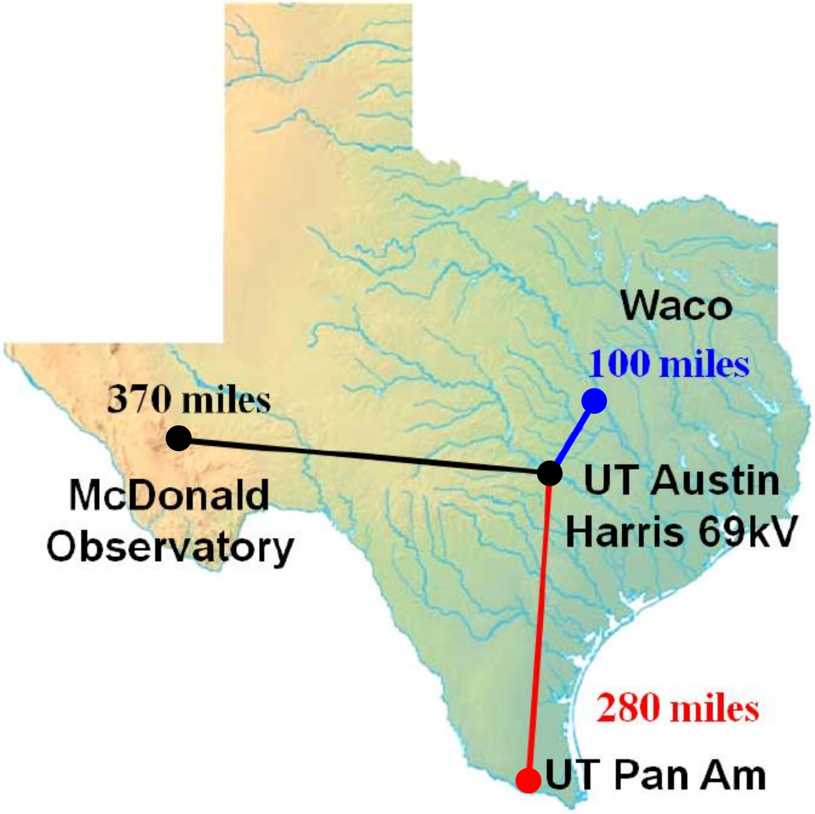
4. **The second and third steps are repeated at different box sizes:**

$$\alpha = \lim_{n\to\infty}\frac{\log_{10}(F(n))}{\log_{10}(n)}$$

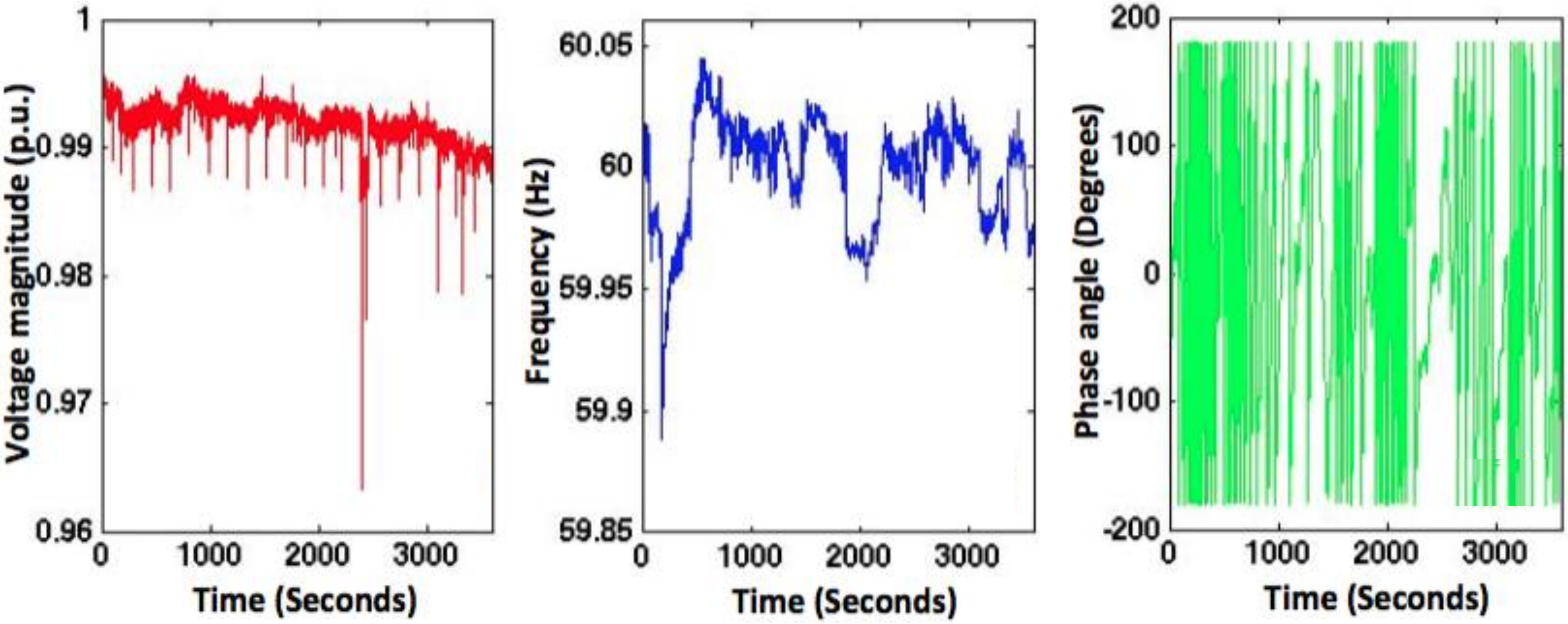


Texas Synchrophasor Network

- Several PMUs are installed at 120V and 69KV over several locations:
  - Baylor University (Waco),
  - Harris Substation, and
  - McDonald Observatory.
- The data we analyzed here are
  - voltage magnitude,
  - frequency, and
  - phase angle.
- The sampling rate of the data is 30 samples/second.



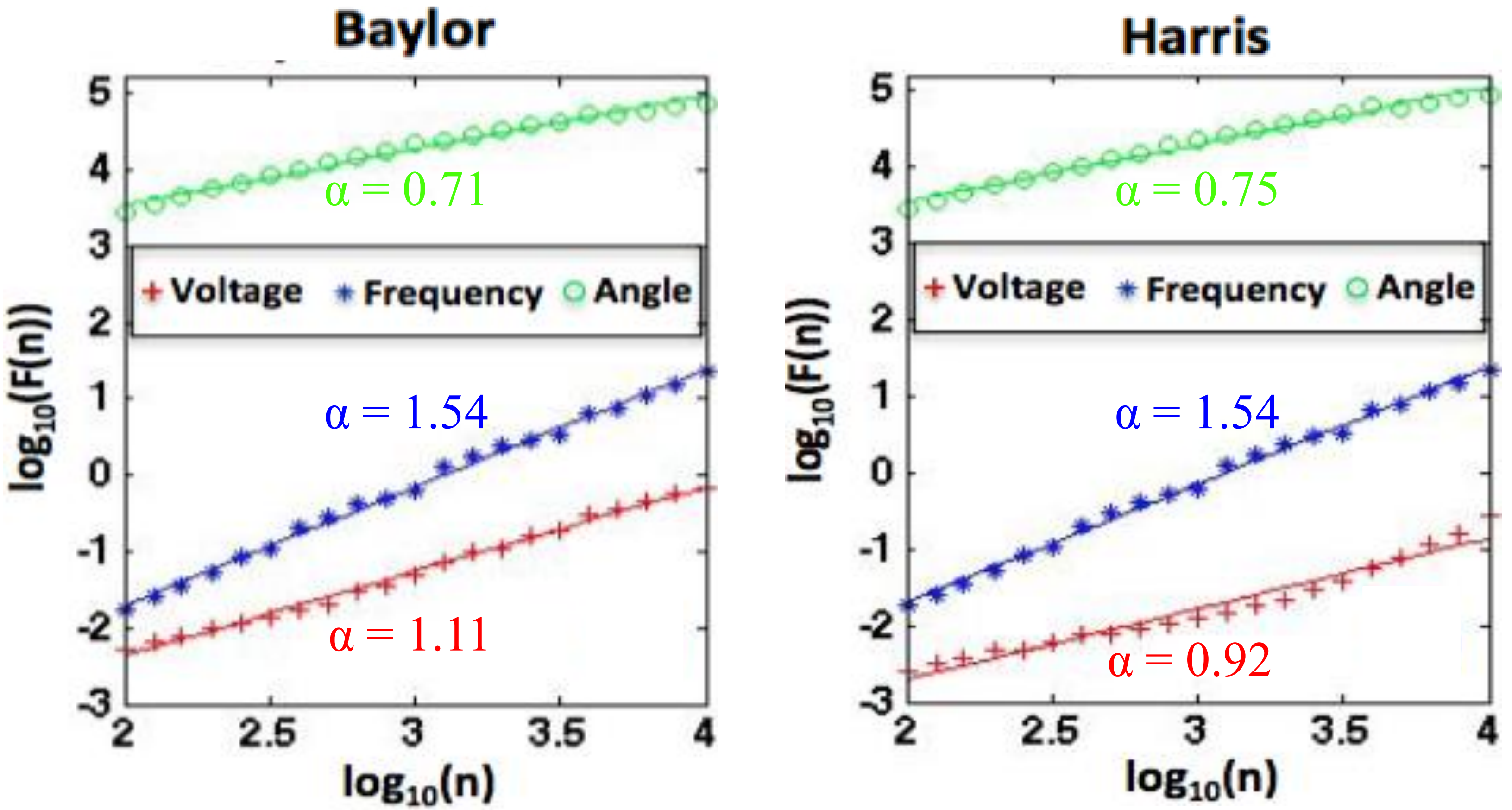
PMU Time Series (Texas)



Hurst Exponent (Texas)

$0.5 \le \alpha \le 1:$ long range with power law

$\alpha > 1:$ long range but no power law



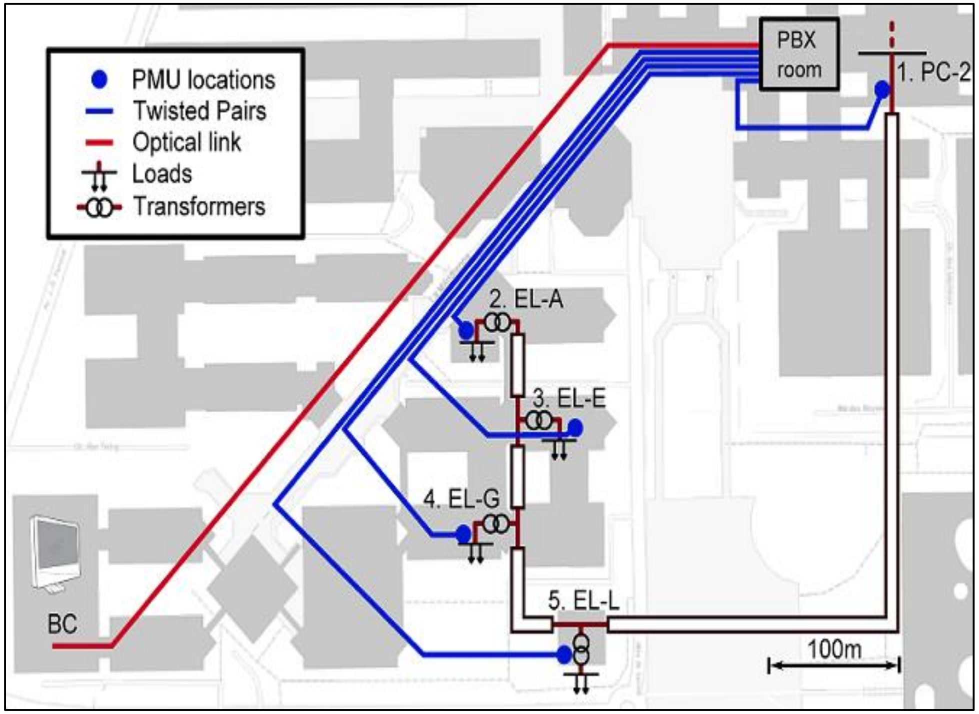
Hurst Exponent (Texas)

Data Set	Baylor			Harris			McDonald		
	V	f	θ	V	f	θ	V	f	θ
#1	1.11	1.54	0.71	0.92	1.54	0.75	1.32	1.54	0.74
#2	1.11	1.53	0.66	0.81	1.53	0.63	1.30	1.53	0.64
#3	1.05	1.45	0.67	0.91	1.45	0.76	1.37	1.45	0.73
#4	0.91	1.49	0.63	0.89	1.49	0.64	1.32	1.49	0.64

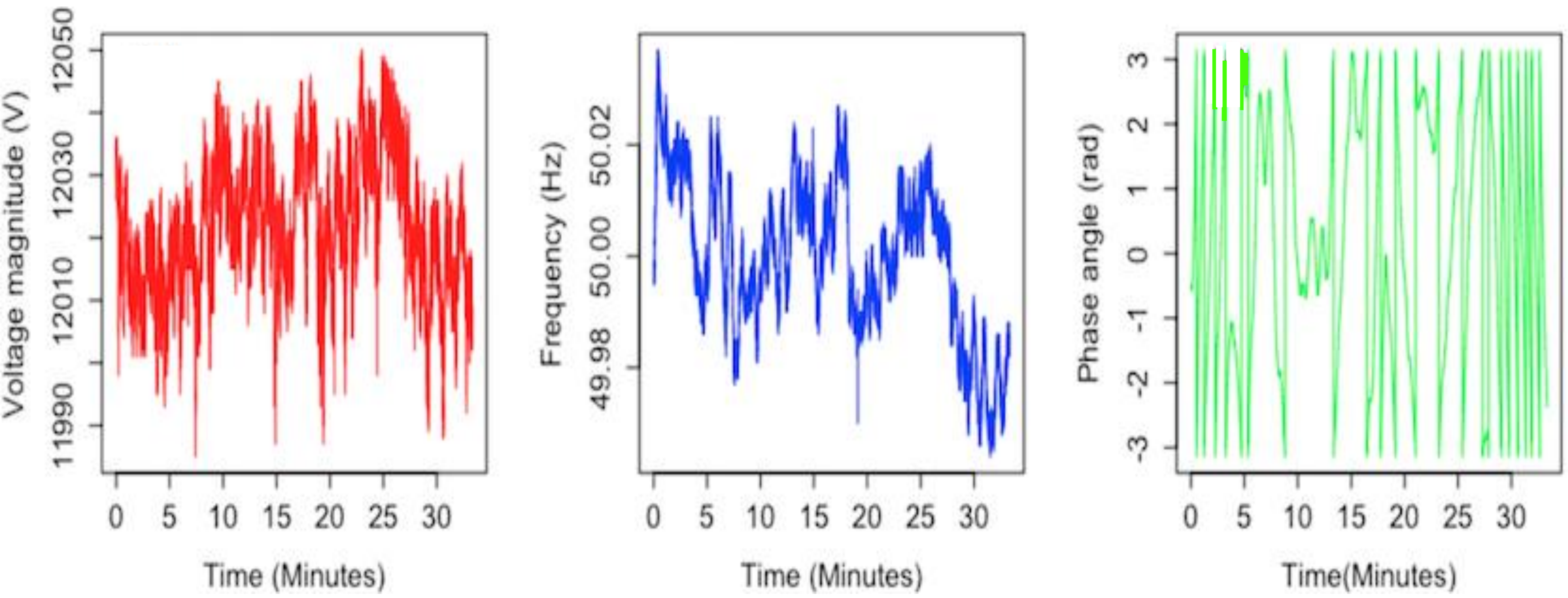
- Frequency and angle data are consistent across the 3 stations.
- Voltage definitely has higher Hurst exponent at McDonald… Why???
  - Proximity of wind farm?
  - Is the Hurst exponent of voltage a sign of *penetration of renewables* in the larger grid?

