NANYANG TECHNOLOGICAL UNIVERSITY

新加坡南洋理工大学

Time Series Approaches to Understanding Physical Phenomena

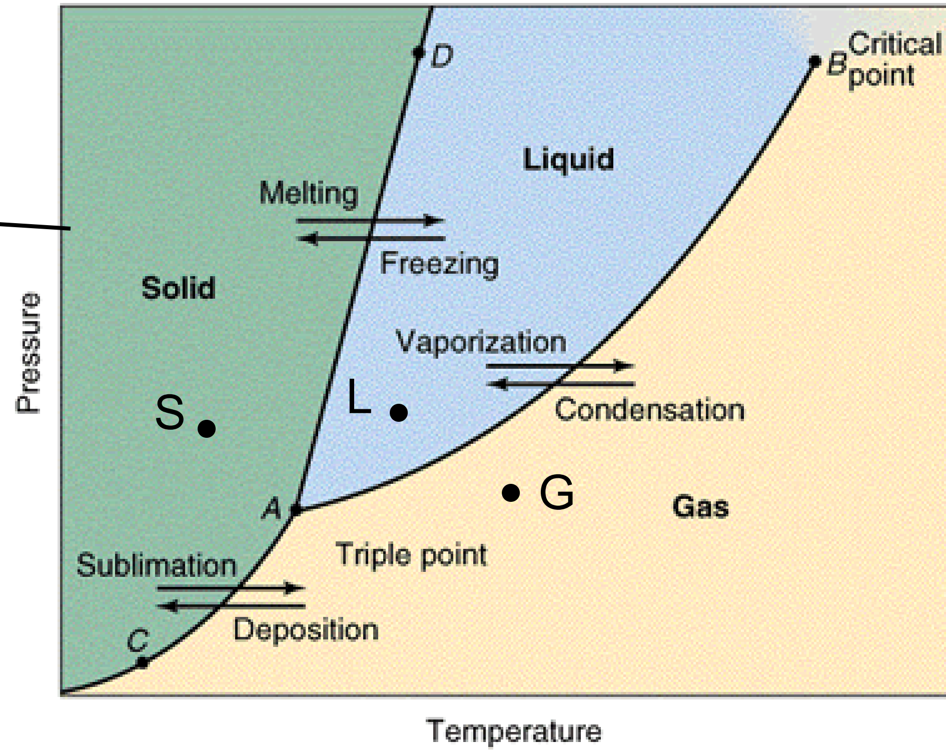
CHEONG Siew Ann

张 寿 安

cheongsa@ntu.edu.sg

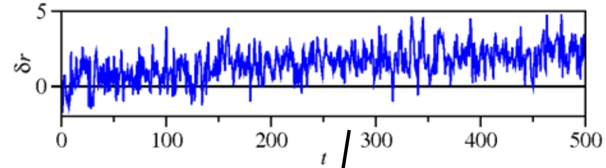
http://www1.spms.ntu.edu.sg/~cheongsa/

Macroscopic Thermal Physics

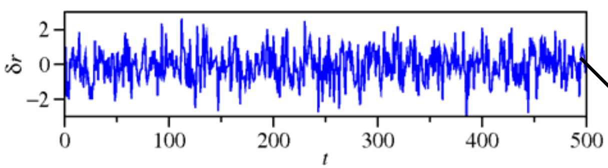


- Macroscopic order parameters differentiate
  - Solid (S)
  - Liquid (L)
  - Gas (G)

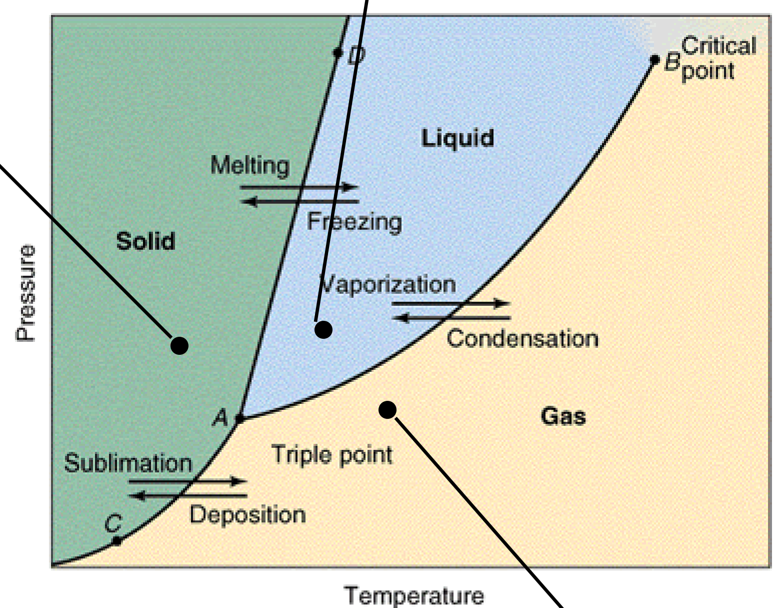
Microscopic Statistical Physics



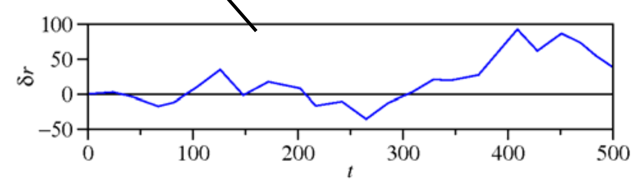
diffusive trajectories,
δr^2 increases with time



δr fluctuates about 0,
$\delta r^2 = \alpha T$ time-independent

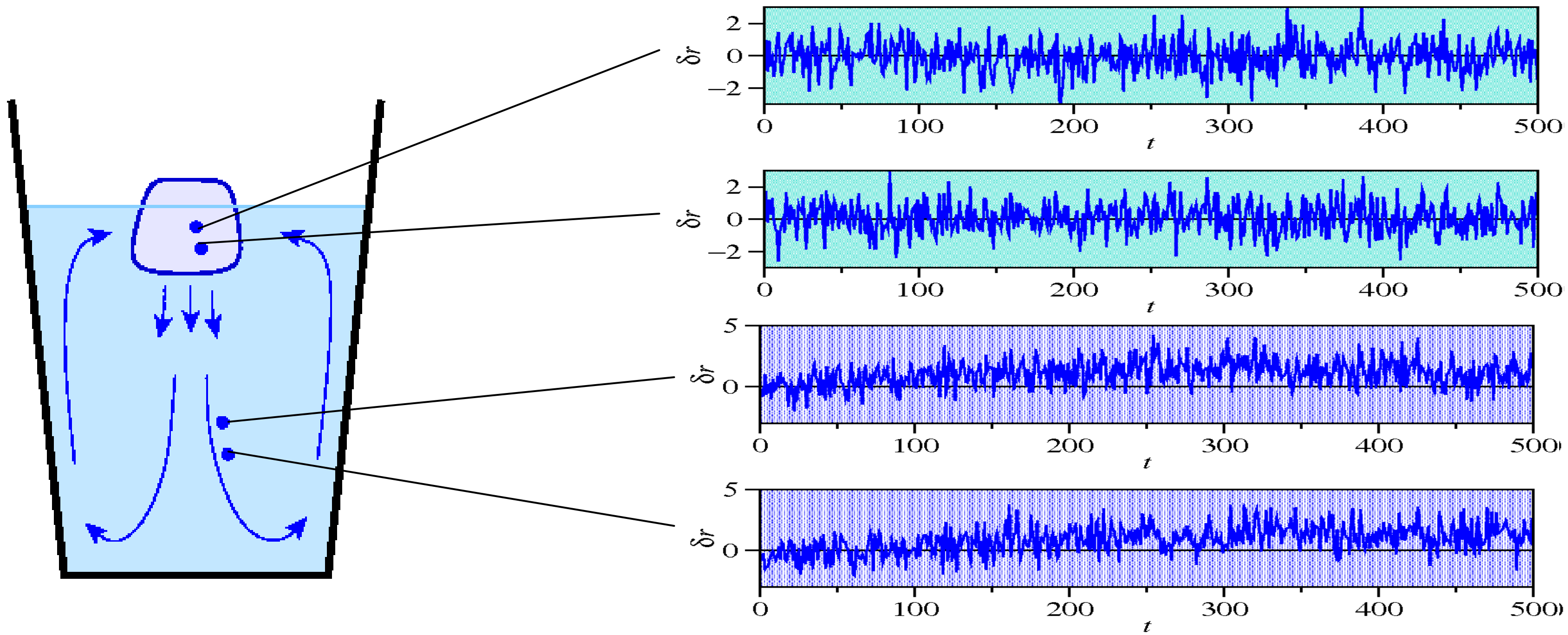


ballistic trajectories,
infrequent collisions



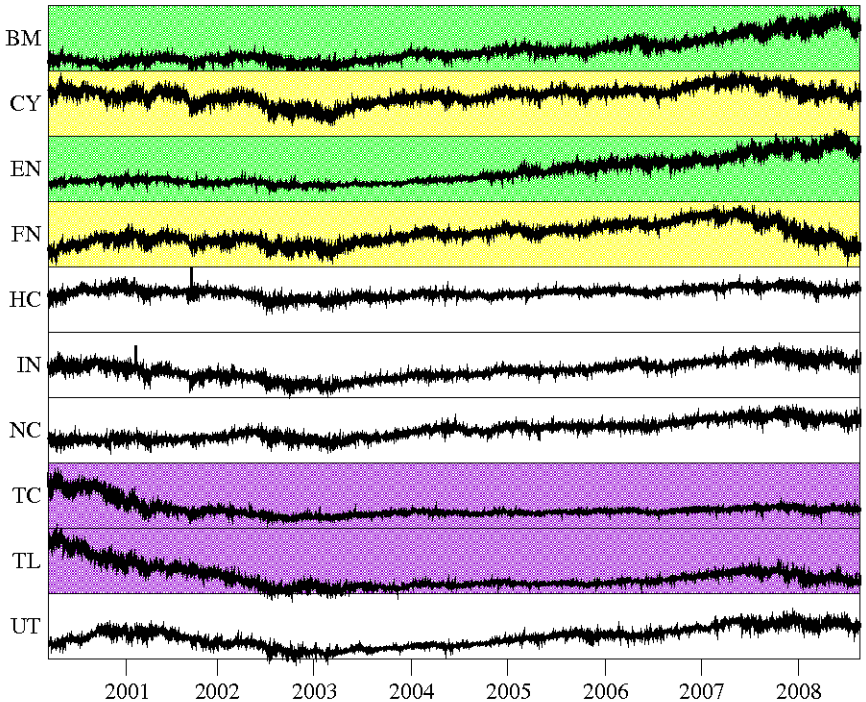
- S, L, G time series distinguishable
- S, L, G phase within single time series distinguishable

Macro ↔ Micro



- Group statistically similar time series
- Discover presence of different phases

