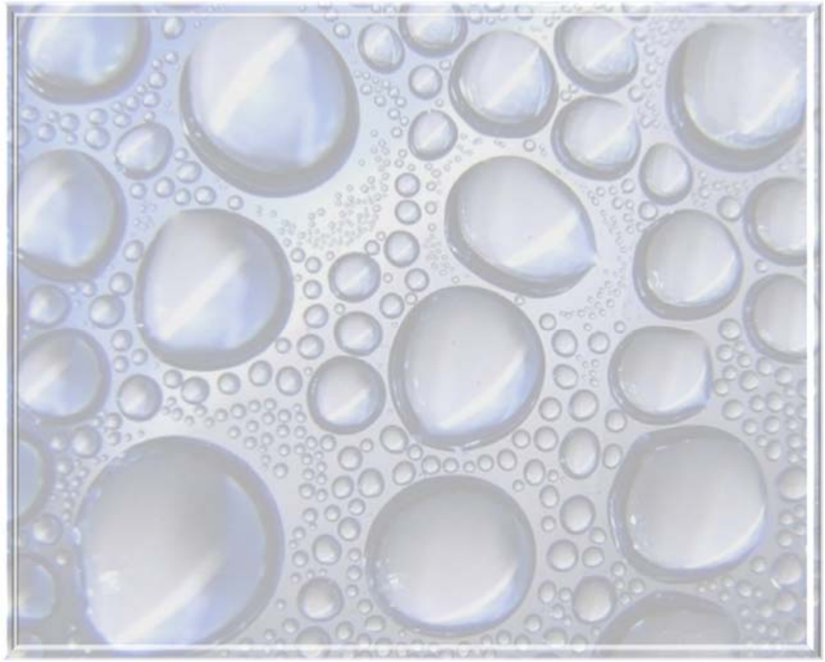
Design of Future Integrated Systems: A Cyber-physical Systems Approach



Radu Marculescu
Carnegie Mellon University
Pittsburgh, PA 15213, USA

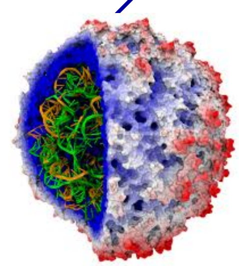
PATMOS-VARI, Karlsruhe
Sept 9, 2013

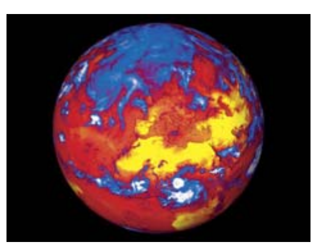
Carnegie Mellon

Integrated systems are critical to IT revolution

Embedded computing (cell phones, automotive, gaming)

Cyber-physical computing (healthcare, transportation, smart grid)

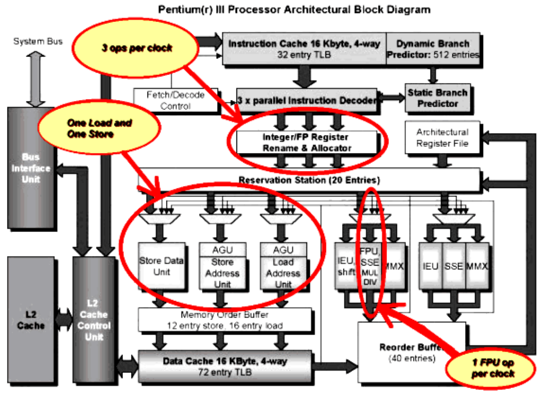




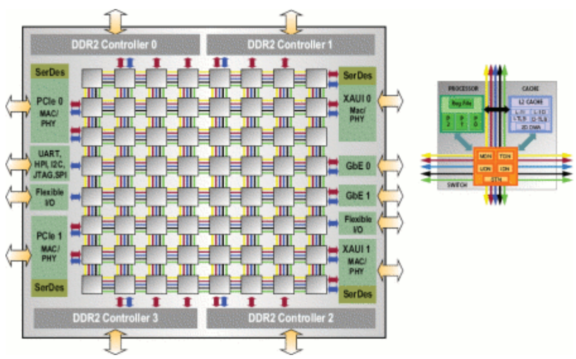
High performance computing (scientific applications, forecasting, data centers)