PMU-Based Monitoring in EPFL
(Ecole Polytechnique Fédérale de Lausanne)

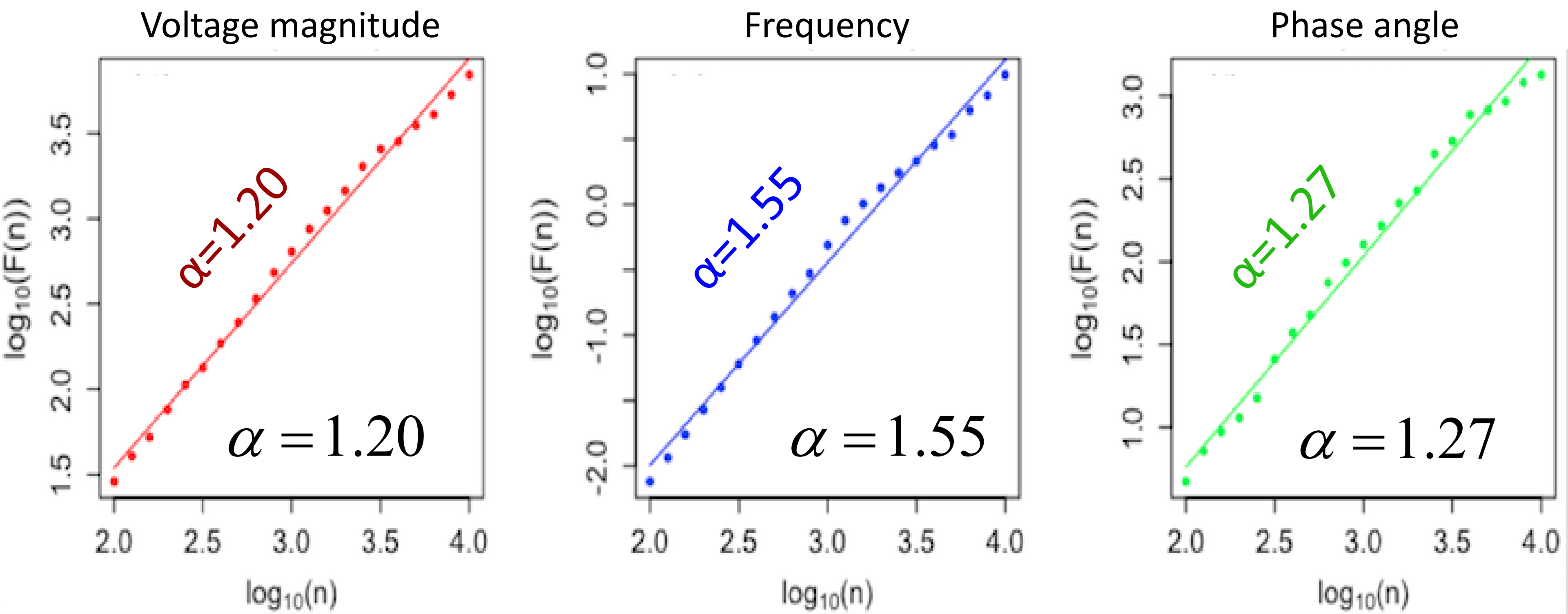
- PMUs installed in EPFL campus perform real time monitoring of the EPFL pilot smart grid.
- The PMUs were installed on medium voltage buses (12KV)
- The sampling rate is 50 samples/second



PMU Time Series (EPFL)

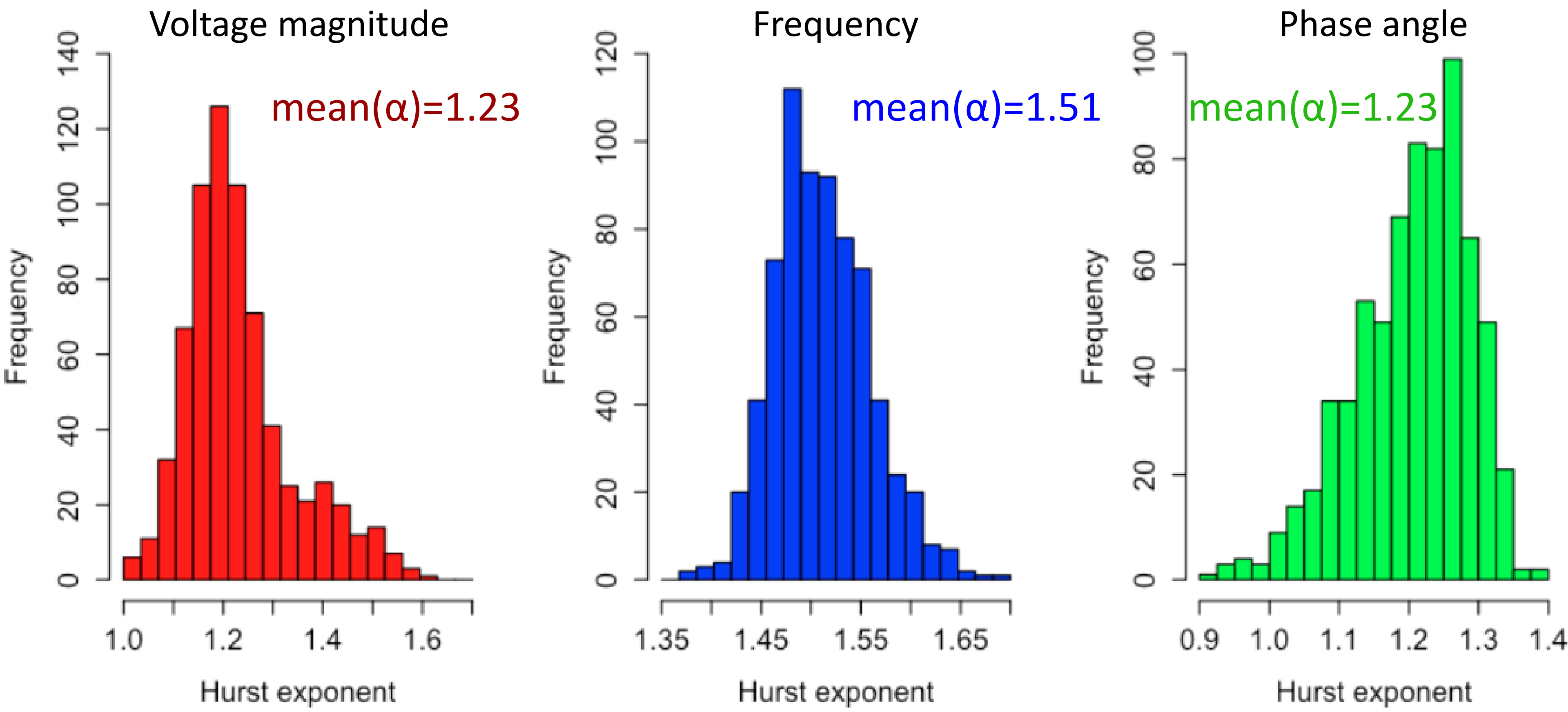


Hurst Exponents (EPFL)



Amazing consistency between the frequency α in Texas (1.54) and Switzerland (1.55)

Hurst Exponent Histograms (EPFL)



Plan of Action

- Fractal PMU signal analysis
  - Texas & EPFL (Switzerland) normal PMU data
- Why are the PMU signals fractal???
  - Fractional dynamics load modeling
  - Hidden feedbacks in power grid
  - Strong connectivity of power grid graph, *aggregating all loads*
- Early warning of imminent blackout
  - Indian blackout PMU data
  - Shift in AR(1) coefficient and Hurst exponent.

Static versus Dynamic Load Models

- Static load model:

$$P_L = K_p V_L^{p_v} \qquad Q_L = K_q V_L^{q_v}$$

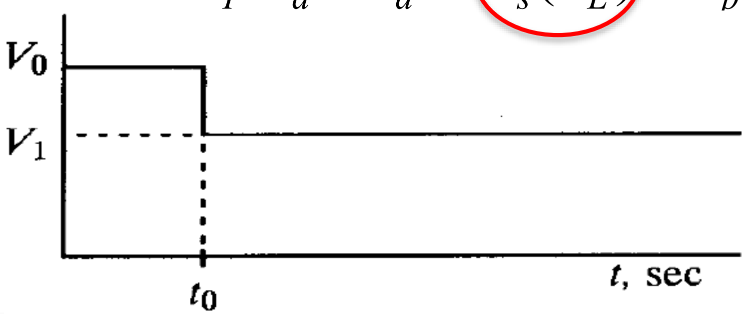
  - Constant Power ➡ $p_v = q_v = 0$
  - Constant Current ➡ $p_v = q_v = 1$
  - Constant Impedance ➡ $p_v = q_v = 2$

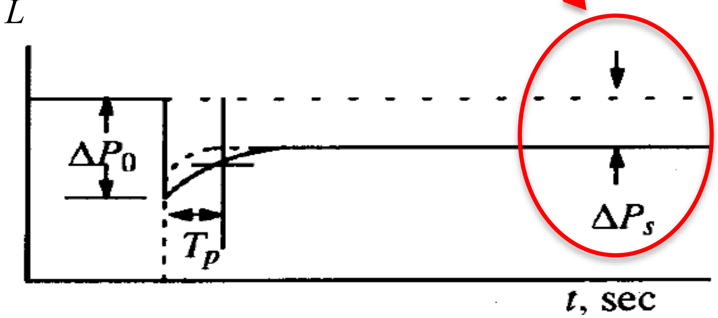
- Dynamic load model (Hill):

$$T_P \dot{P}_d + P_d = P_s(V_L) + k_p(V_L)\dot{V}_L$$

Should be

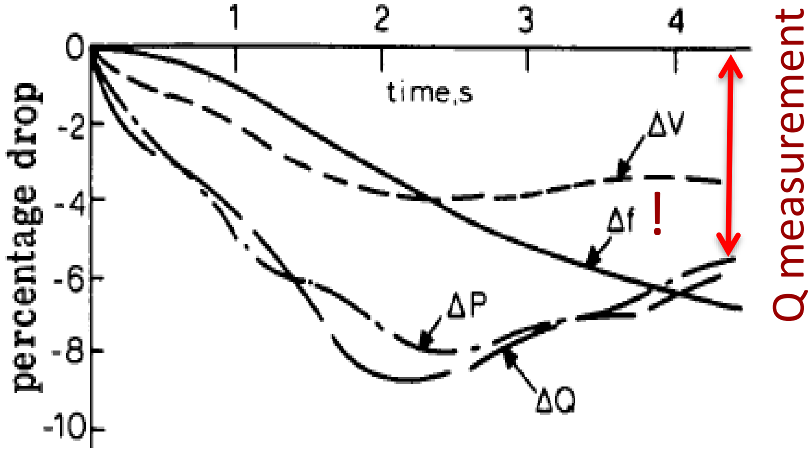
$P_L(V_L, \omega)$



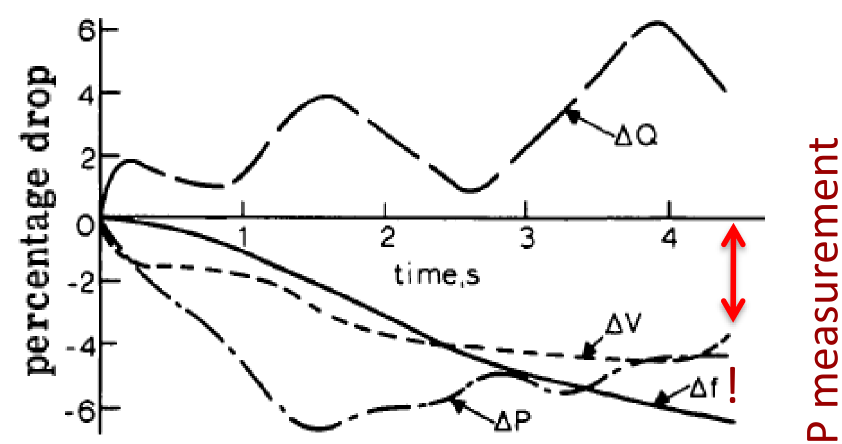


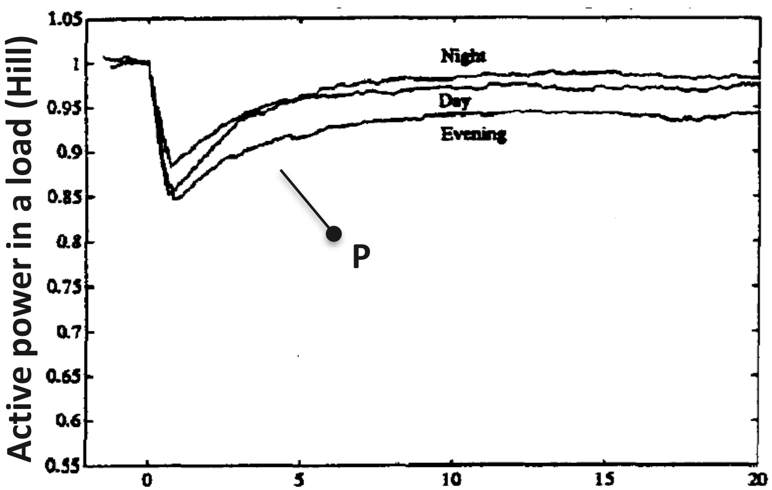
Berg Data-Driven Load Modeling Experiment
in a real microgrid