Cross Correlations Between Time Series



Dow Jones US economic sector indices

$$C_{ij} = \frac{\left\langle \left(x_i - \bar{x}_i\right)\left(x_j - \bar{x}_j\right)\right\rangle}{\sigma_i \sigma_j} = \left\langle \frac{\delta x_i}{\sigma_i}\frac{\delta x_j}{\sigma_j}\right\rangle$$

↓

$$D_{ij} = \left\langle \theta\left(\frac{\delta x_i}{\sigma_i} - 1\right)\theta\left(\frac{\delta x_j}{\sigma_j} - 1\right)\right\rangle$$

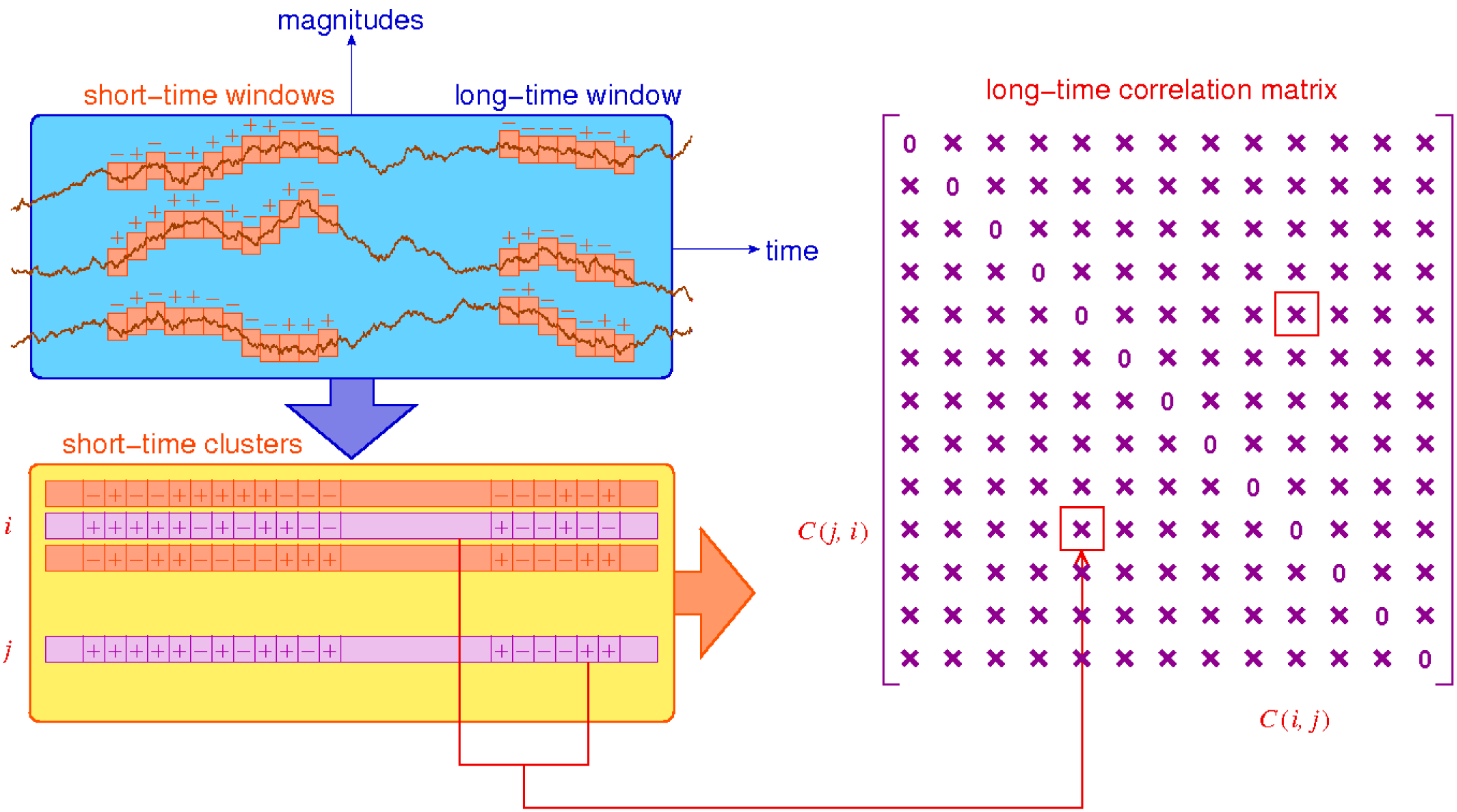
digital cross correlations

↓

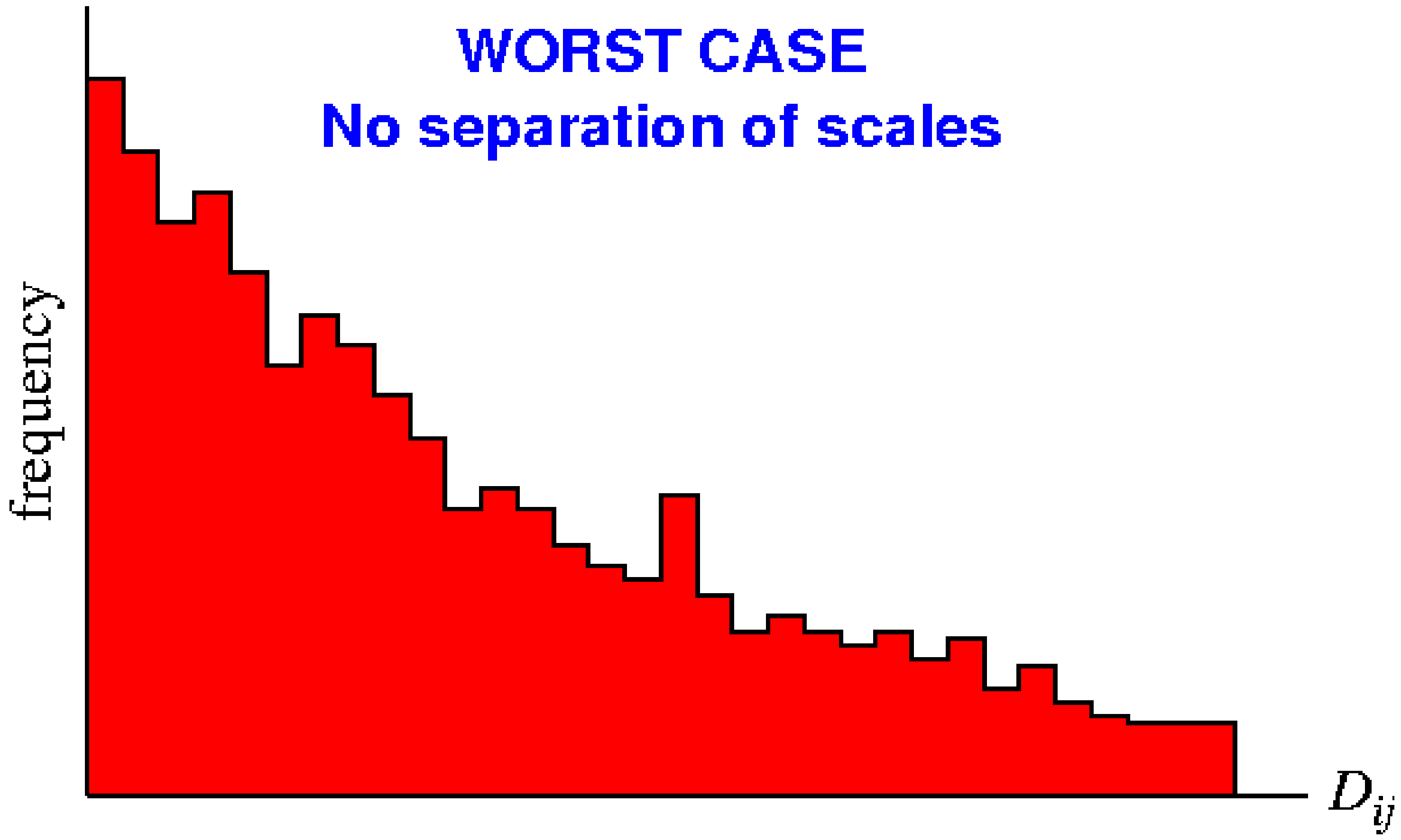
$$\widetilde{D}_{ij} = \left\langle \theta\left(\Delta x_i \Delta x_j\right)\right\rangle \text{ or } \widetilde{D}_{ij} = \sum_{t=1}^{N} \theta\left(\Delta x_{it} \Delta x_{jt}\right)$$