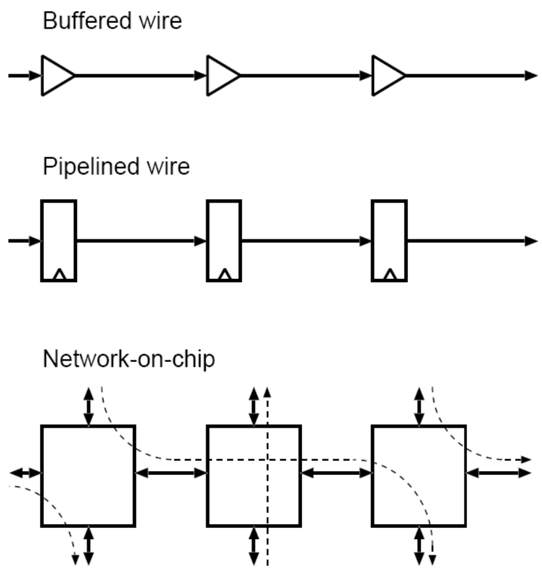
Three paradigm shifts, i.e. low-power (~90s), network-centric (~2K), and cyber-physical design (~2010). Exciting times ahead…

Low-power Design. Computation vs. communication



The Intel Pentium !!!

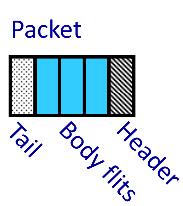


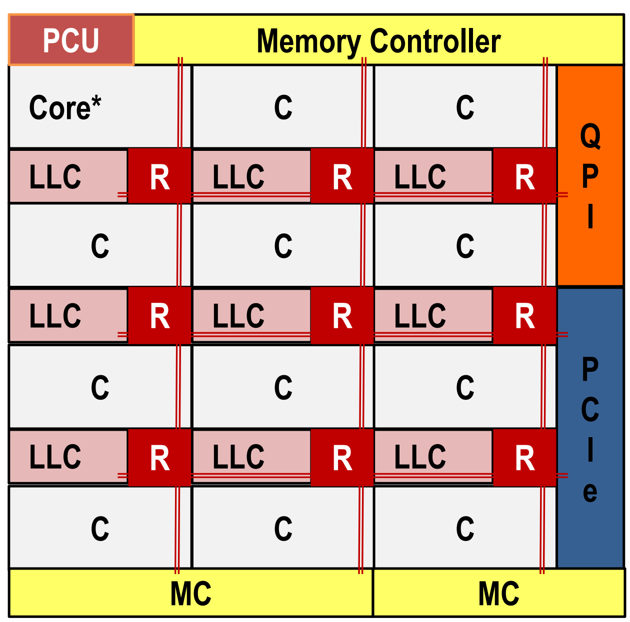


[D. Greenfield, S. Moore, Computer J., 2008]

Multicore platforms are large scale distributed systems at nanoscale; they are dominated by communication costs

- Last level cache (LLC)
- Memory controller (MC) & channels
- I/O controller(s)
- QPI controller,
- Power control unit (PCU), etc.

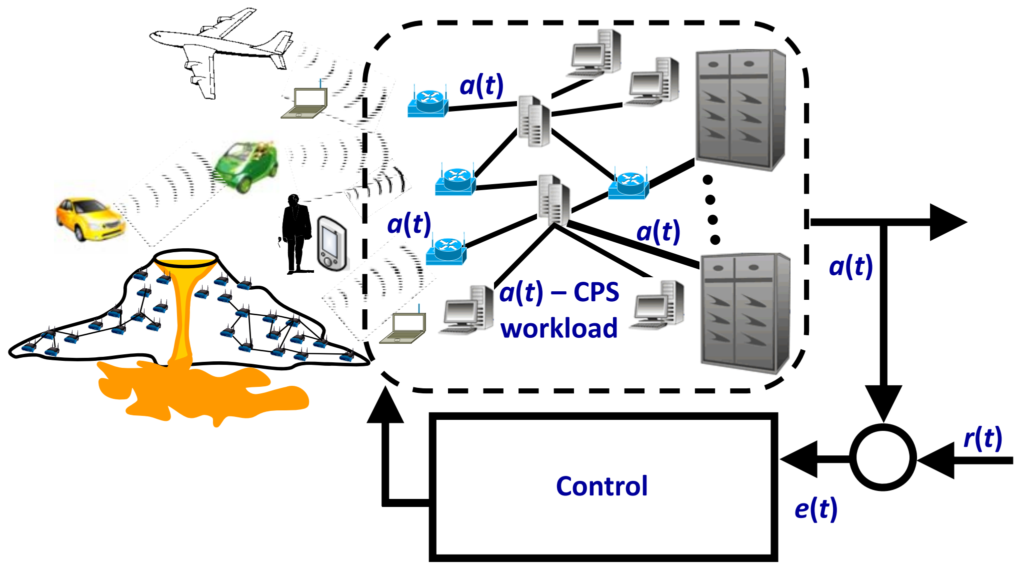




[U. Ogras, Tutorial ASPLOS 2012]