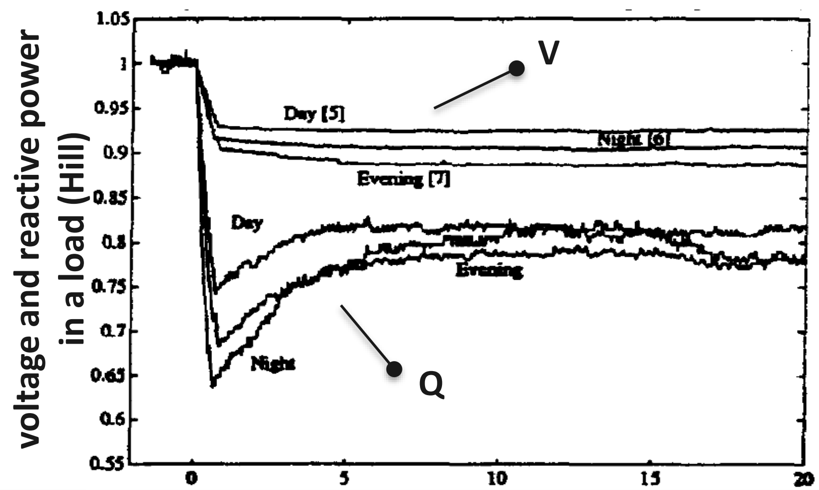
Electric Reduction Furnace (Berg)



Mechanical Plant (Berg)







Berg load model involves frequency to a noninteger exponent

$$\vec{S}_L = P_L + jQ_L \qquad P_L = K_P V_L^{p_v} \omega^{p_\omega} \qquad Q_L = K_Q V_L^{q_v} \omega^{q_\omega}$$

Load Type	p_v	p_ω	q_v	q_ω
Filament lamp	1.6	0	0	0
Fluorescent lamp	1.2	-1.0	3.0	-2.8
Heater	2.0	0	0	0
Induction motor (HL)	0.2	1.5	1.6	-0.3
Induction motor (FL)	0.1	2.8	0.6	1.8
Reduction furnace	1.9	-0.5	2.1	0
Aluminum plant	1.8	-0.3	2.2	0.6
Regulated aluminum plant	2.4	0.4	1.6	0.7

Impedance Describing Function

$$\vec{Z}_L = \frac{\vec{V}_L}{\vec{I}_L} = \frac{\vec{V}_L \vec{V}_L^*}{\vec{I}_L \vec{V}_L^*} = \frac{V_L^2}{\vec{S}_L^*} = \frac{V_L^2}{P_L - jQ_L} = \frac{1}{K_p V_L^{p_v - 2} \omega^{p_\omega} - jK_q V_L^{q_v - 2} \omega^{q_\omega}}$$

Load Type	Describing Function
Filament lamp	$(K_p V_L^{-0.4} - jK_q V_L^{-2})^{-1}$
Fluorescent lamp	$(K_p V_L^{-0.8} \omega^{-1} - jK_q V_L \omega^{-2.8})^{-1}$
Heater	$(K_p - jK_q V_L^{-2})^{-1}$
Induction motor (HL)	$(K_p V_L^{-1.8} \omega^{1.5} - jK_q V_L^{-0.4} \omega^{-0.3})^{-1}$
Induction motor (FL)	$(K_p V_L^{-1.9} \omega^{2.8} - jK_q V_L^{-1.4} \omega^{1.8})^{-1}$
Reduction furnace	$(K_p V_L^{-0.1} \omega^{-0.5} - jK_q V_L^{0.1})^{-1}$
Aluminum plant	$(K_p V_L^{-0.2} \omega^{-0.3} - jK_q V_L^{0.2} \omega^{0.6})^{-1}$
Regulated aluminum plant	$(K_p V_L^{0.4} \omega^{0.4} - jK_q V_L^{-0.4} \omega^{0.7})^{-1}$

Analytic Extension of Describing Function

$$Y_L = \frac{1}{Z_L} = L(V_L)\omega^p + jW(V_L)\omega^q$$

Crude way:

Leaves some coefficients complex, not completely in line with formal circuit theory

$$\omega \to \omega - j\sigma$$

Better way:

Coefficients are kept real, in line with formal circuit theory;

However, positive realness does not hold unless the load is a heater

$$Y_L \approx A(V_L) \times (j\omega)^\alpha + B(V_L) \times (j\omega)^\beta \xrightarrow{\text{extension}} A(V_L)s^\alpha + B(V_L)s^\beta$$

where $A(.)$ and $B(.)$ are real valued.

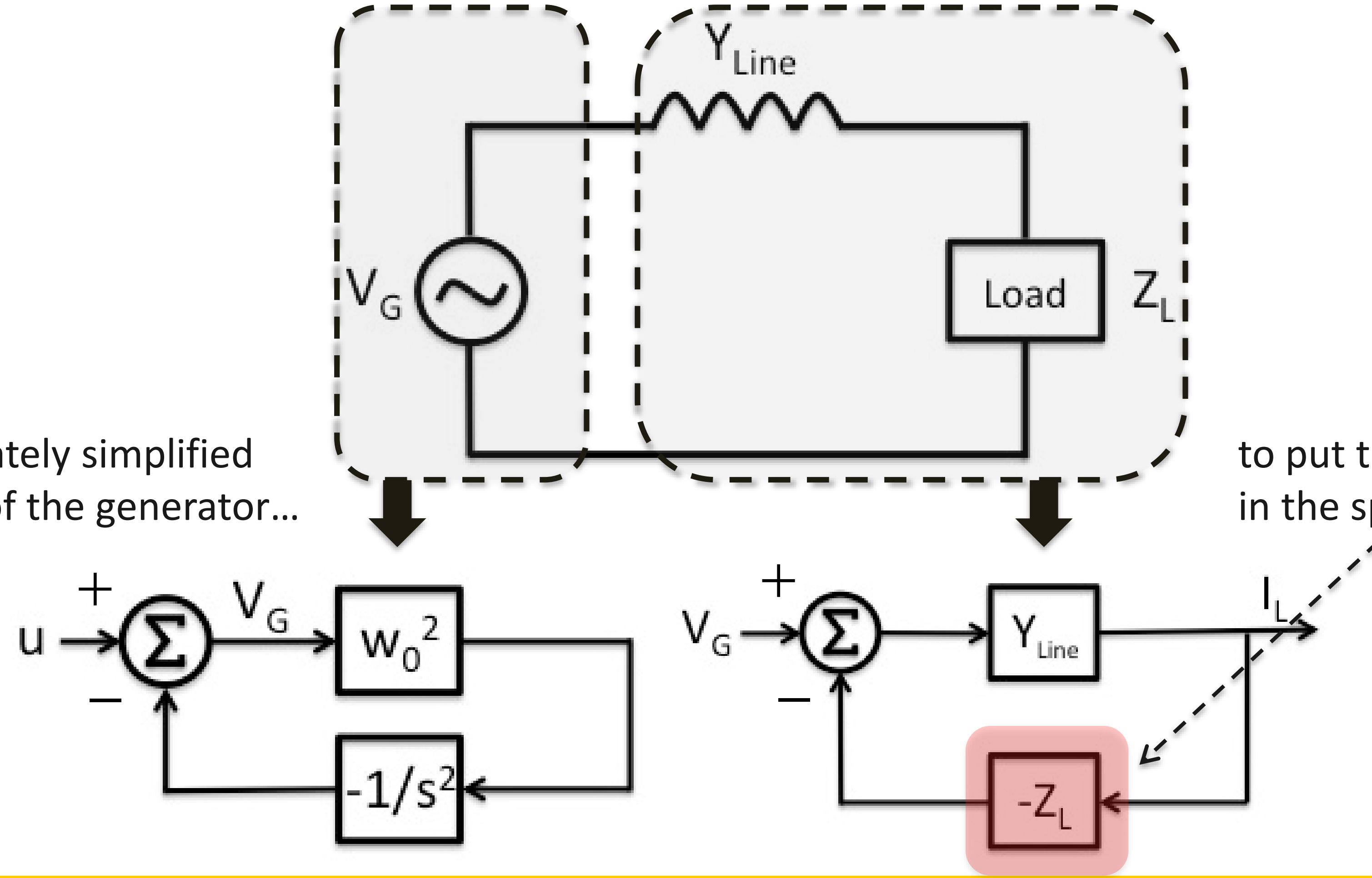
Can we replace s by $\frac{d}{dt}$???

Yes, but subject to correct interpretation:

related {
- Caputo, D_* (initial conditions in terms of integer derivatives)
- Riemann-Liouville, D (initial conditions in terms of fractional derivatives)
- Grunwald-Leitnikov, d (close to ARFIMA model)

$$\underbrace{\begin{pmatrix} a_1 D_{(*)}^{\alpha_1} + b_1 D_{(*)}^{\beta_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & a_n D_{(*)}^{\alpha_n} + b_n D_{(*)}^{\beta_n} \end{pmatrix}}_{\text{Distribution network}} \underbrace{\begin{pmatrix} V_1 \\ \vdots \\ V_n \end{pmatrix}}_{\substack{\text{State} \\ \text{(load voltages)}}} = \underbrace{A(Y_{\text{Line}})}_{\text{Transmission network}} \begin{pmatrix} V_1 \\ \vdots \\ V_n \end{pmatrix} + B(Y_{\text{Line}}) \underbrace{\begin{pmatrix} V_{G_1} \\ \vdots \\ V_{G_m} \end{pmatrix}}_{\text{Generation}}$$

Hidden Feedback in Power Systems



Deliberately simplified model of the generator…

to put the load in the spotlight.

Feedback Model of Power System

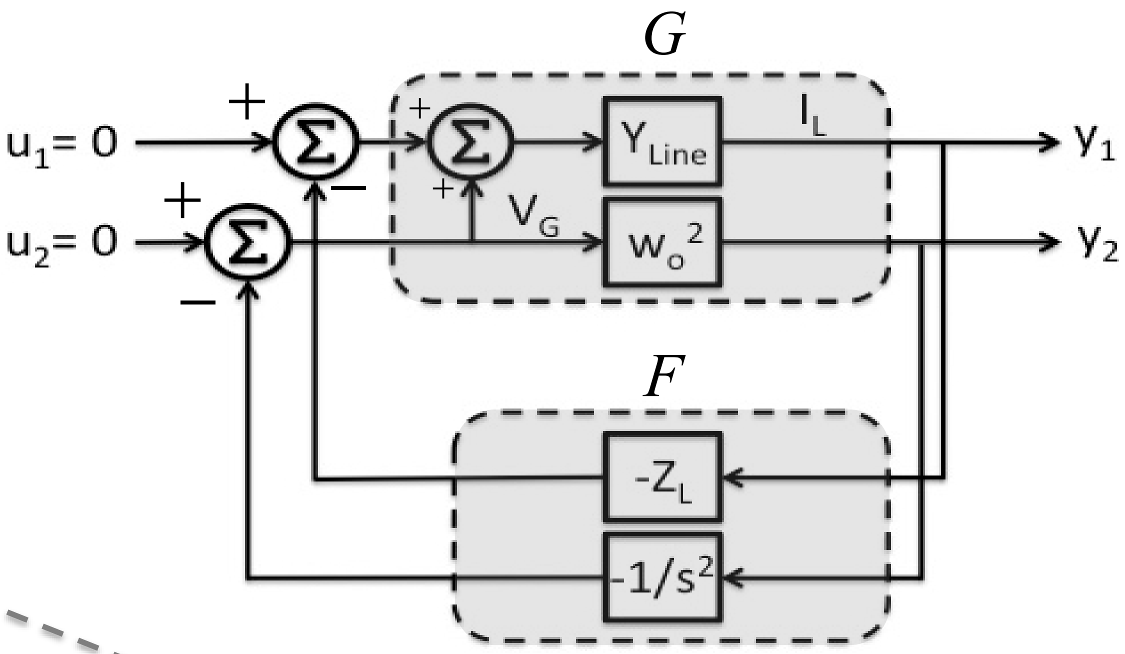
$$y = (I - GF)^{-1} Gu$$

$$u = \begin{bmatrix} u_1 & u_2 \end{bmatrix}^t$$

$$y = \begin{bmatrix} y_1 & y_2 \end{bmatrix}^t$$

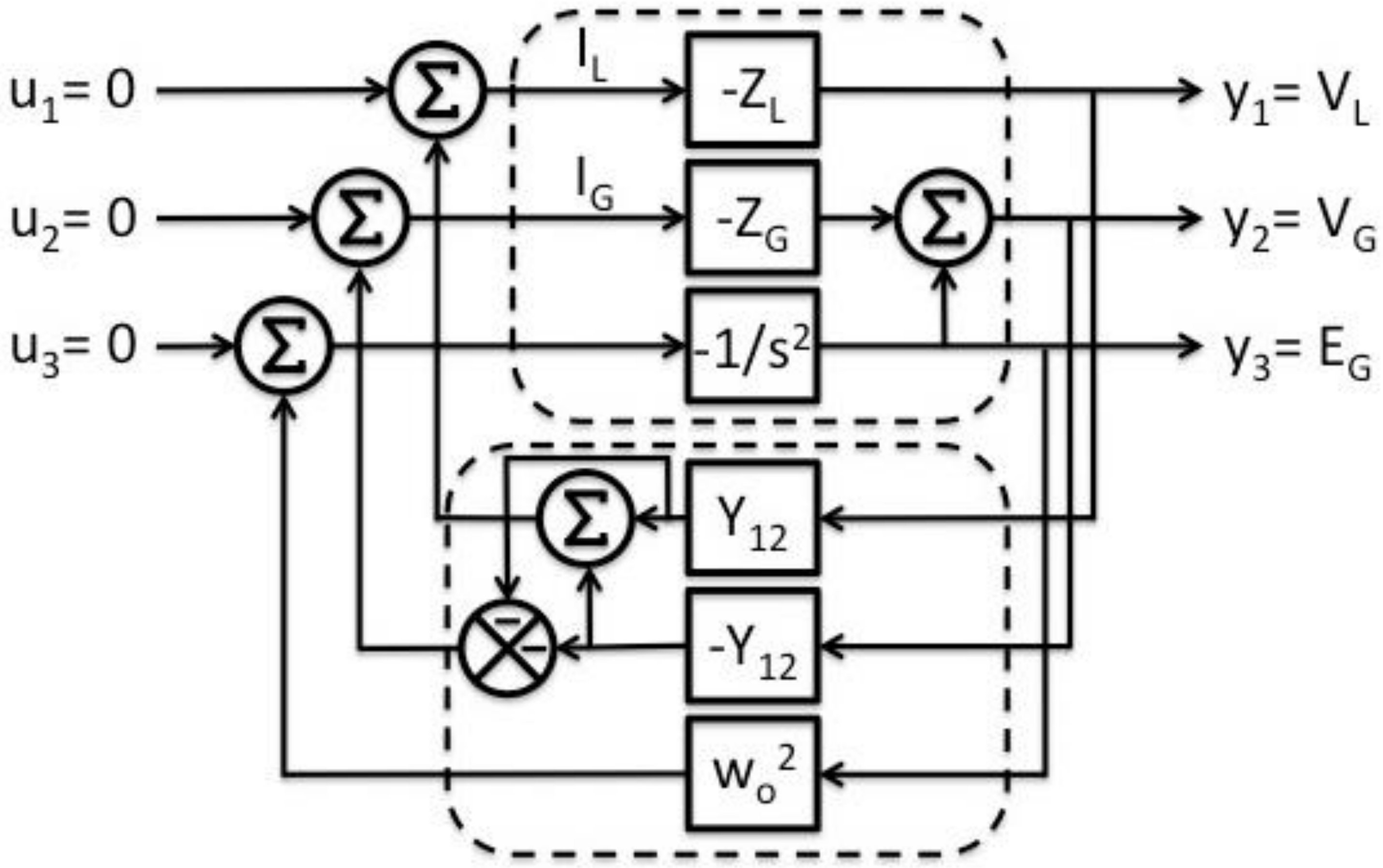
$$G = \begin{bmatrix} Y_{Line} & Y_{Line} \\ 0 & \omega_0^2 \end{bmatrix}$$