comovement digital cross correlations

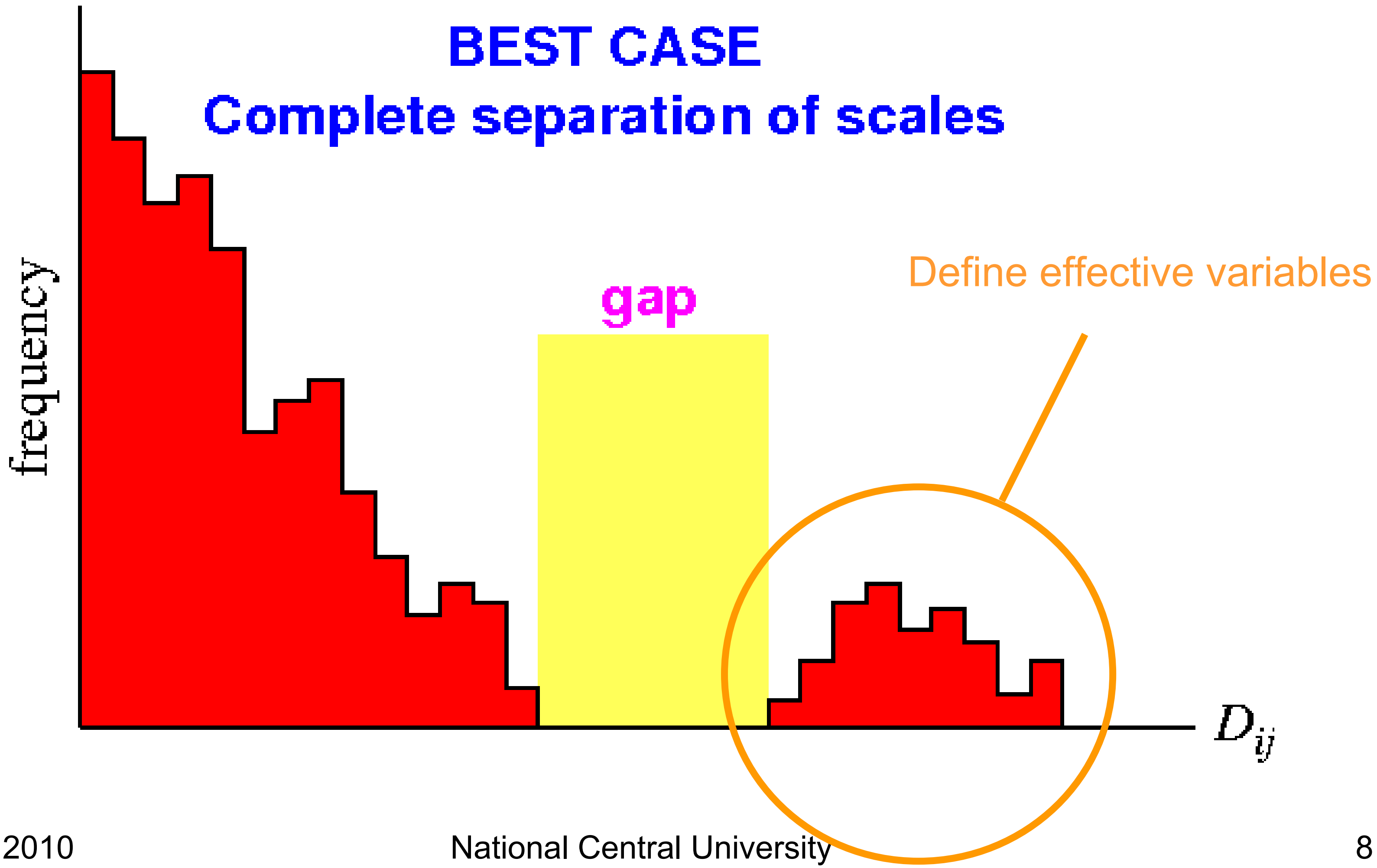
Long-Time Correlation Matrix



No Separation of Scales



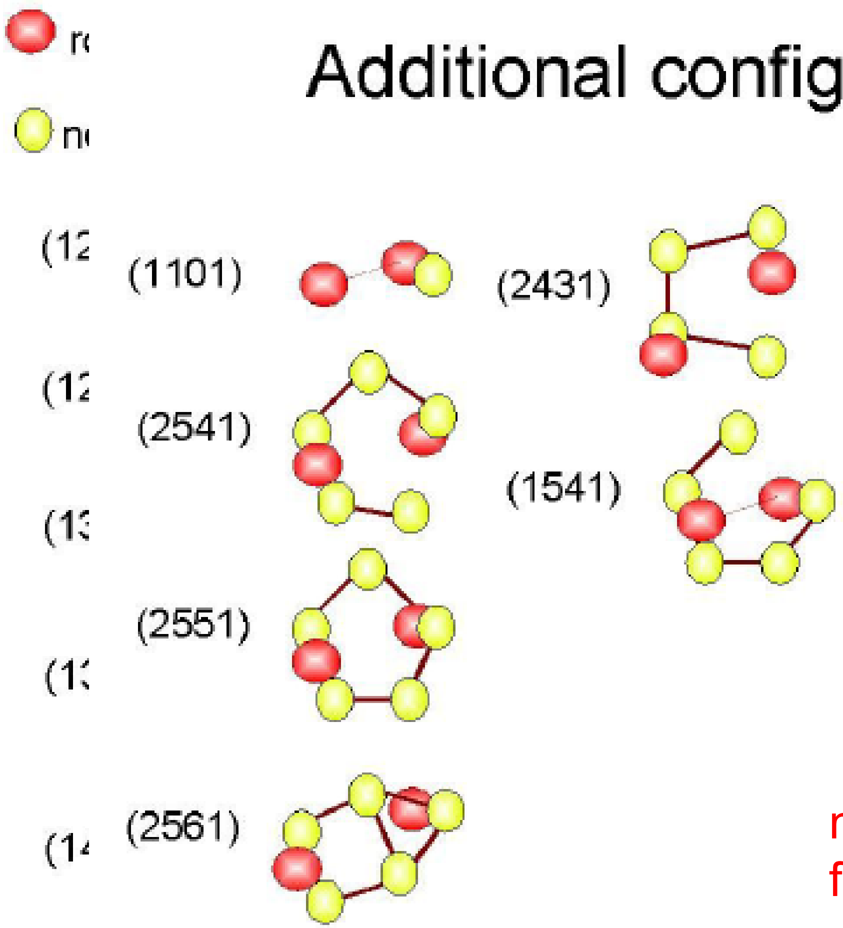
Complete Separation of Scales



Melting of Metallic Nanocluster

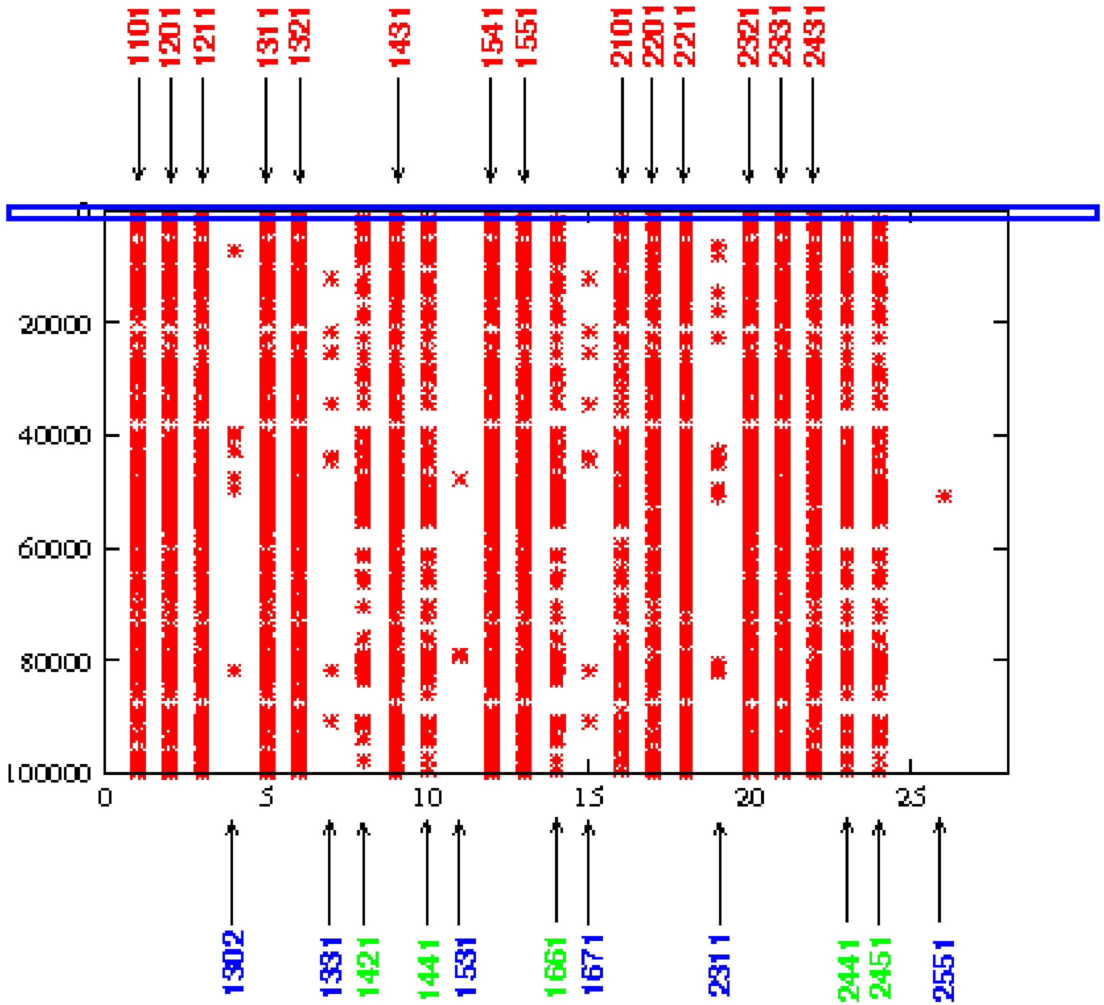
- Prof S K Lai
- Isothermal molecular dynamics (MD) simulations
  - Small number of metal atoms
- Determination of melting temperature
  - Specific heat
  - Lindemann parameter
  - Velocity autocorrelation function