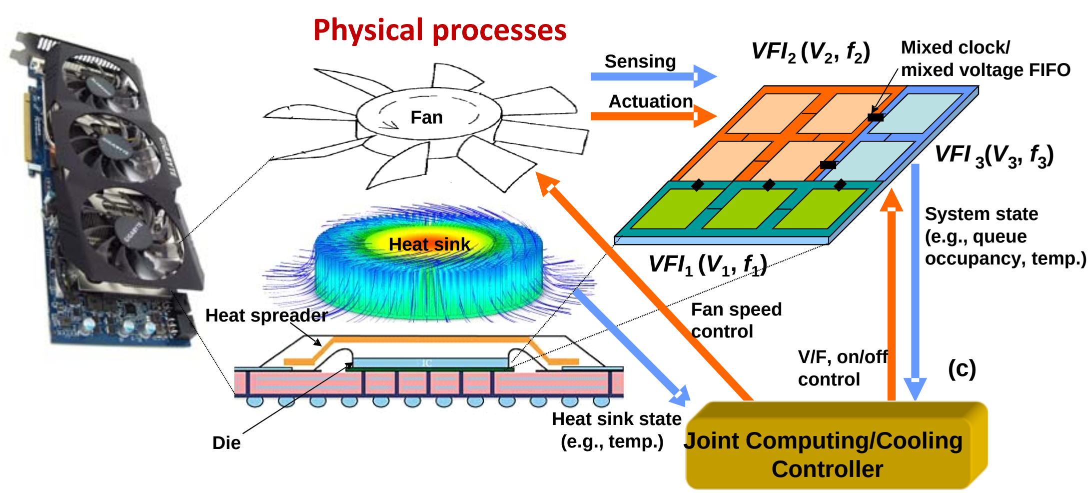
Need to understand the behavior of thousand core systems.
Network (routers+links) is the missing link in understanding.

CPS consist of networks of computational devices and networks that monitor and control physical processes



Multi-scale behavior has been observed in many technological, economical, and biological systems

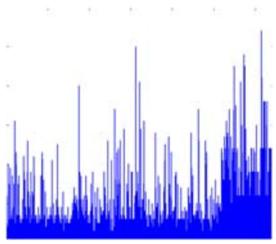
MPSoCs as CPS: Current multicore platforms bring together computation, communication, and control



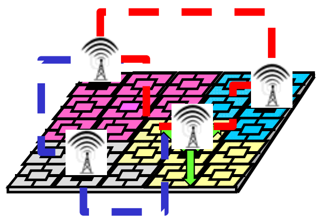
Processing cores interconnected via routers and links; they operate at various voltages/frequencies thus saving power

This presentation focuses on the new CPS paradigm and its impact on future integrated systems

Structure
Architecture and small world effects

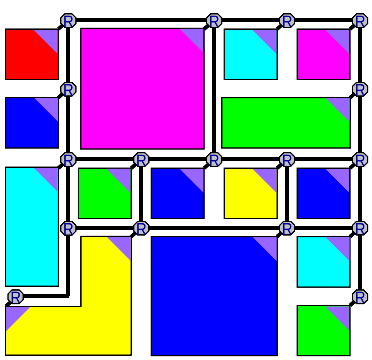


Dynamics
Workloads and multiscale behavior

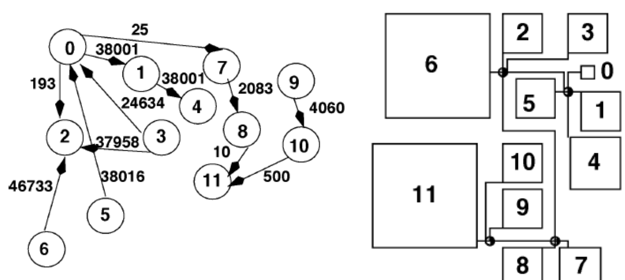


Control
Power and resource management

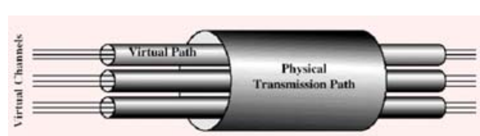
Our first insight into communication-based design came through architecture (topology, buffer, etc.) optimization

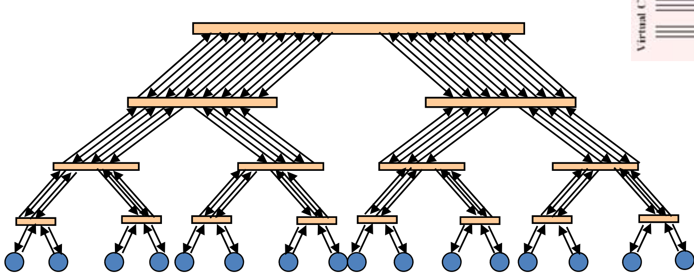


[Bolotin et al. DATE 2007]