$$F = \begin{bmatrix} -Z_L & 0 \\ 0 & -s^{-2} \end{bmatrix}$$

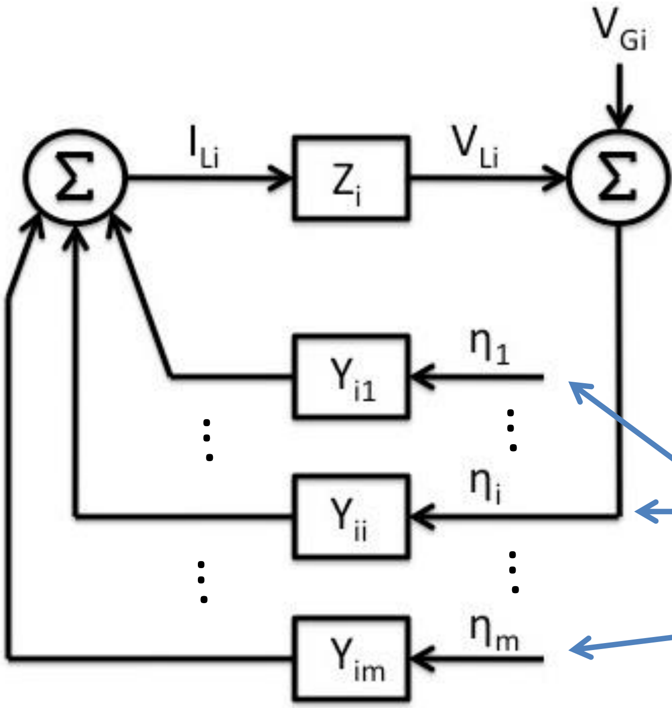


Simplification:
No back-action of the load to the generator

Towards more Complicated Feedback Models of Power System



Time to Conceptualize



714 IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS, VOL. CAS-23, NO. 12, DECEMBER 1976

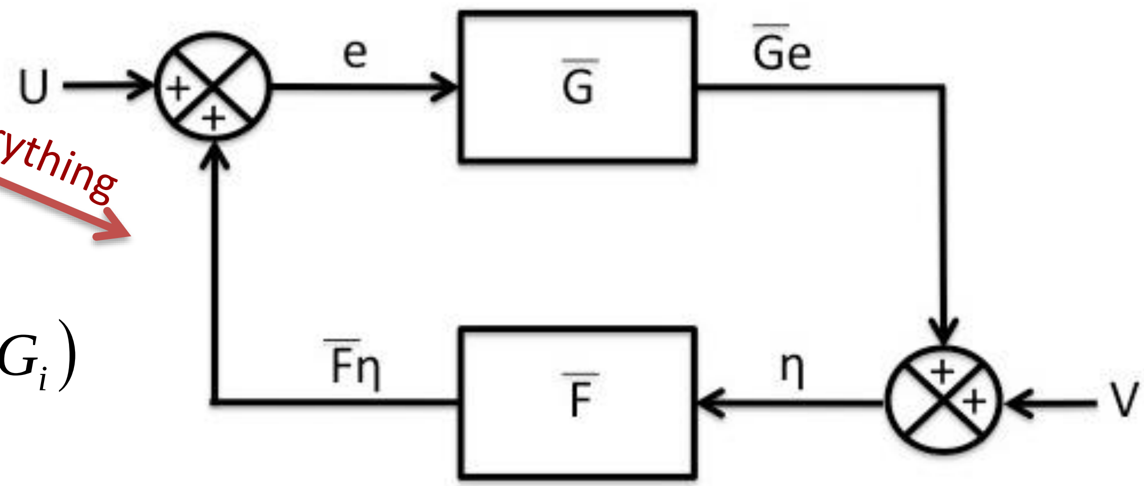
Input–Output Stability Theory of Interconnected Systems Using Decomposition Techniques

FRANK M. CALLIER, MEMBER, IEEE, WAN S. CHAN, STUDENT MEMBER, IEEE, AND CHARLES A. DESOER, FELLOW, IEEE

Nominal impedance, line Z_i, Y_{ii}

Connecting lines 1, 2, …, ≠i, …, m

Lump everything

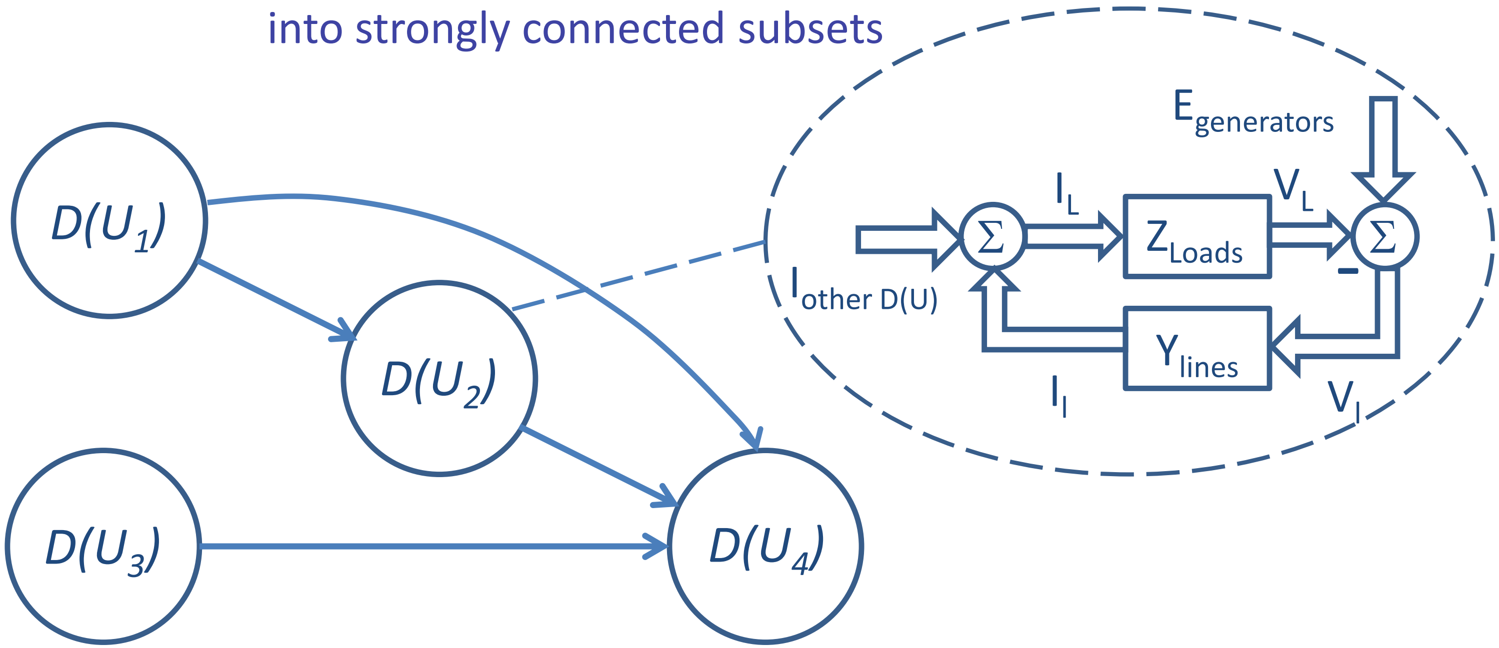


Are we sure that

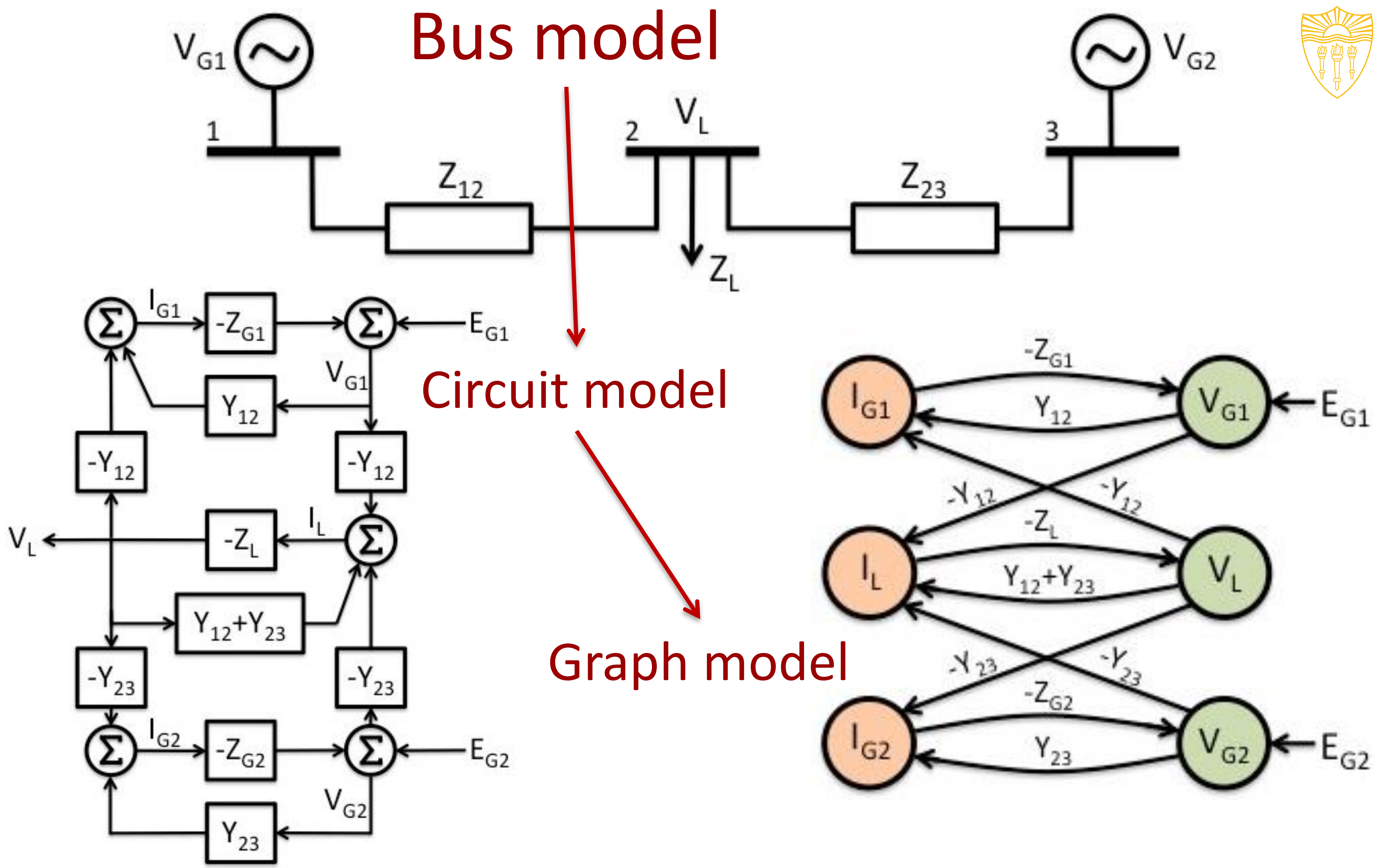
$$\det\left(I - \overline{F}\,\overline{G}\right) \neq \Pi_i \det\left(I - F_i G_i\right)$$

Decomposition of Digraph into Strongly Connected Components *D(U_i)*

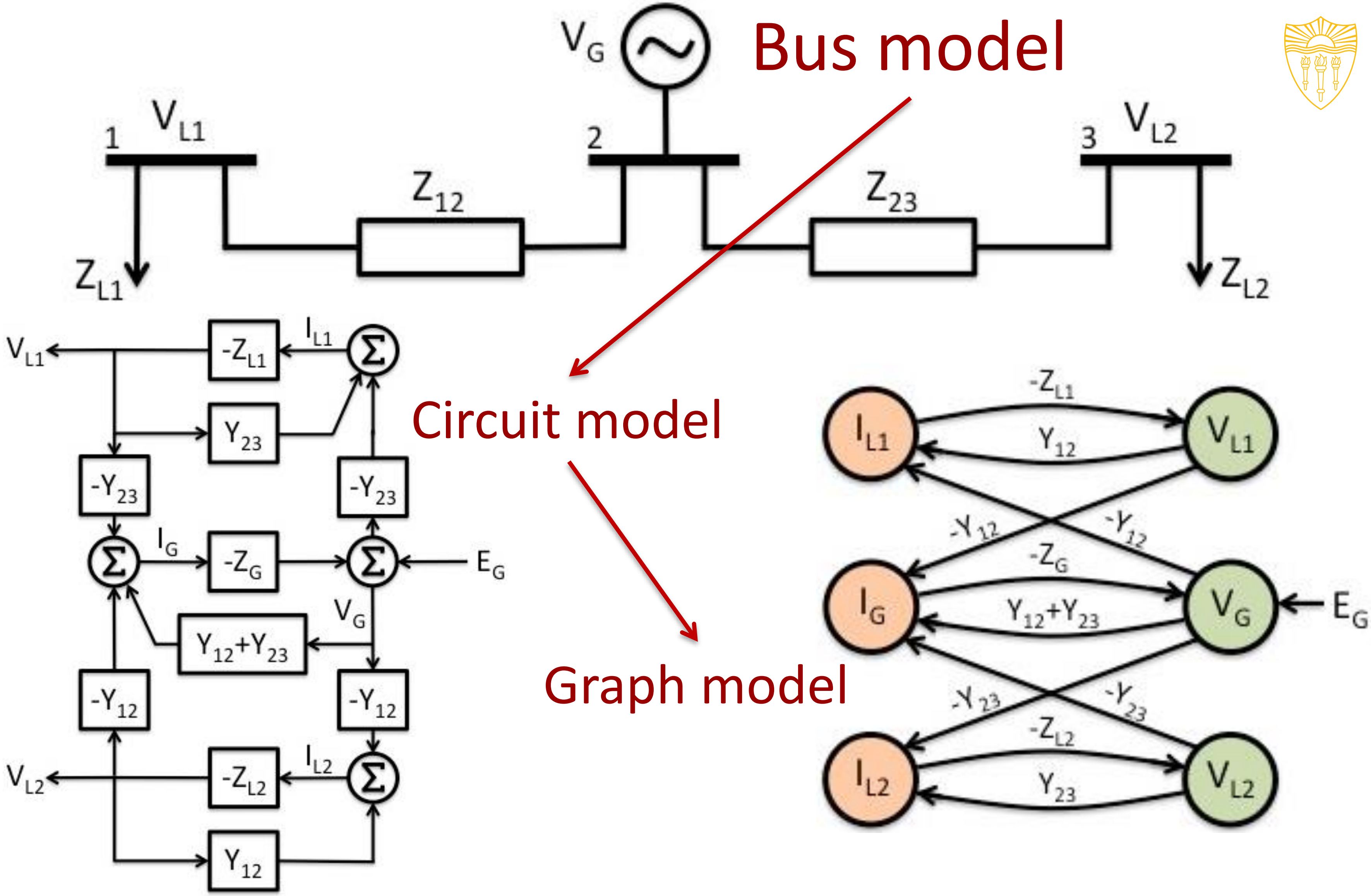
Feedback connections, if any, are lumped into strongly connected subsets