Common-Neighbor Configurations



monitor their abundances as functions of time

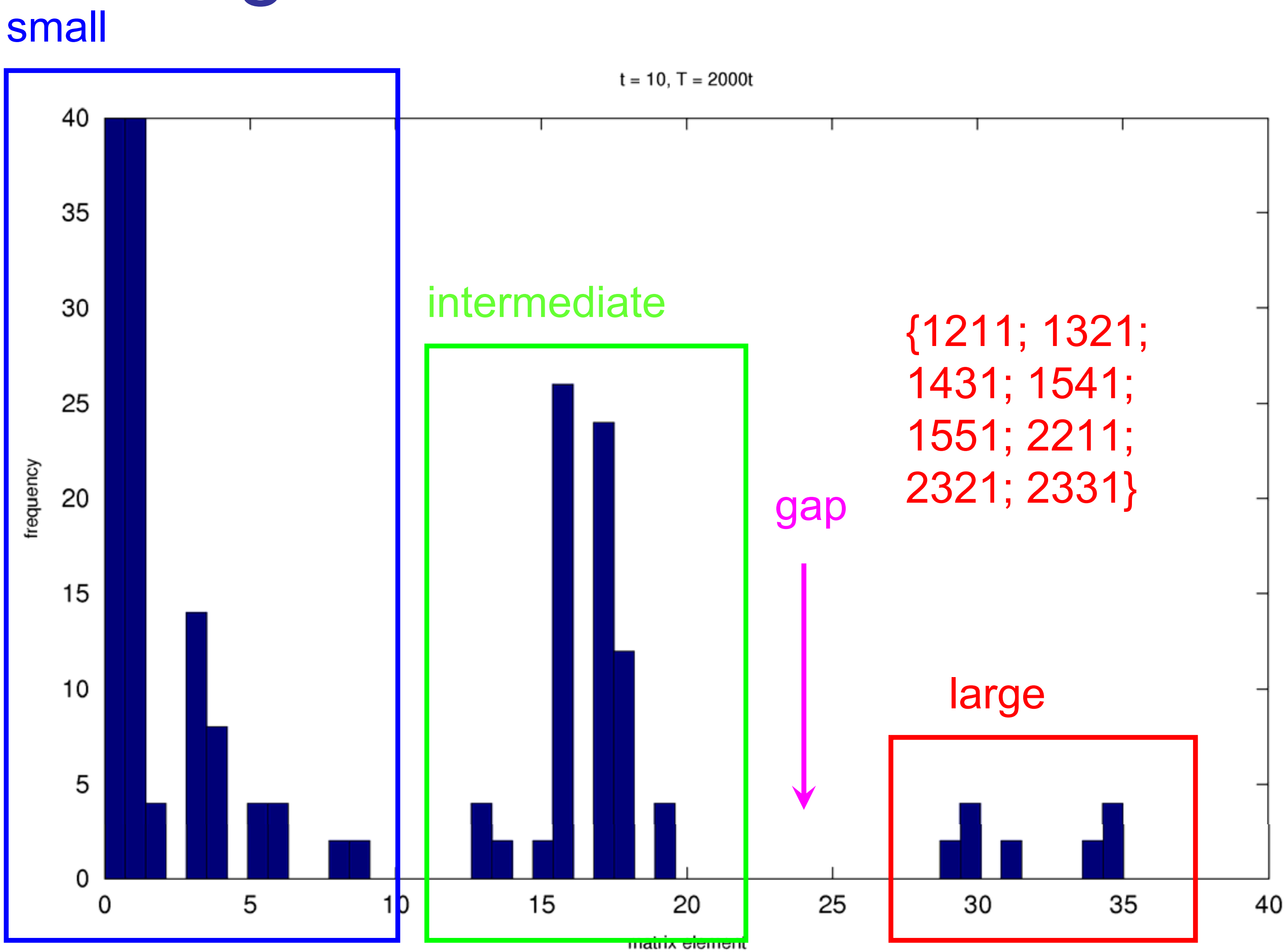
Ag_{14} at 300 K



T = 2000

91 × 91 comovement digital cross correlations matrix

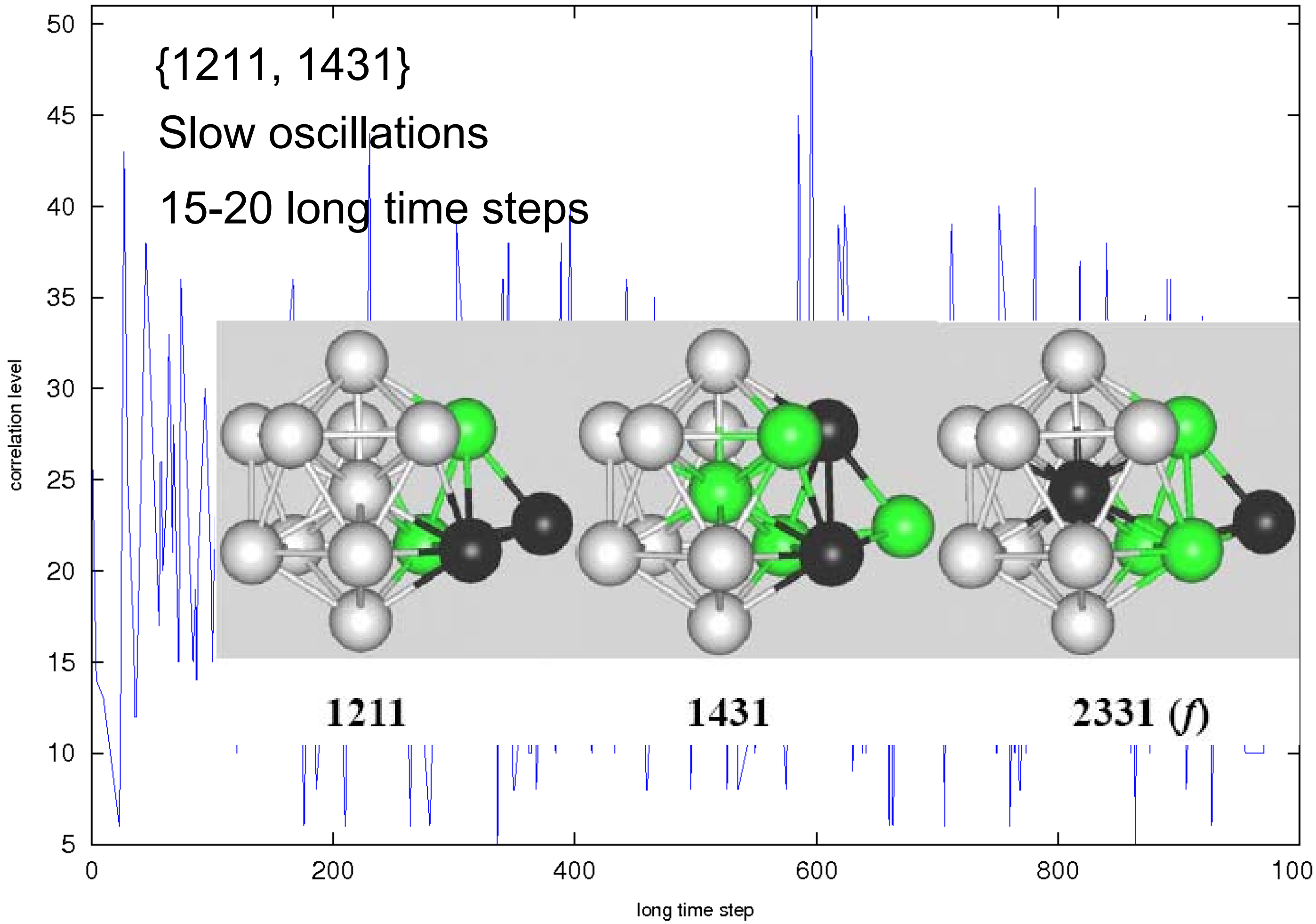
Histogram of Cross Correlations



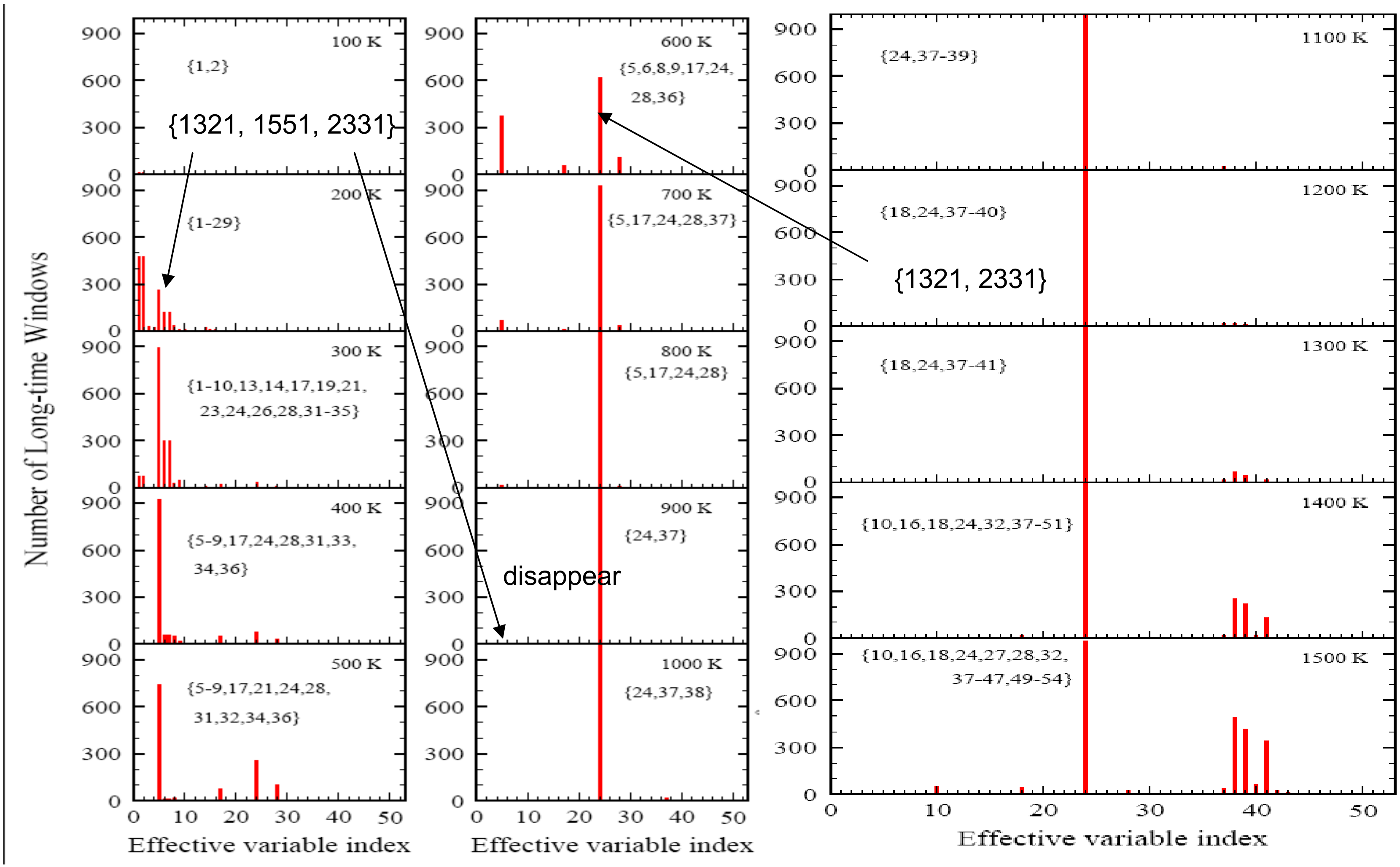
Effective Variables

	1321	1551	2331	1541	2211	2321	1211	1431
1321	0	35	35	1	3	1	17	18
1551	35	0	34	0	3	1	16	17
2331	35	34	0	1	2	0	17	17
1541	1	0	1	0	30	29	15	18
2211	3	3	2	30	0	30	13	16
2321	1	1	0	29	30	0	13	14
1211	17	16	17	15	13	13	0	31
1431	18	17	17	18	16	14	31	0