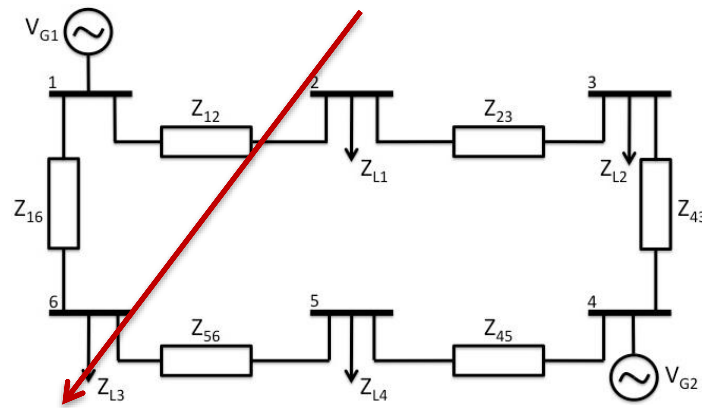
No large scale feedback connections at the large scale of the structure graph



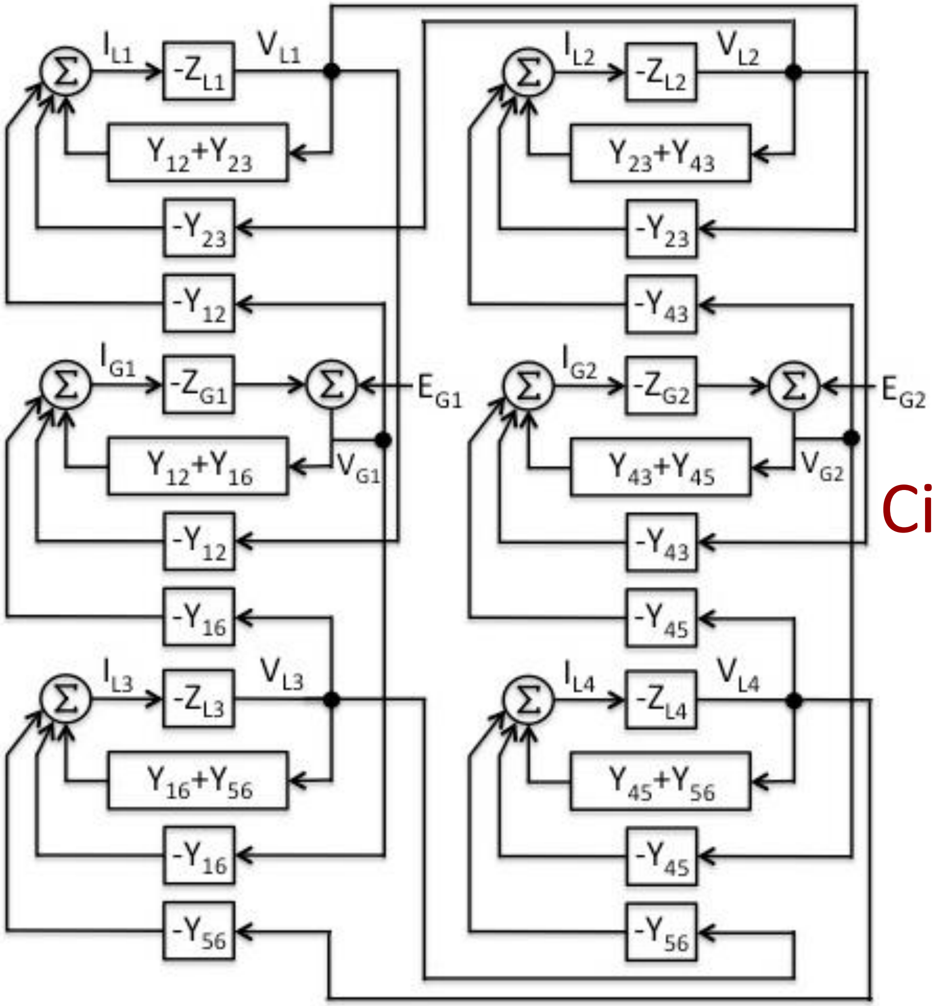
Bus model
Circuit model
Graph model
V_{G1}
V_{G2}
V_L
1
2
3
Z_{12}
Z_{23}
Z_L
I_{G1}
$-Z_{G1}$
E_{G1}
V_{G1}
Y_{12}
$-Y_{12}$
$-Y_{12}$
I_L
$-Z_L$
V_L
$Y_{12}+Y_{23}$
$-Y_{23}$
$-Y_{23}$
I_{G2}
$-Z_{G2}$
E_{G2}
V_{G2}
Y_{23}
I_{G1}
V_{G1}
$-Z_{G1}$
Y_{12}
E_{G1}
$-Y_{12}$
$-Y_{12}$
I_L
V_L
$-Z_L$
$Y_{12}+Y_{23}$
$-Y_{23}$
$-Y_{23}$
I_{G2}
V_{G2}
$-Z_{G2}$
Y_{23}
E_{G2}

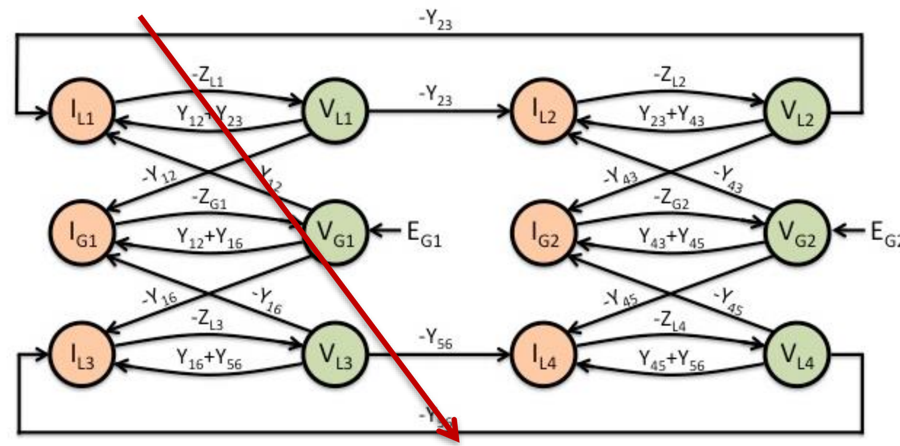

Bus model
V_G
1
V_{L1}
2
3
V_{L2}
Z_{12}
Z_{23}
Z_{L1}
Z_{L2}
Circuit model
V_{L1}
$-Z_{L1}$
I_{L1}
Σ
Y_{23}
$-Y_{23}$
$-Y_{23}$
I_G
$-Z_G$
E_G
V_G
$Y_{12}+Y_{23}$
$-Y_{12}$
$-Y_{12}$
V_{L2}
$-Z_{L2}$
I_{L2}
Y_{12}
Graph model
I_{L1}
V_{L1}
$-Z_{L1}$
Y_{12}
$-Y_{12}$
$-Y_{12}$
I_G
V_G
$-Z_G$
$Y_{12}+Y_{23}$
E_G
$-Y_{23}$
$-Y_{23}$
I_{L2}
V_{L2}
$-Z_{L2}$
Y_{23}