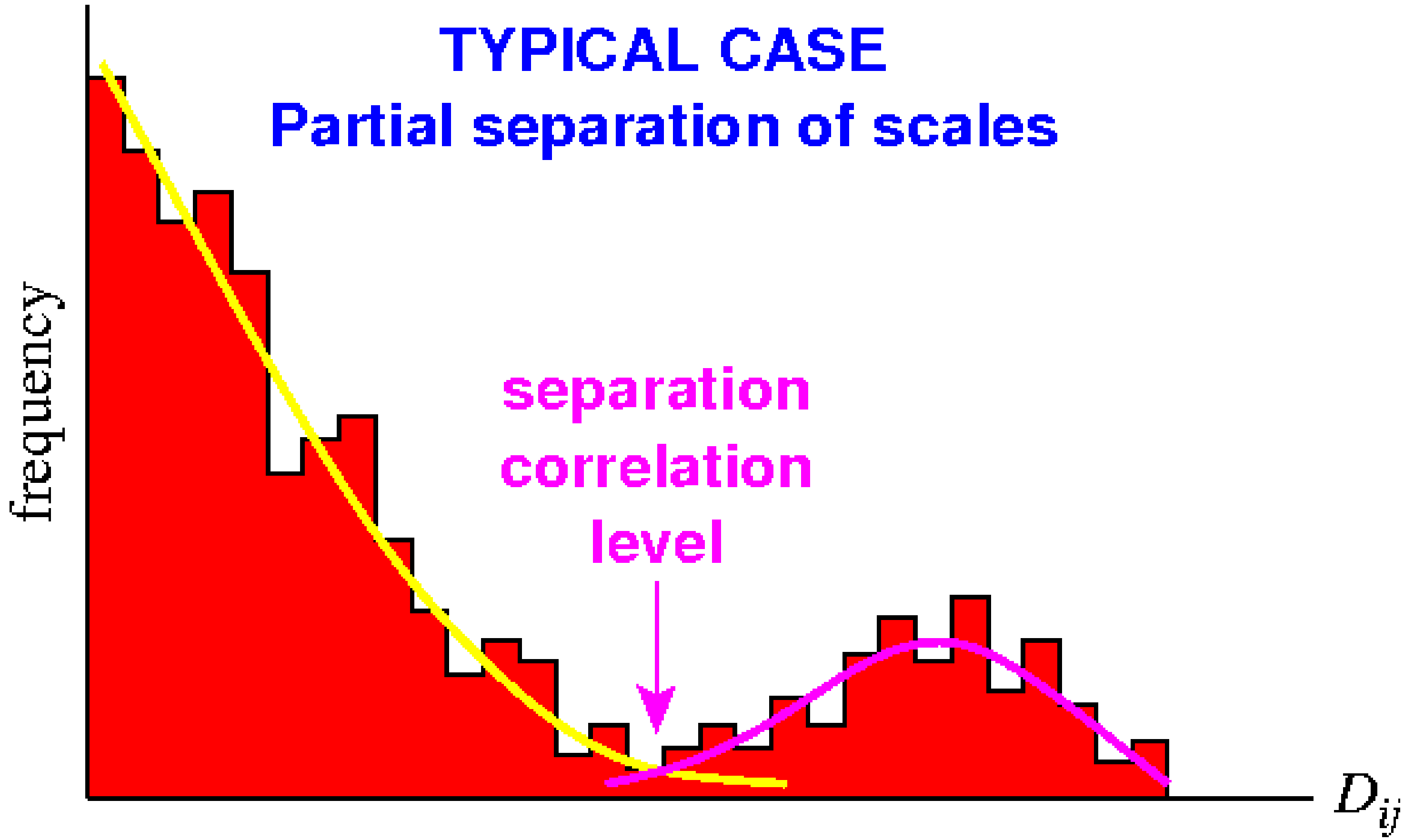
Dynamics of Effective Variables



Temperature Dependence



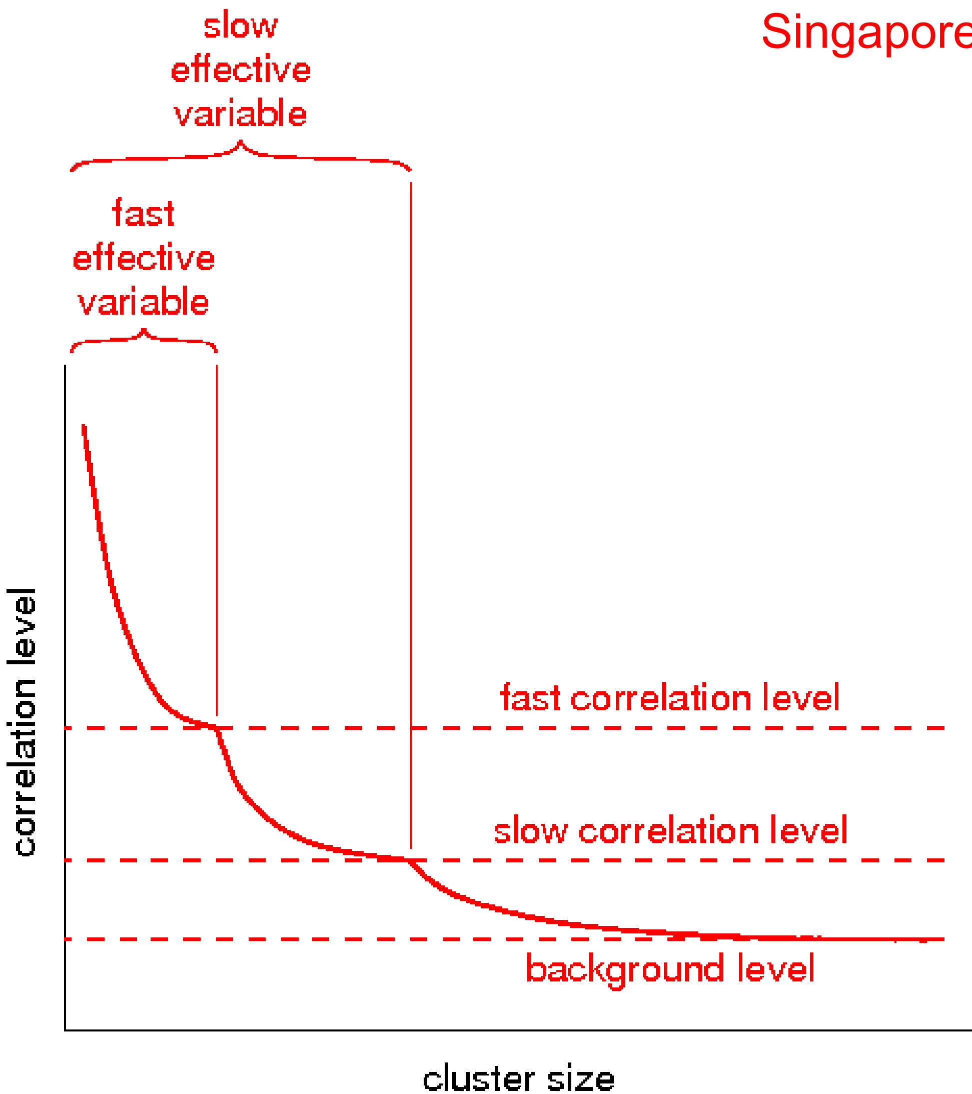
Separation of Scales



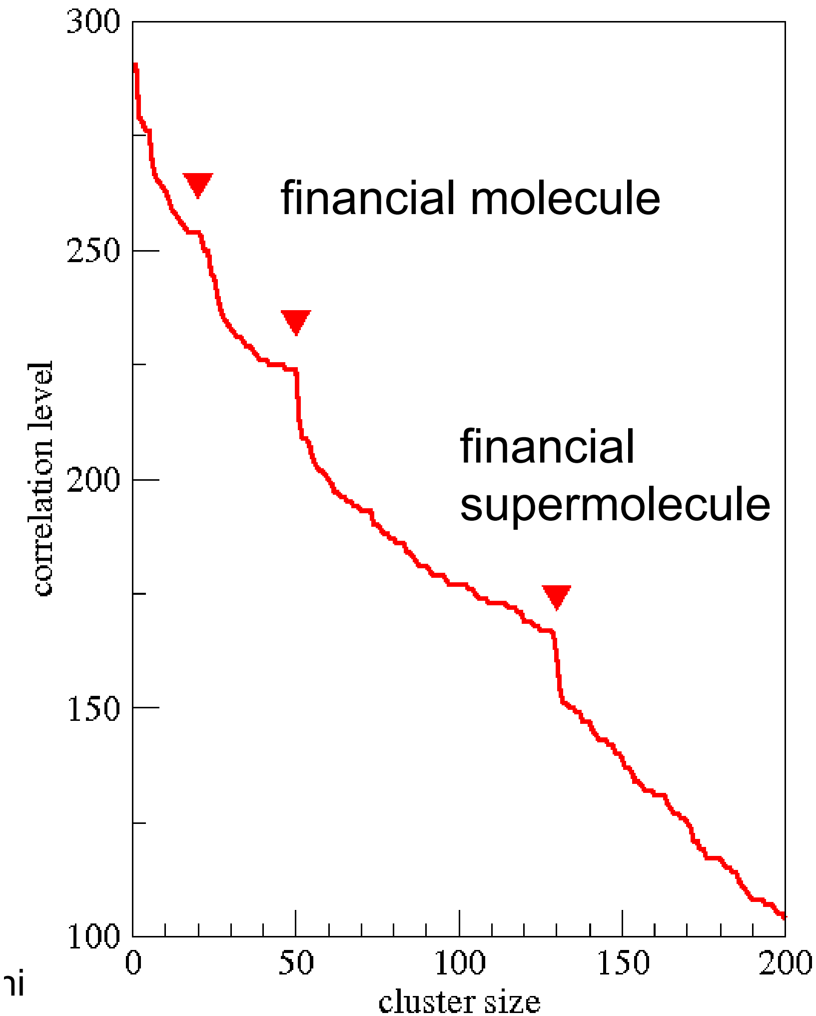
Partial Hierarchical Clustering

- Seed cluster
  - Find $D_{i^*j^*} = \max_{ij} D_{ij}$
  - Use $c = \{i^*, j^*\}$ as seed cluster
- Grow cluster
  - Add k^* to cluster if
    - $D_{k,c} = \min_{l \in c} D_{kl}$
    - $D_{k^*,c} = \max_k D_{k,c}$
  - Iterate
- Cluster boundary
  - Plot correlation level $D_{k^*,c}$ against cluster size

Hierarchy of Effective Variables



Singapore Stock Exchange (SGX): 2006-2007

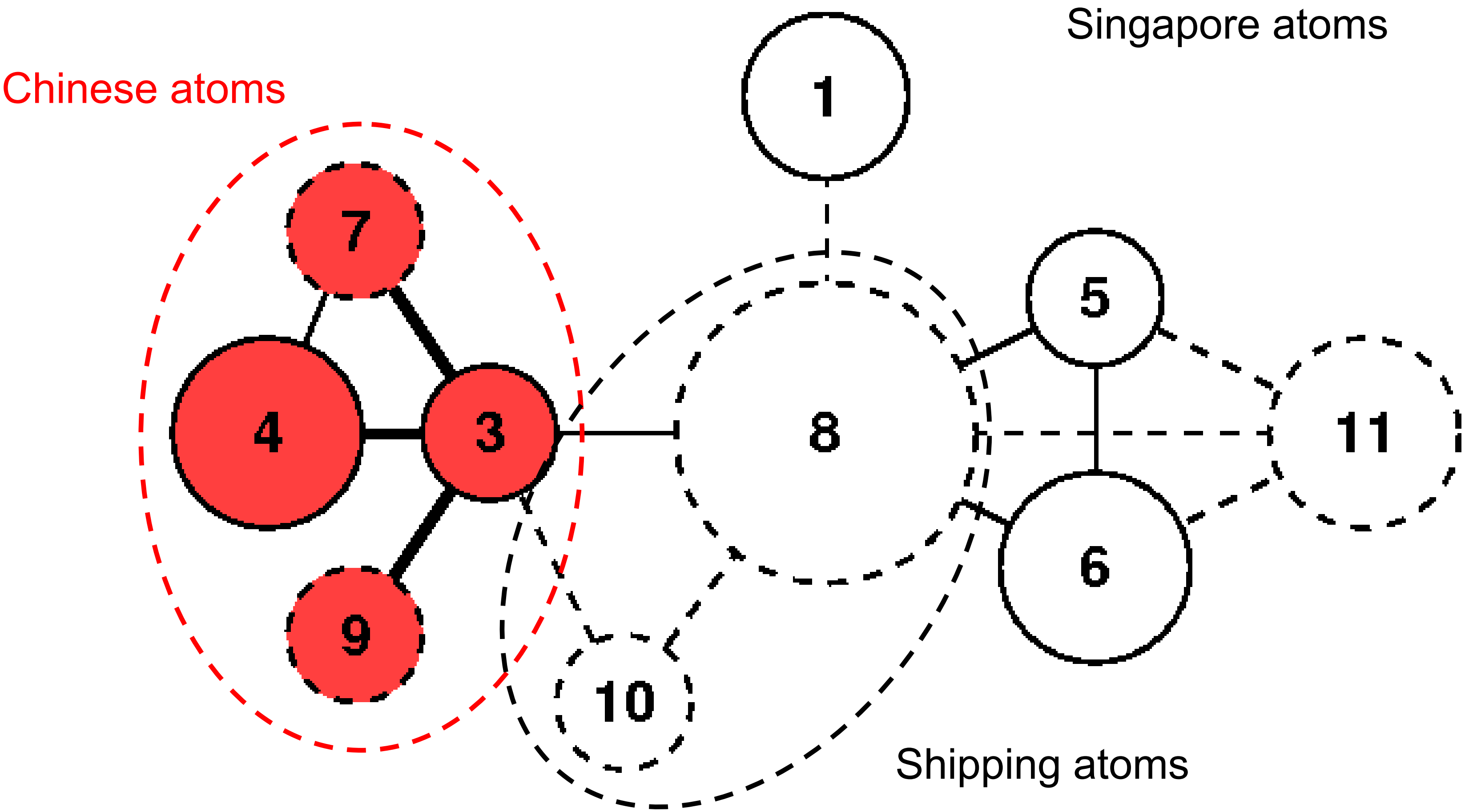


SGX Financial Atoms

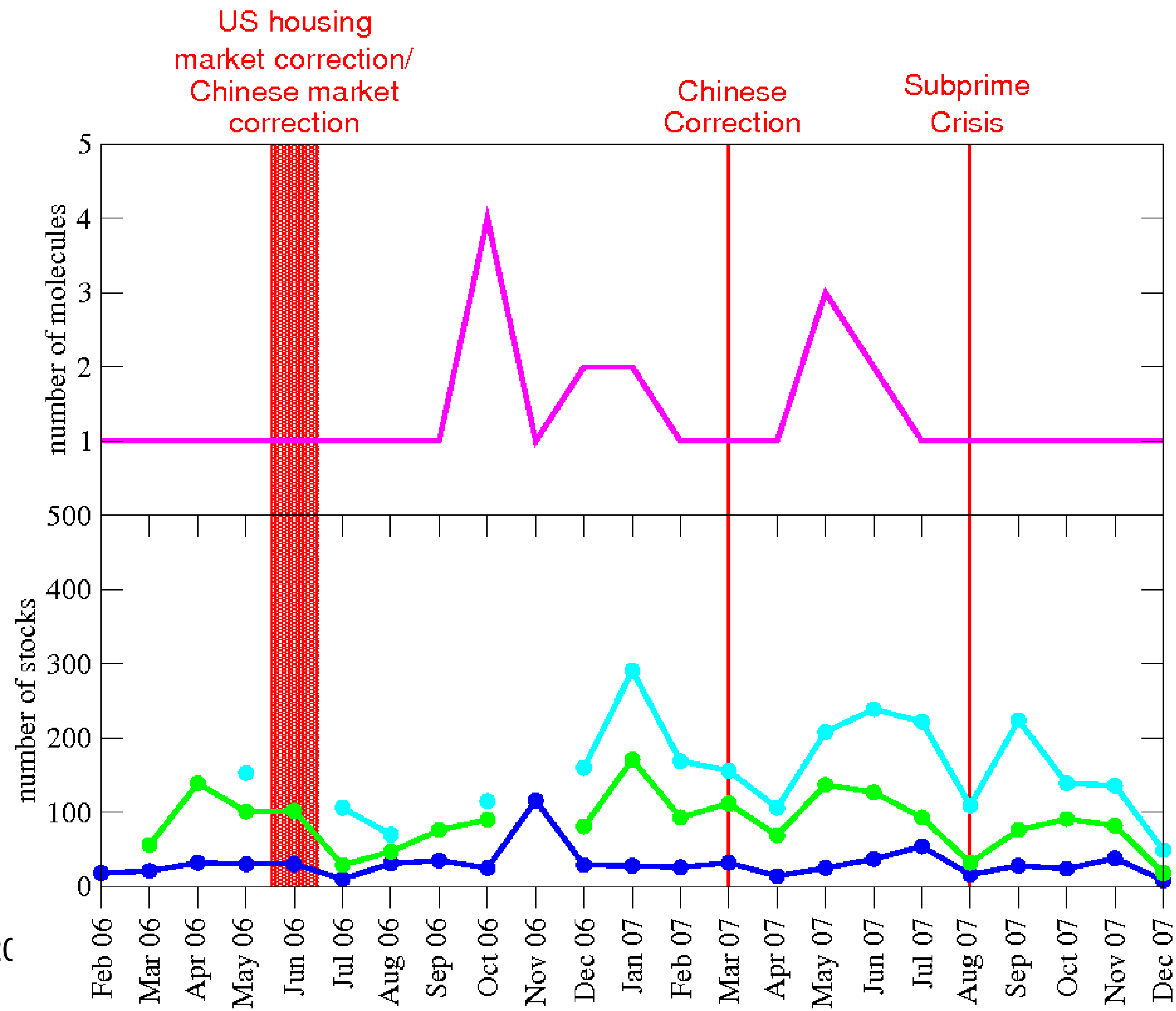
SGX1	SGX2	SGX3	SGX4	SGX5	SGX6
Singtel	Singapore Airlines	Celestial Nutri-foods	Mirach Energy	CapitaLand	DBS Gp Hldg
Singtel 10	Singapore Airlines 200	China Sun Bio-chem Tech Gp	Sky China Petroleum Svcs	City Develop-ment	United Overseas Bank
Singtel 100			Ferrochina		Overseas Chinese Banking Corp
			China Sky Chemical Fibre Co		Wing Tai Hldg

SGX Financial Molecule

SGX financial molecule 2006-2007



Dynamics of Effective Variables



Segmentation vs Clustering

- Time Series Clustering
  - Discover effective mesoscopic variables in given time window
  - Discover slow time evolution of effective variables by sliding time window
- Time Series Segmentation
  - Discover number/type of macroscopic phases
  - Discover lifetimes of macroscopic phases
  - Discover time scales of transitions between macroscopic phases

Modeling Nonstationary Time Series

- Assume non-stationary time series
  - $\mathbf{x} = (x_1, x_2, \ldots, x_N)$
  - M stationary segments
  - In segment m, data points drawn from (μ_m, σ_m^2) Gaussian distribution
- Recursive segmentation
  - One time series → two segments
  - Each segment → two subsegments
  - Iterate + optimize
  - Terminate

Jensen-Shannon Divergence

- Single-segment likelihood for **x** = (x_1, x_2, …, x_N)

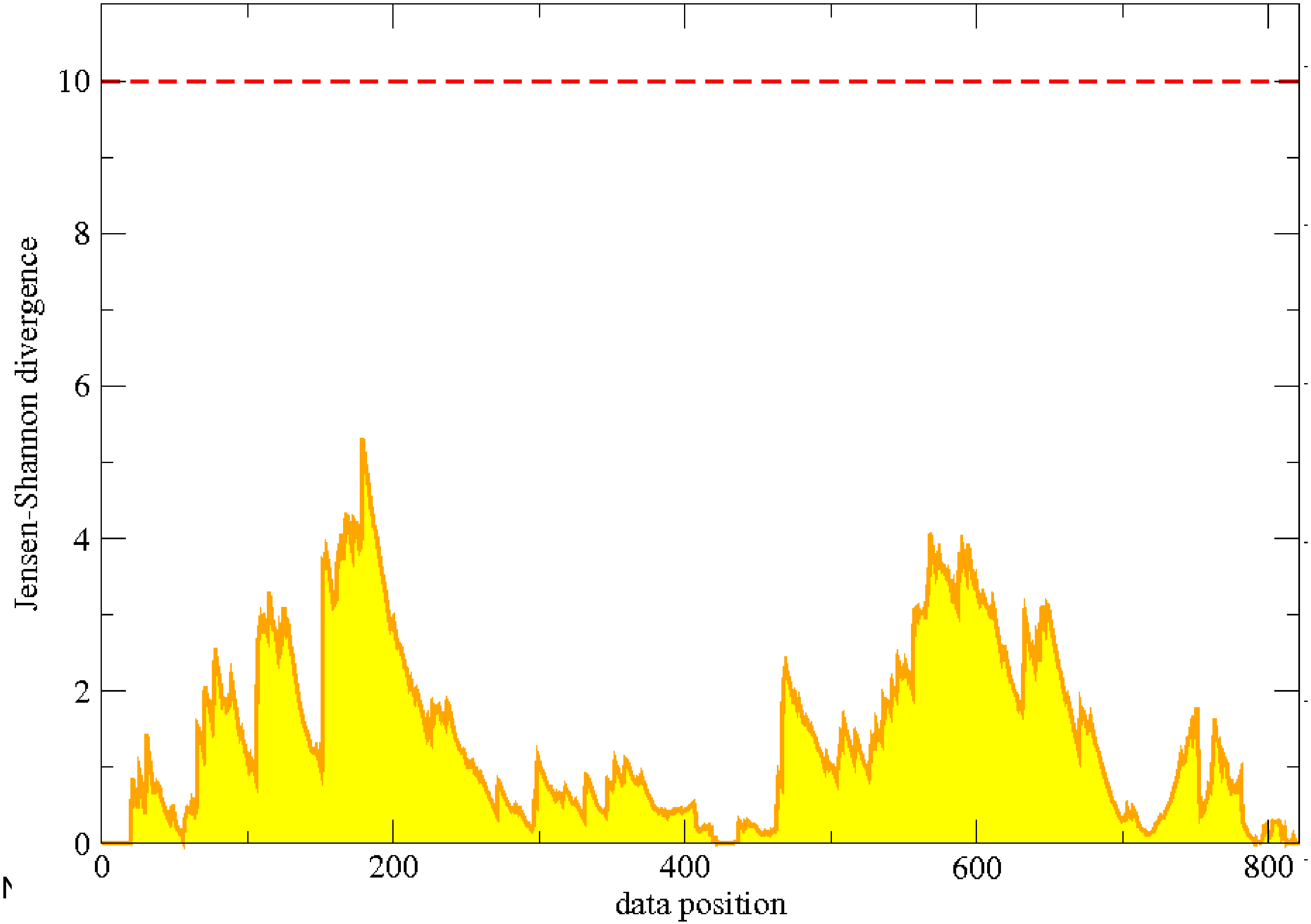
$$L_1 = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x_i - \mu)^2}{2\sigma^2} \right]$$