Bus model

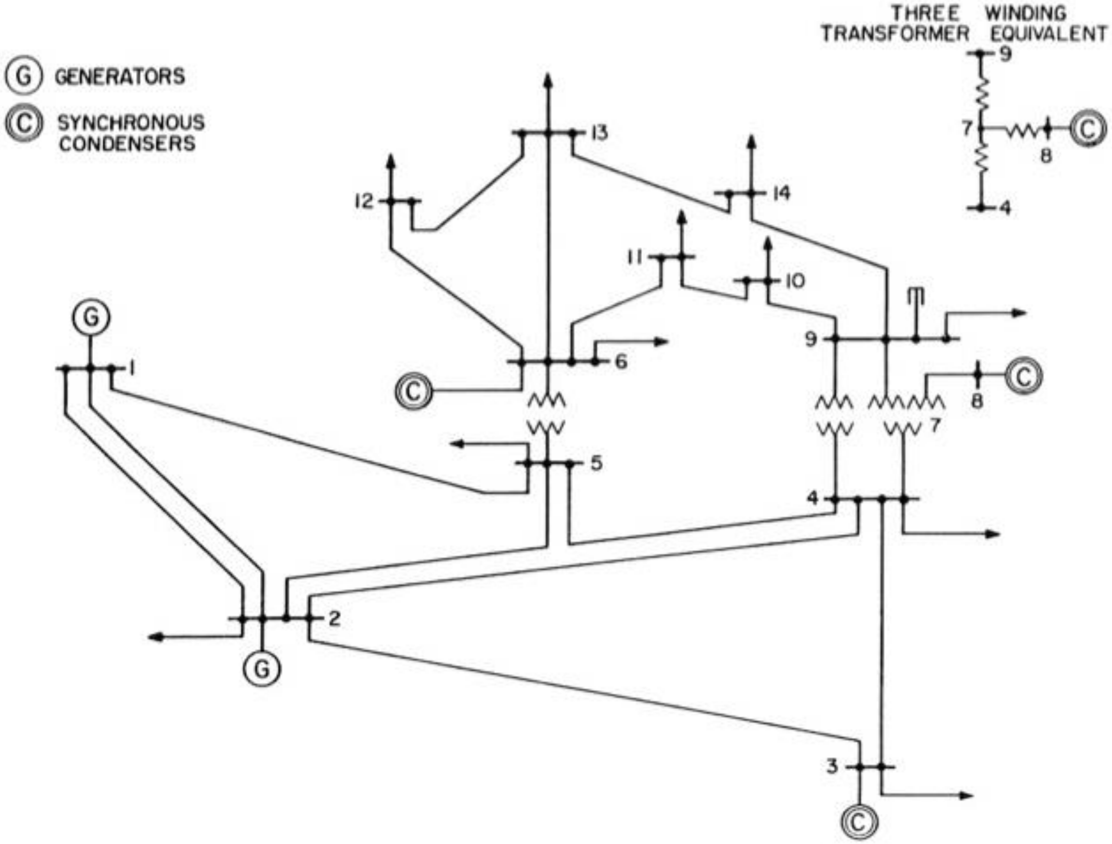


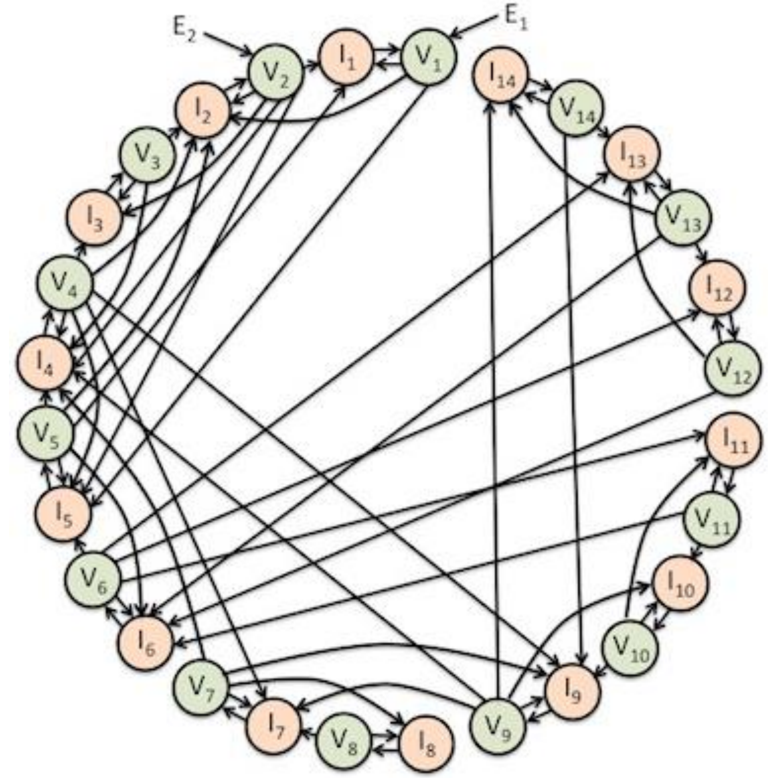
Circuit model



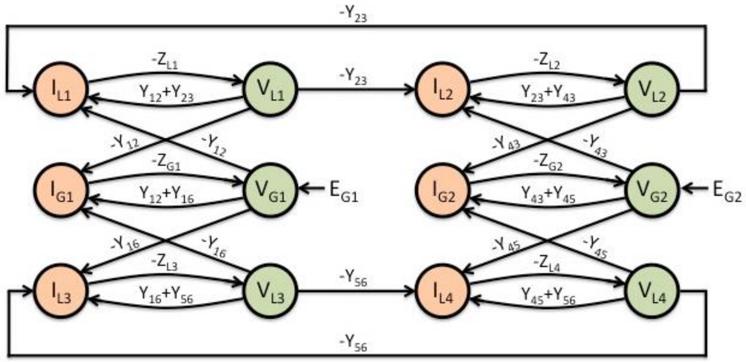


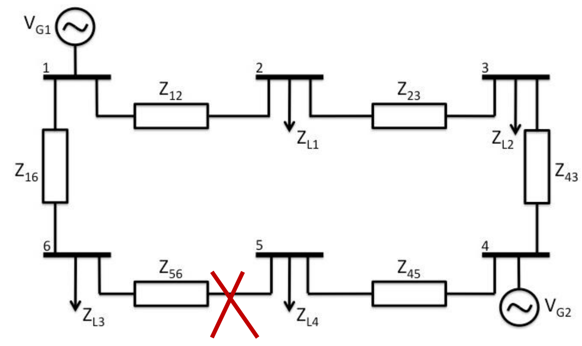
Graph model


GENERATORS
SYNCHRONOUS CONDENSERS
THREE WINDING TRANSFORMER EQUIVALENT
1
2
3
4
5
6
7
8
9
10
11
12
13
14

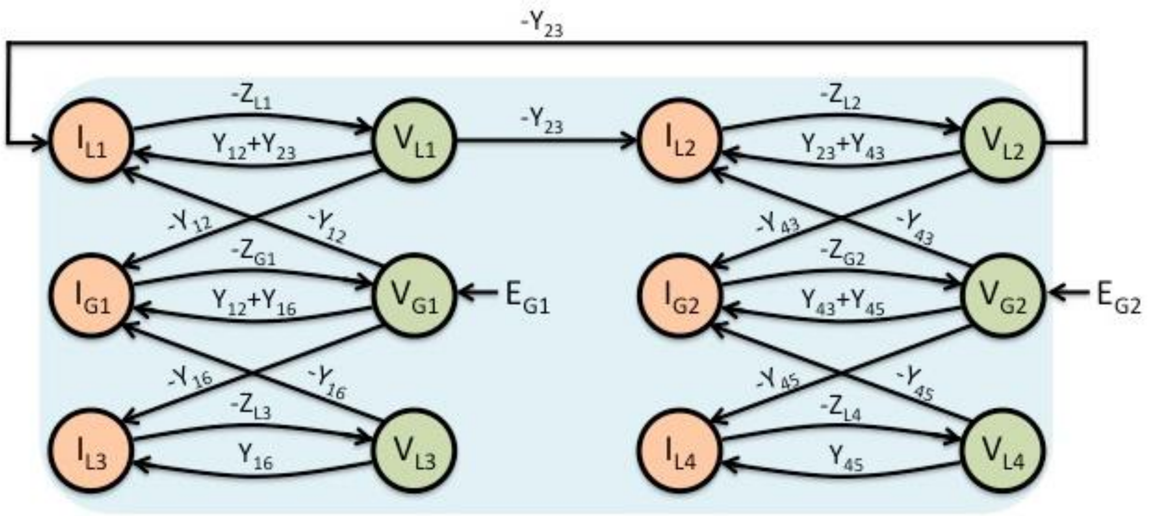

E_1
E_2
V_1
I_1
V_2
I_2
V_3
I_3
V_4
I_4
V_5
I_5
V_6
I_6
V_7
I_7
V_8
I_8
V_9
I_9
V_{10}
I_{10}
V_{11}
I_{11}
V_{12}
I_{12}
V_{13}
I_{13}
V_{14}
I_{14}

Effect of Single Contingency



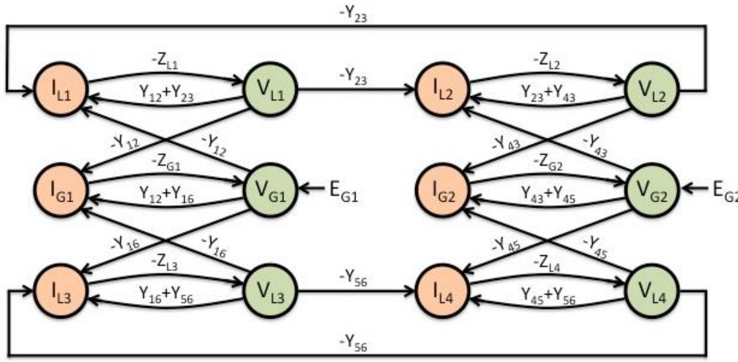


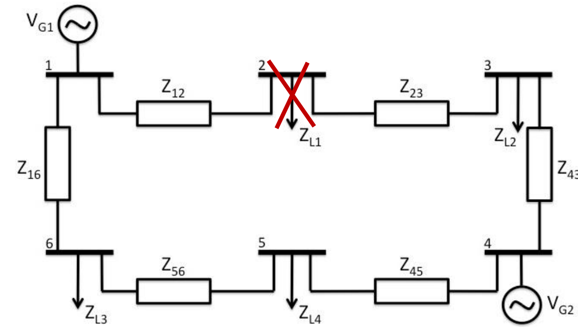
Single transmission line 5-6 tripping:



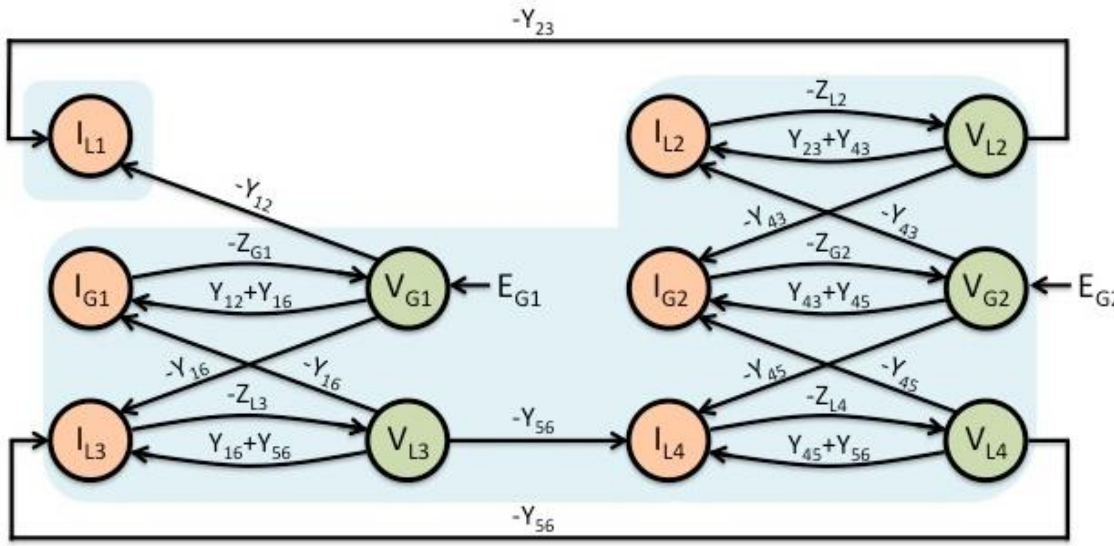
No loss of strong connectivity!

Effect of Single Contingency



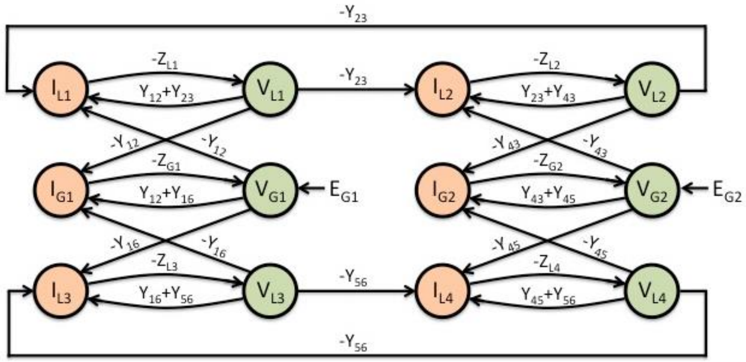


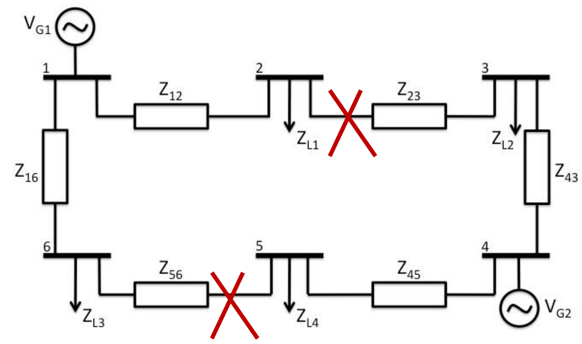
Three-phase fault at Load 1:



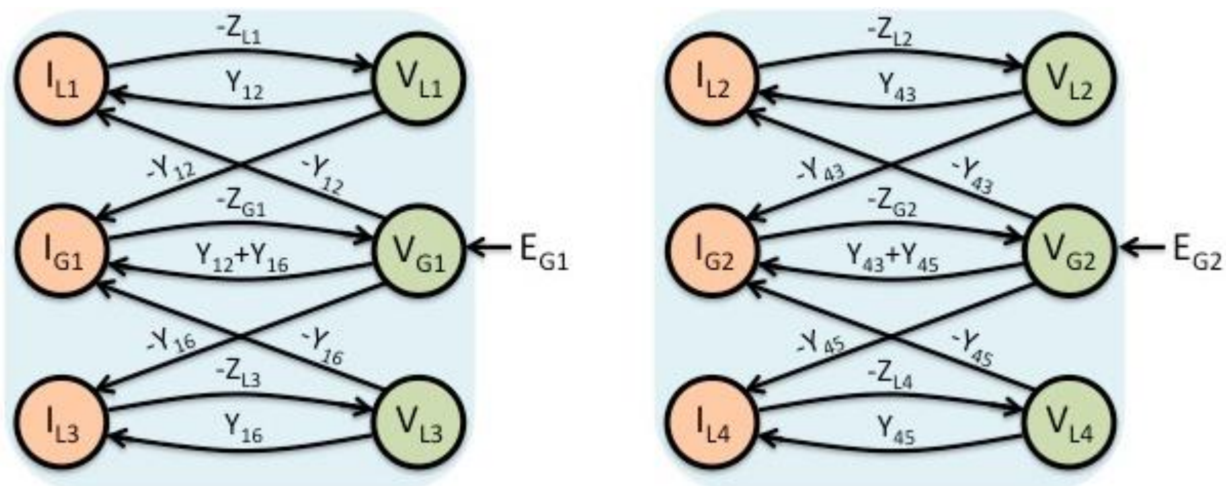
Loss of strong connectivity: two strongly connected components!

Effect of Double Contingency



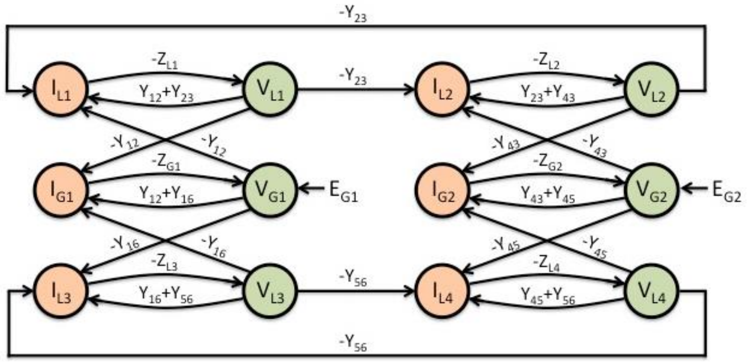


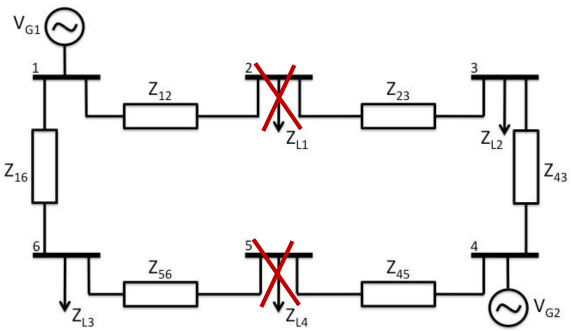
Double transmission line 5-6, 2-3 tripping:



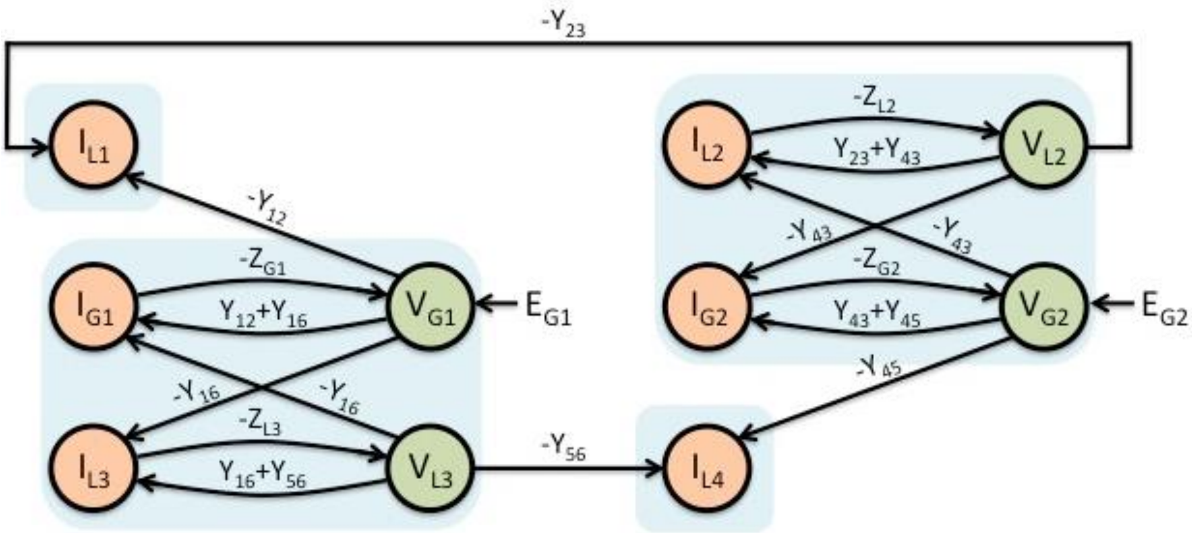
Loss of connectivity: two connected components!

Effect of Double Contingency





Two three-phase faults at Loads 1 and 4:



Loss of strong connectivity: four strongly connected components!