- Two-segment likelihood for **x** = (x_1, …, x_t, x_{t+1}, …, x_N)

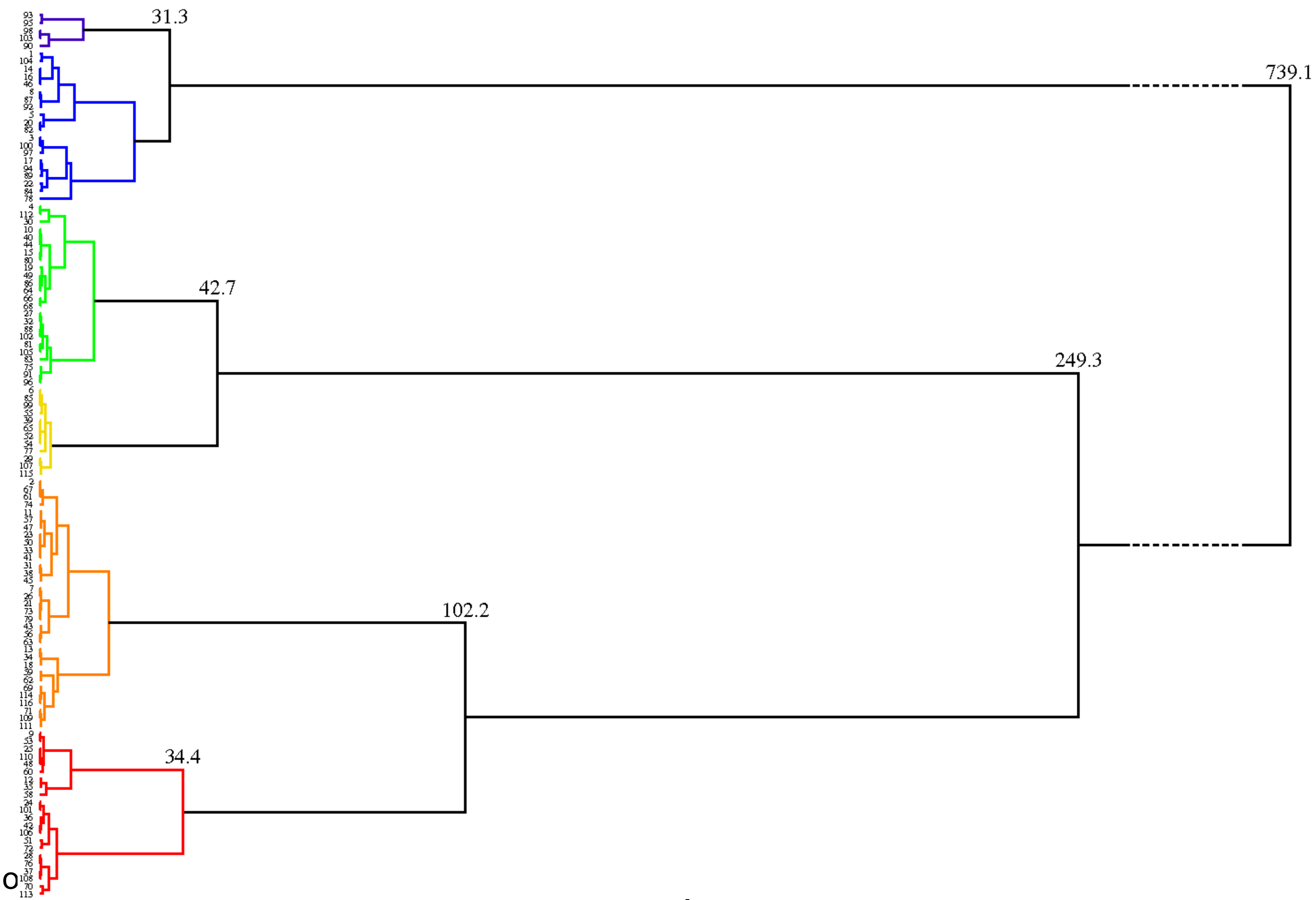
$$L_2(t) = \prod_{i=1}^{t} \frac{1}{\sqrt{2\pi\sigma_L^2}} \exp\left[-\frac{(x_i - \mu_L)^2}{2\sigma_L^2} \right] \prod_{i=t+1}^{N} \frac{1}{\sqrt{2\pi\sigma_R^2}} \exp\left[-\frac{(x_i - \mu_R)^2}{2\sigma_R^2} \right]$$

- ML estimates $\hat{\mu}, \hat{\mu}_L, \hat{\mu}_R, \hat{\sigma}^2, \hat{\sigma}_L^2, \hat{\sigma}_R^2$
- Jensen-Shannon divergence $\Delta(t) = \ln \frac{L_2(t)}{L_1} \geq 0$