Main Theorem

Theorem: Under the conditions that

- the bus system is connected,
- all generators have nonvanishing internal impedance,

and the contingencies are restricted to

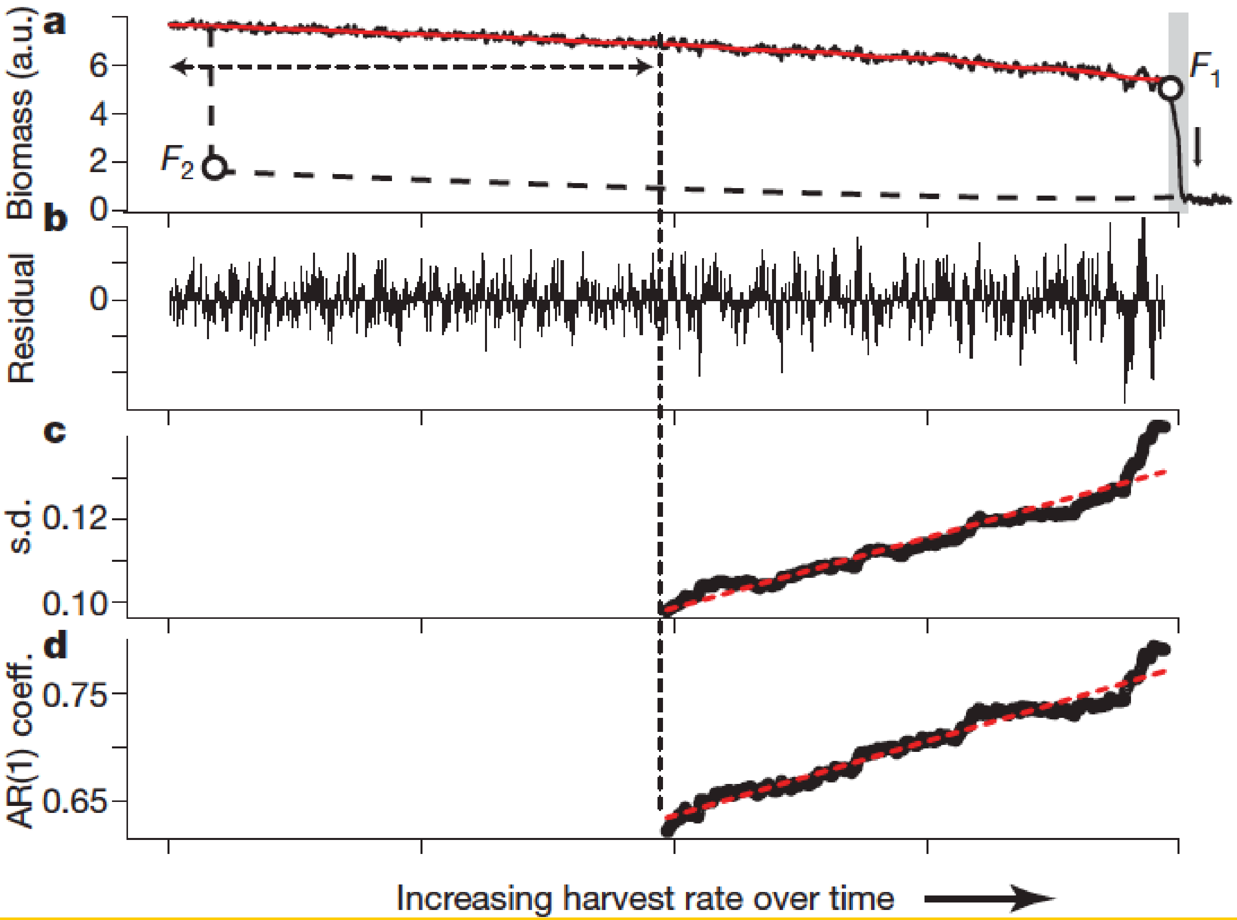
- single transmission line tripping,

the graph model is strongly connected.

Plan of Action

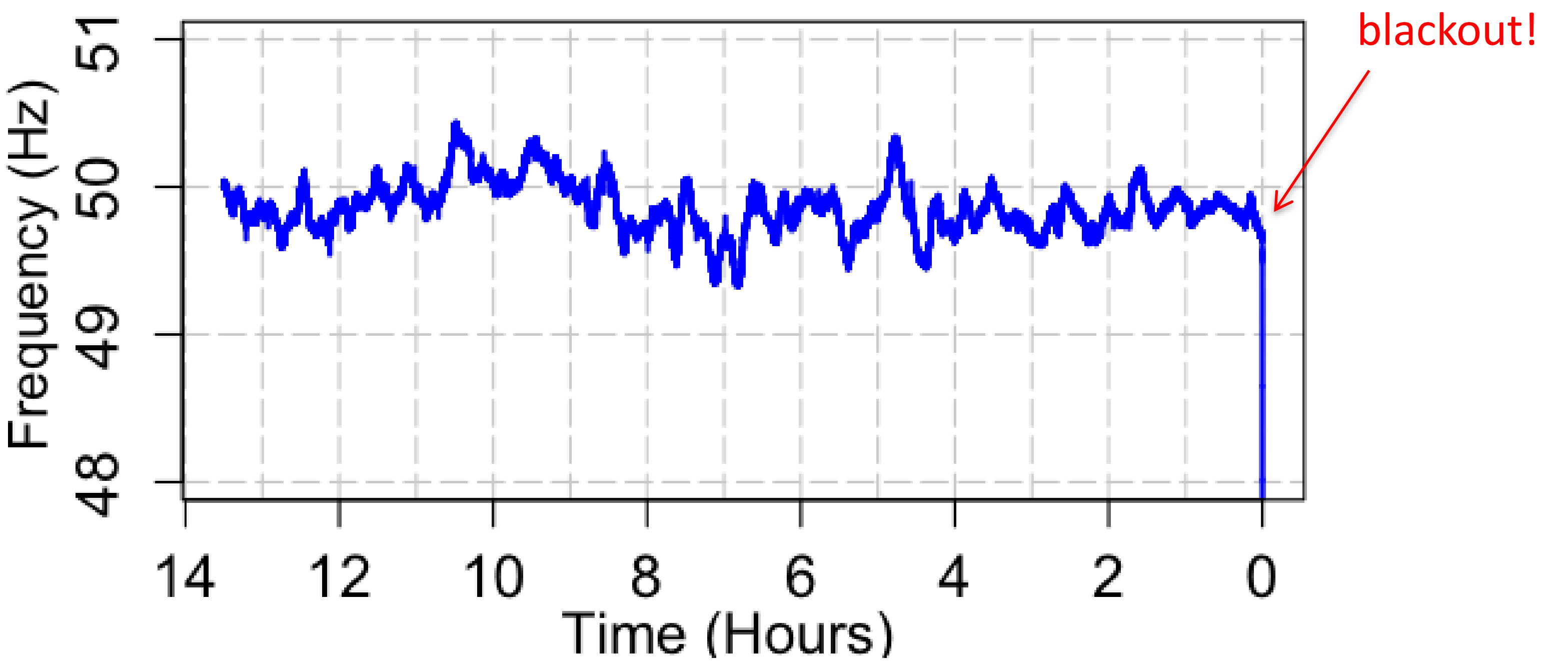
- Fractal PMU signal analysis
  - Texas & EPFL (Switzerland) normal PMU data
- Why are the PMU signals fractal???
  - Fractional dynamics load modeling
  - Hidden feedbacks in power grid
  - Strong connectivity of power grid graph, *aggregating all loads*
- Early warning of imminent blackout
  - Indian blackout PMU data
  - Shift in AR(1) coefficient and Hurst exponent.

Critical Transition in Harvested Population

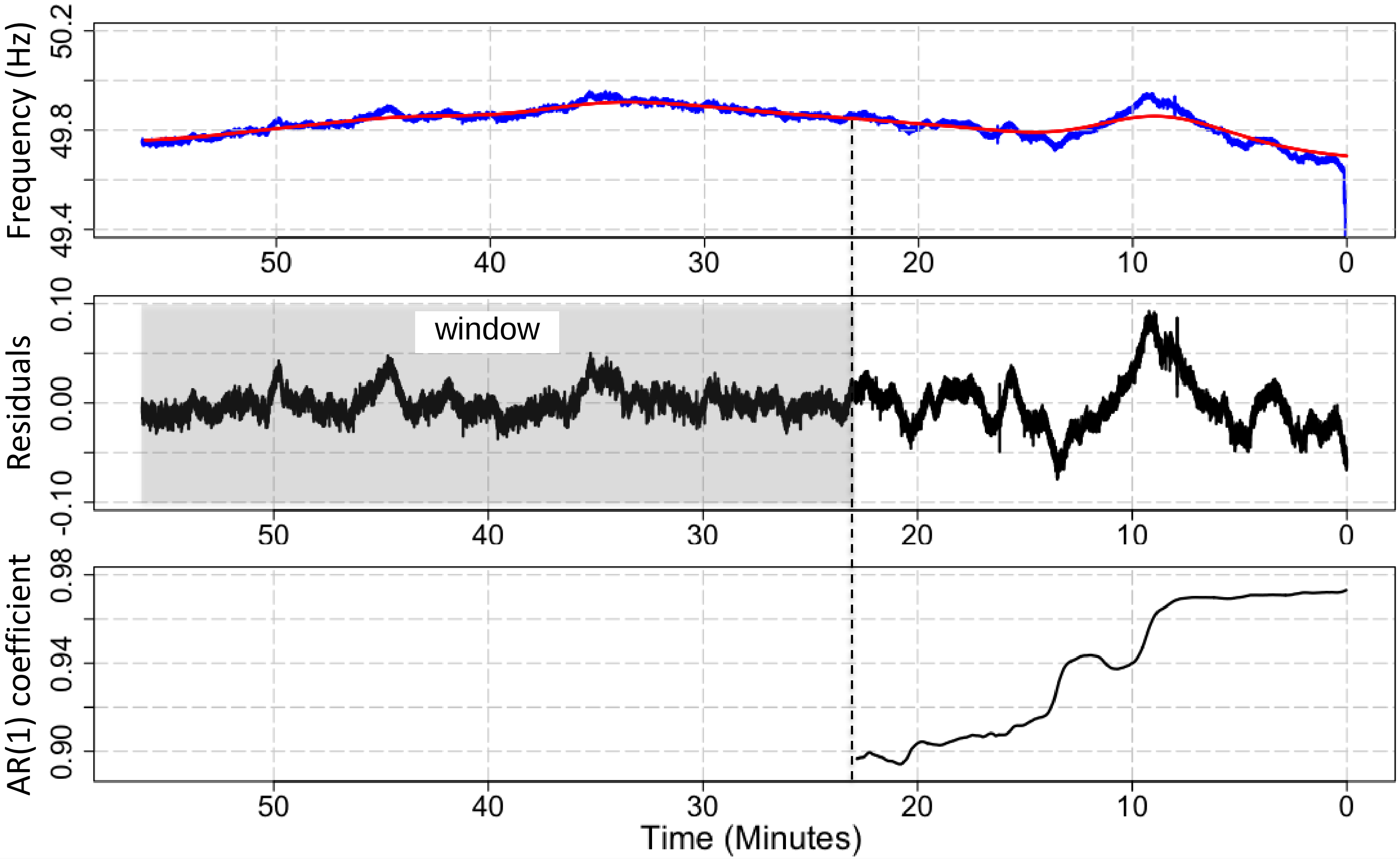


2012 Indian Blackout

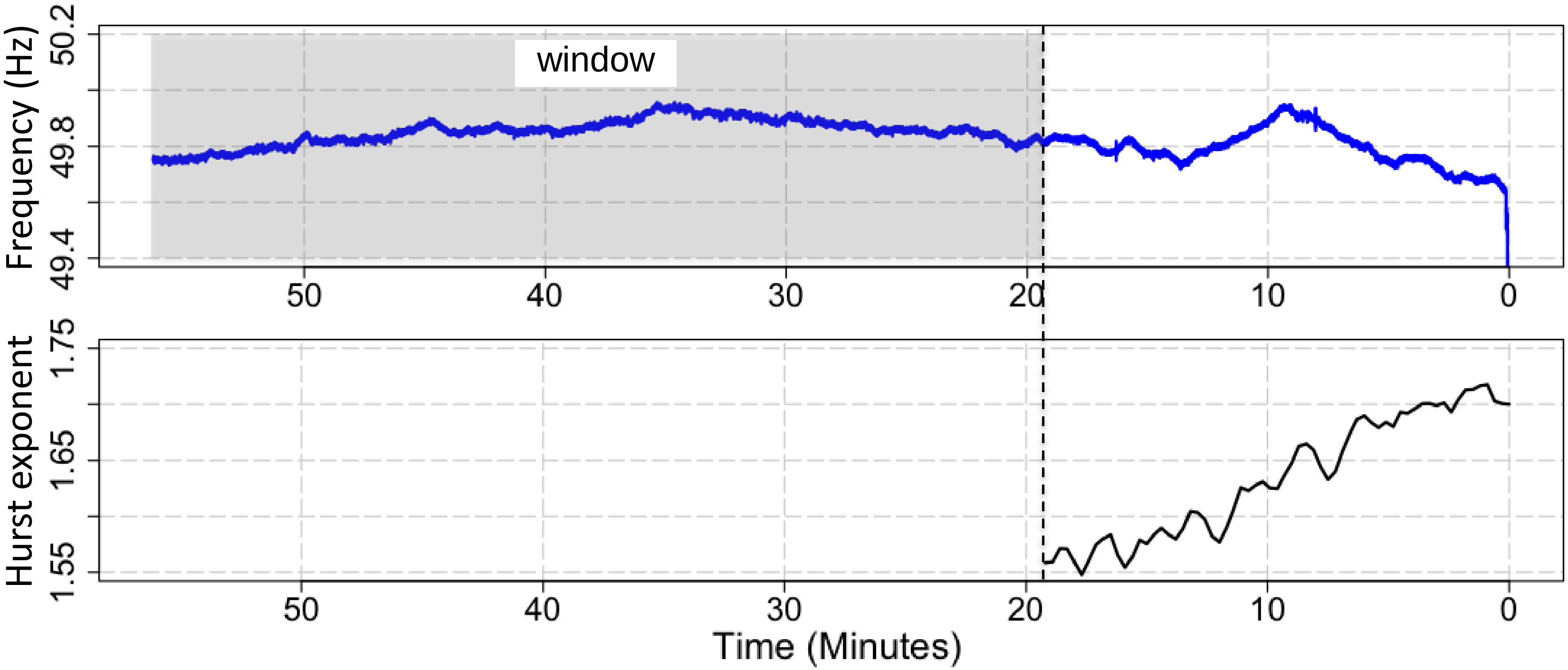
➢ **The blackout occurred on July 30, 2012 and affected more than 300 million people living in Northern India.**