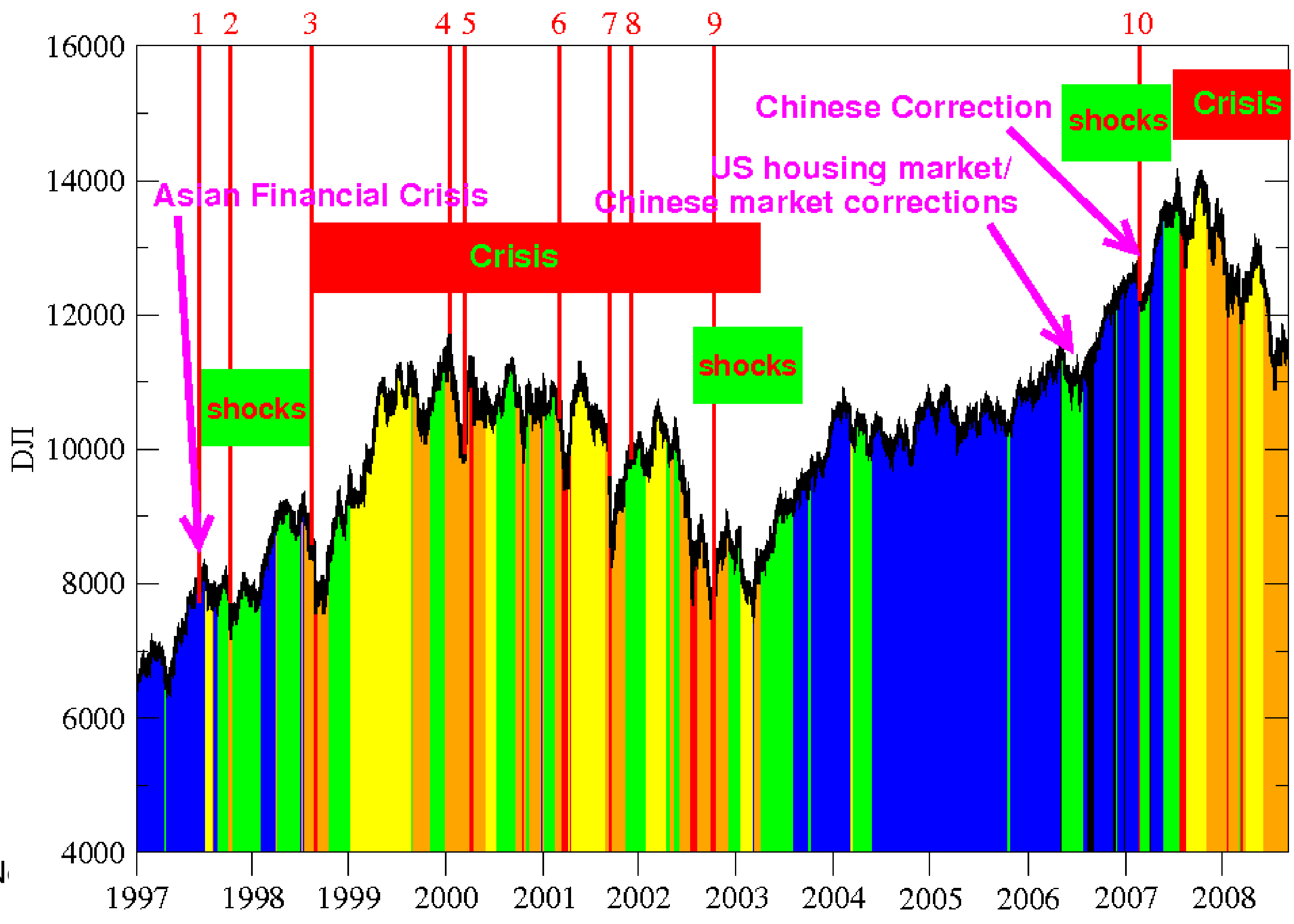
Recursive Segmentation



Segment Clustering



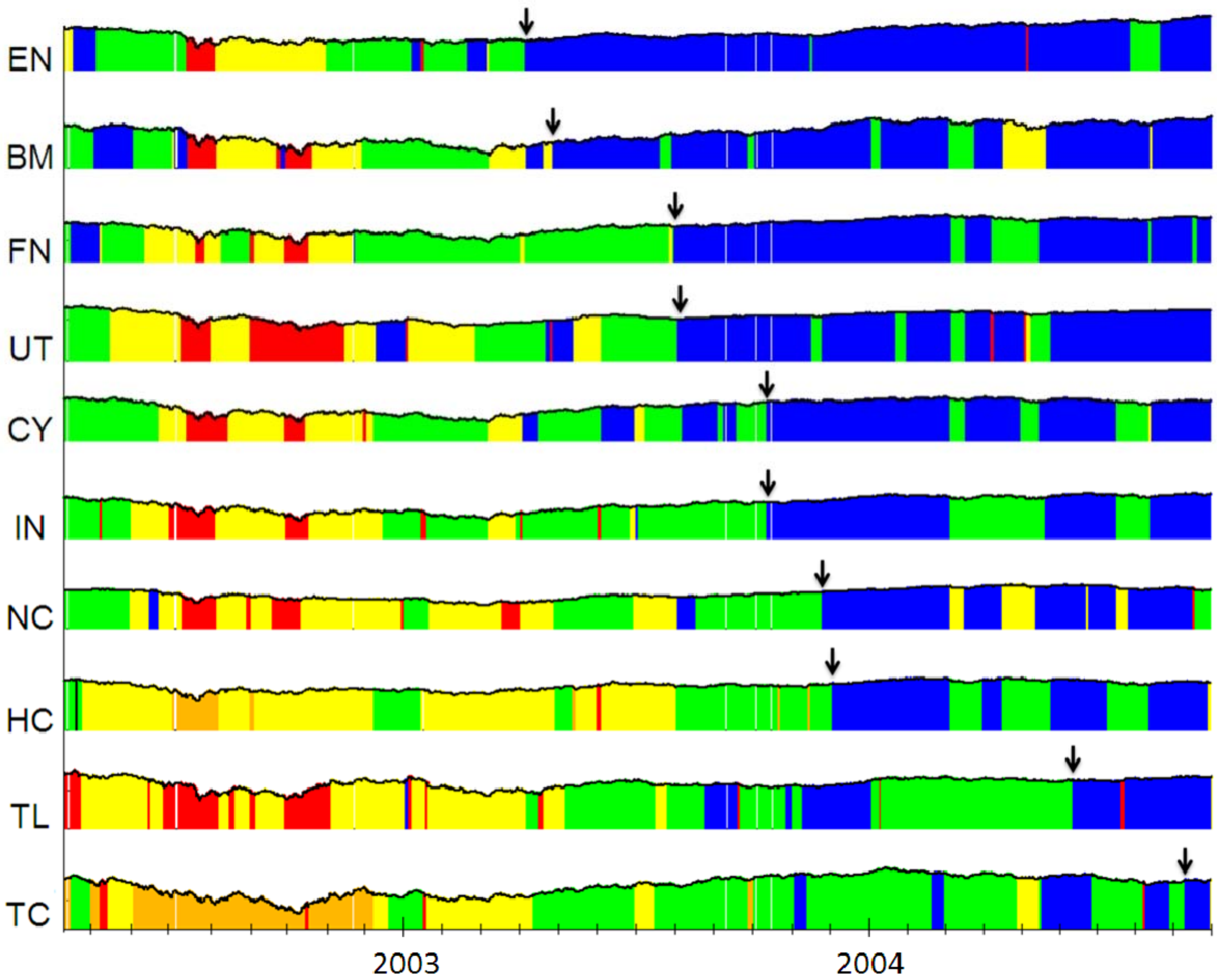
Temporal Distribution of Segments



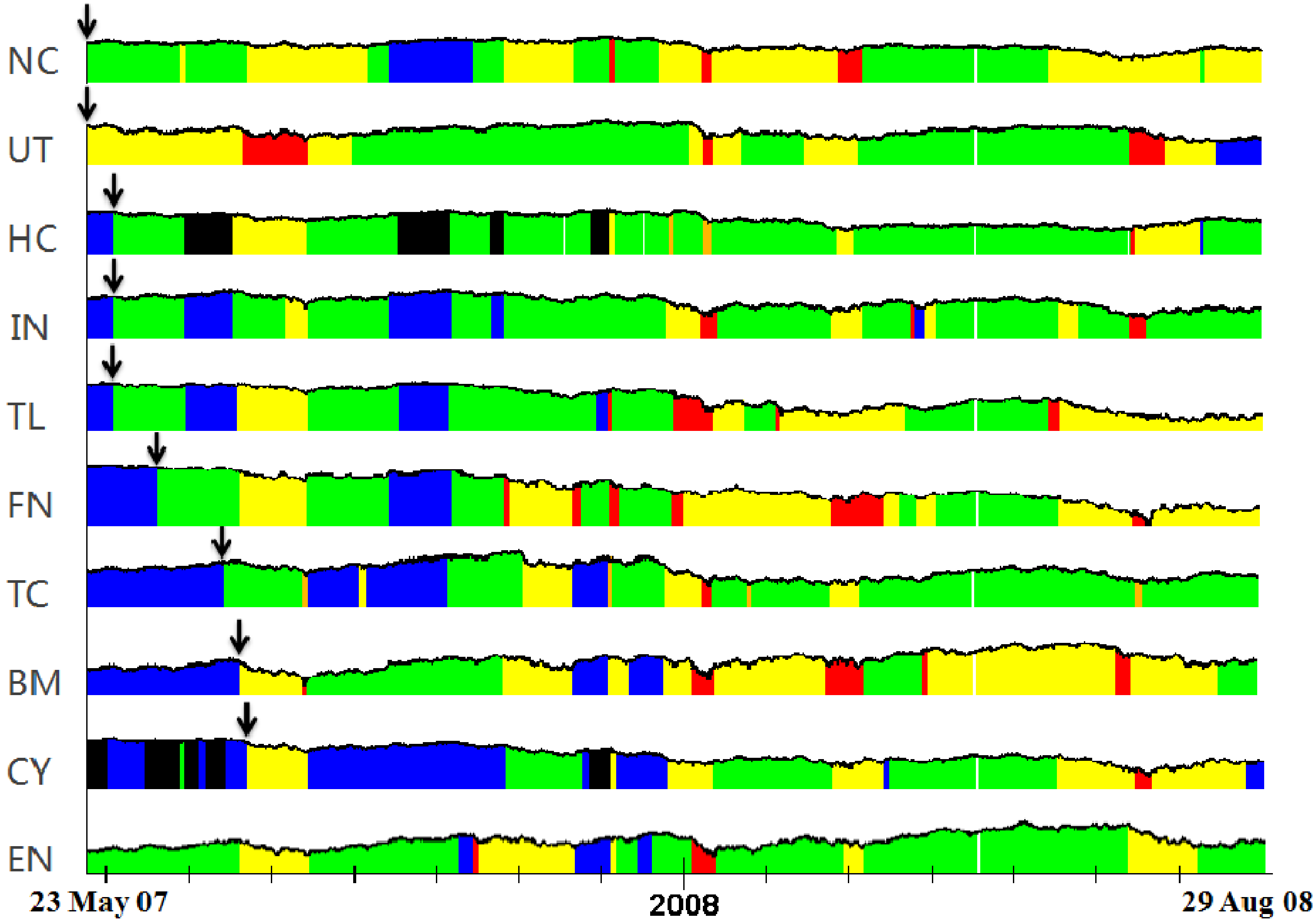
Cross Section Study

k	***Symbol***	***Economic Sector***
1	BM	Basic Materials
2	CY	Consumer Services
3	EN	Oil & Gas
4	FN	Financials
5	HC	Healthcare
6	IN	Industrials
7	NC	Consumer Goods
8	TC	Technologies
9	TL	Telecommunications
10	UT	Utilities

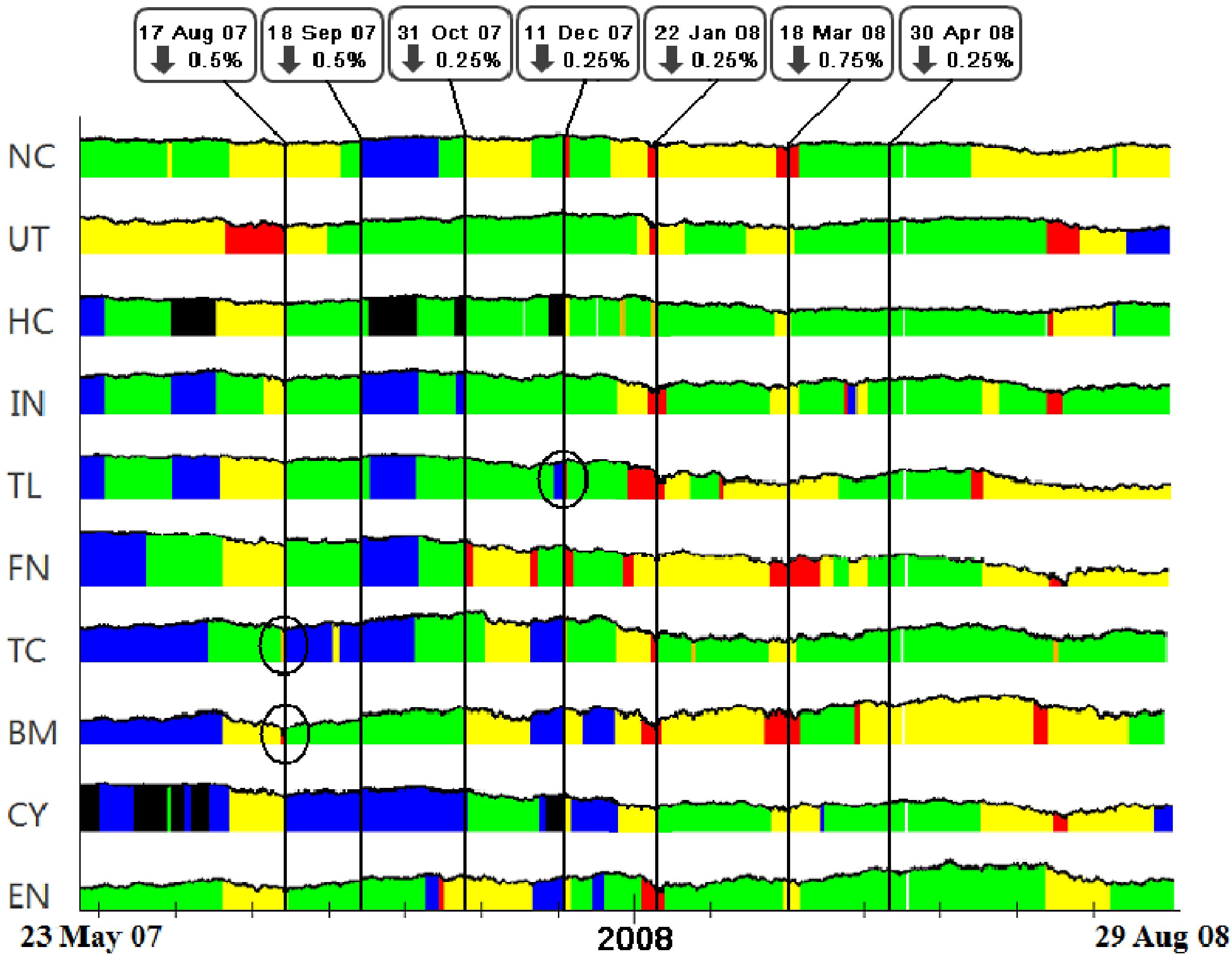
Economic Recovery



Onset of Crisis



Response to Extrinsic Shocks



Conclusions

- Study macroscopic phases and their dynamics from microscopic time series
- Time series clustering
  - Metallic Nanoclusters
    - Slow effective variables
    - Temperature dependence and melting
  - SGX
    - Financial atoms and molecules
    - Chemical picture of market crashes
- Time series segmentation
  - US economy
    - Crisis and growth
    - Chemical picture of slow dynamics using MSTs

Acknowledgments

- Time Series Clustering
  - Prof Lai S K
  - GOO Yik Wen
  - LIAN Tong Wei
  - ONG Wei Guang
  - CHOI Wen Ting
  - Mridula DAMODARAN
- Time Series Segmentation
  - WONG Jian Cheng
  - Gladys LEE Hui Ting
  - ZHANG Yiting
  - Dr Manamohan PRUSTY

Thank You!