Increase in Autoregressive Coefficient before Blackout



Increase in Hurst Exponent before Blackout



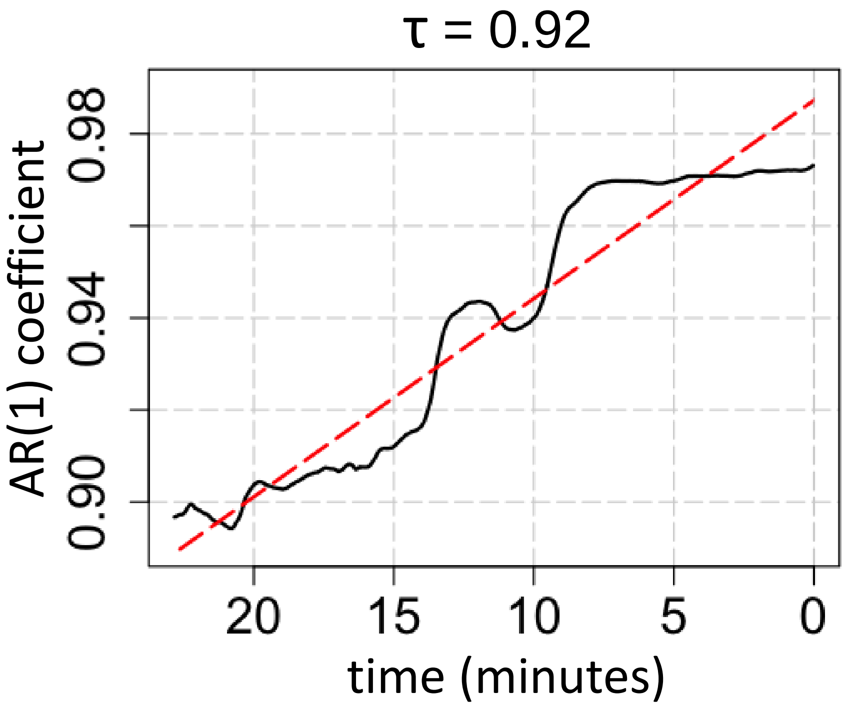
Kendall's tau

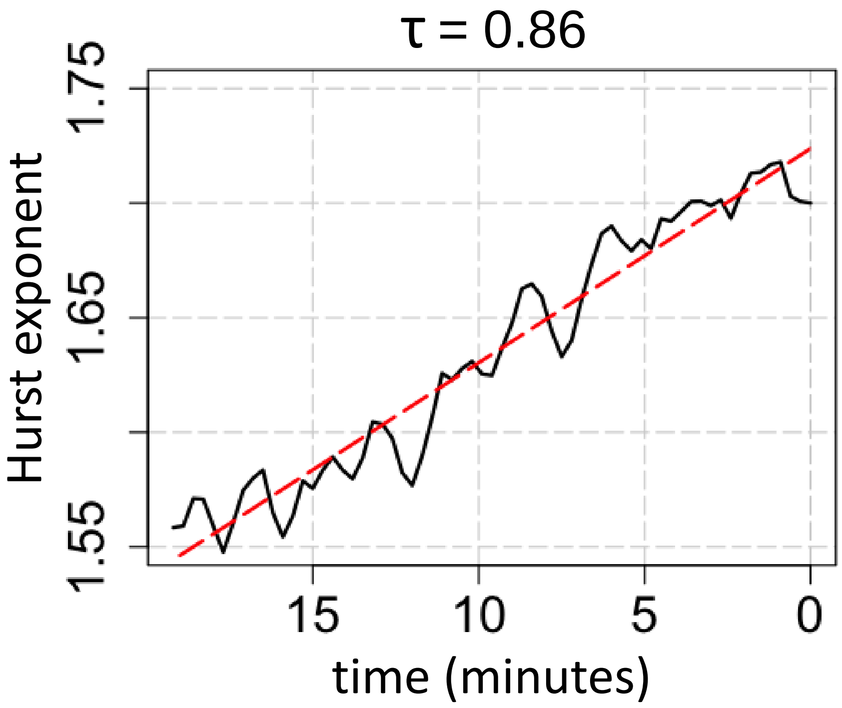
- Kendall's **tau is a rank correlation coefficient that is used to measure—in a statistically meaningful sense—the ordinal association between two datasets, $\{(t_i, \alpha_i)\}$.**
- **Assuming that we have n pairs of x and y data**
  - $((x_1,y_1); (x_2,y_2); \ldots; (x_n,y_n))$,
  - Kendall's **tau is defined as**

$$t = \frac{\#\text{ of concordant pairs} - \#\text{ of discordant pairs}}{n(n-1)/2}$$

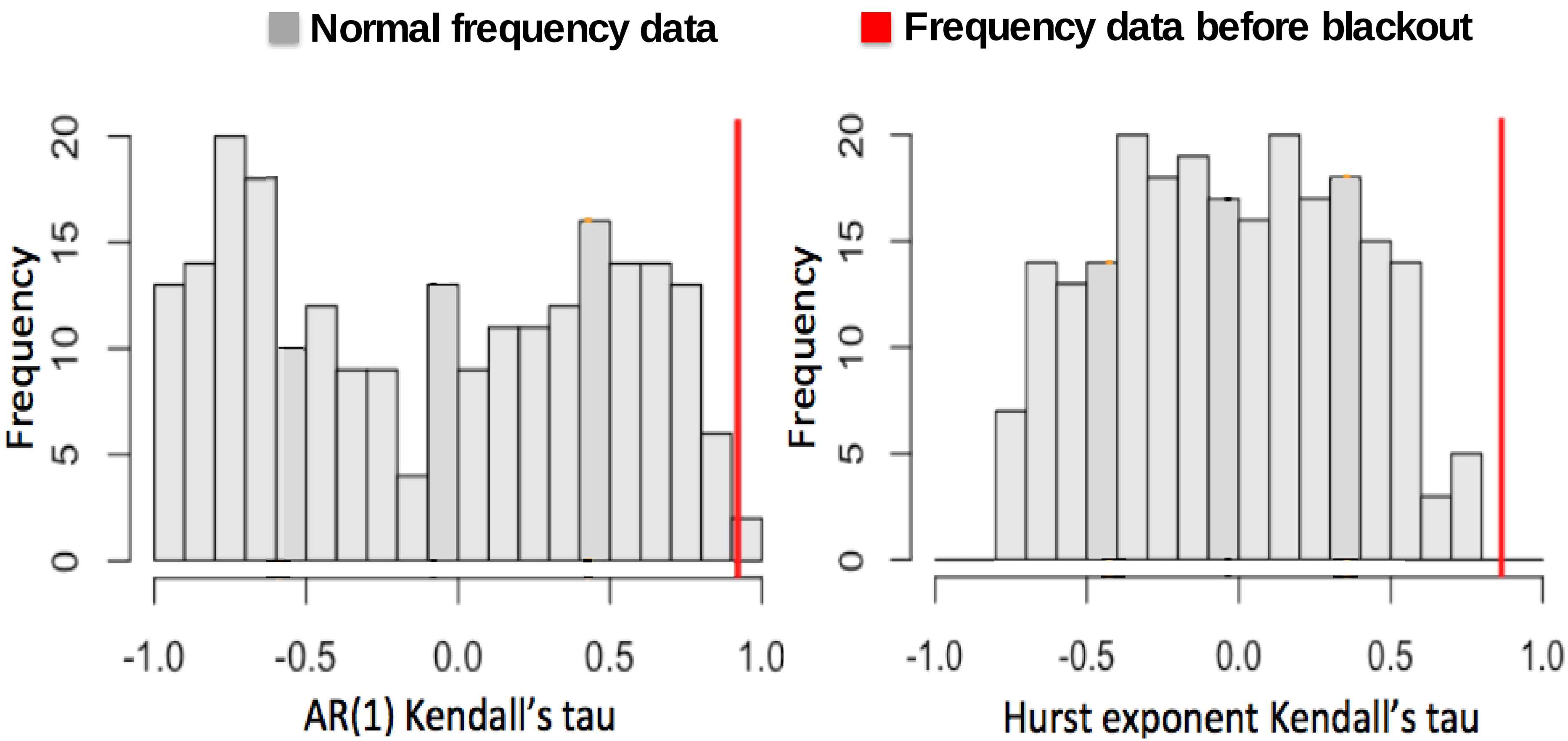
- **Concordant pair** ➡ $x_i > x_j$ & $y_i > y_j$ or $x_i < x_j$ & $y_i < y_j$
- **Discordant pair** ➡ $x_i > x_j$ & $y_i < y_j$ or $x_i < x_j$ & $y_i > y_j$

Kendall's Tau of AR(1) Coefficient versus Hurst Exponent





AR(1) versus Hurst Exponent Sample Distributions

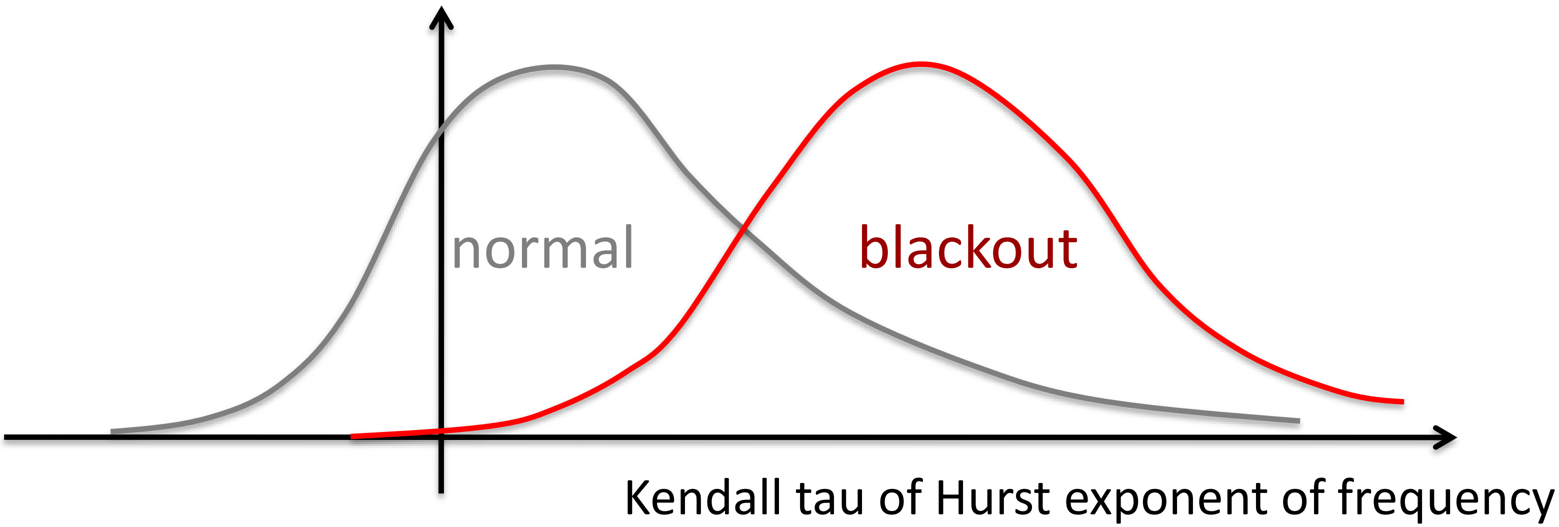


Conclusions

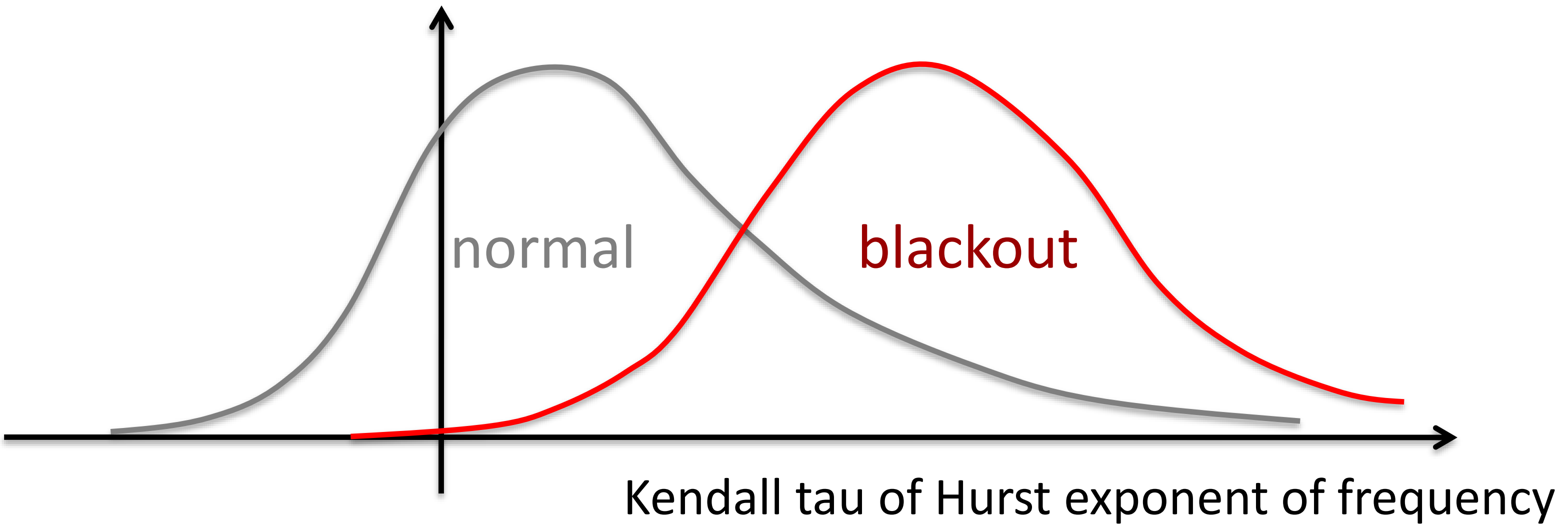
- The frequency appears to be the most relevant data to anticipate blackout.
- The fundamental observation is that the τ(AR(1)) and the τ(α) of the frequency blackout data point *are shifted to the right* of the empirical distributions of the Kendall tau of AR(1) and Hurst exponent of normal frequency data.
  - The shift is more pronounced for the Hurst data.
  - *The Hurst exponent of the frequency data appears the best bet to anticipate blackout.*
    - $1,000,000 question: Could it anticipate malware?
      - There is hope to achieve this as it was shown that during Distributed Denial of Service (DDoS) and UDP flooding attacks the Akaike/Kolmogorov informational statistics of the link utilization signals changed!

Future Work

- With more blackout data points, we hope to demonstrate—with enough confidence—that the empirical distributions of the normal and blackout Hurst frequency data *are random draws from different distributions.*



Thank you!
Questions?
joncKhee@usc.edu



Conclusions

- The fractal behavior of the PMU signals is puzzling…
- Its potential for anticipating black-out and/or cyber attacks has been demonstrated.
- So, it is of paramount importance to understand *why the PMU signals are fractal.*
  - The Berg load models provide a clue with their fractional exponents of ω.
  - In the Berg experiment, the load is modeled in its microgrid environment.
  - The aggregation of the loads combines a great many lumped parameter circuit elements *to make distributed parameter elements.*