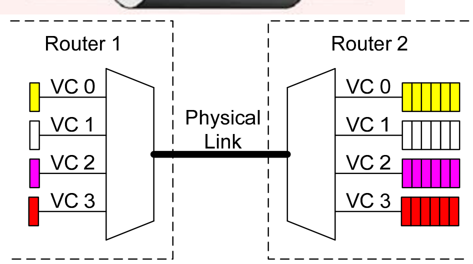


[K. Srinivasan et al. IEEE T-VLSI 2006]

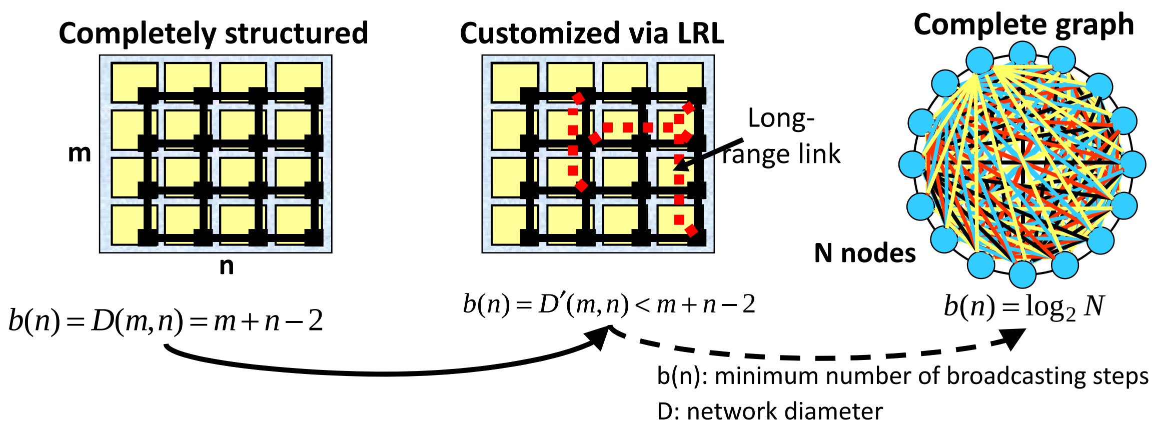




[J. Kim et al. ISCA 2007]

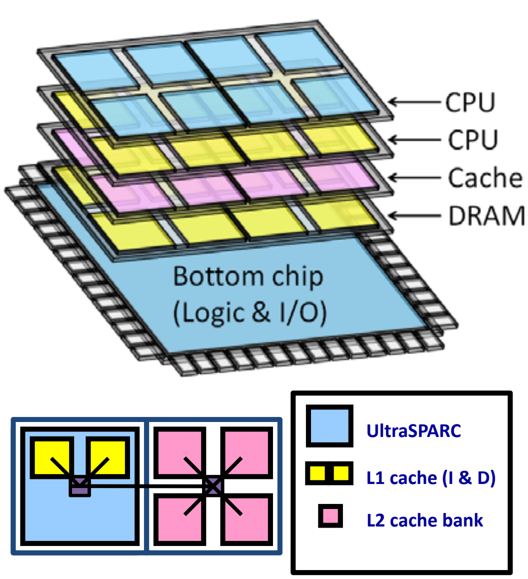


Can induce small-world effects in regular NoCs. This brings huge performance improvements

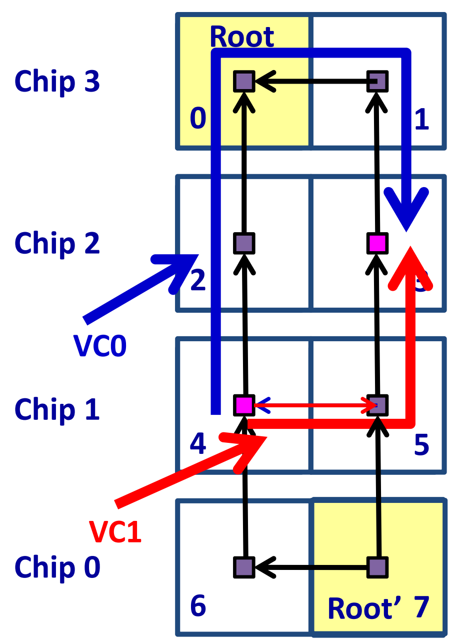


This way, the fundamental idea of Small World networks (aka "six degrees of separation") enters the multicore world

Small world effects can also be exploited to reduce hop count in 3D wireless NoCs and improve performance

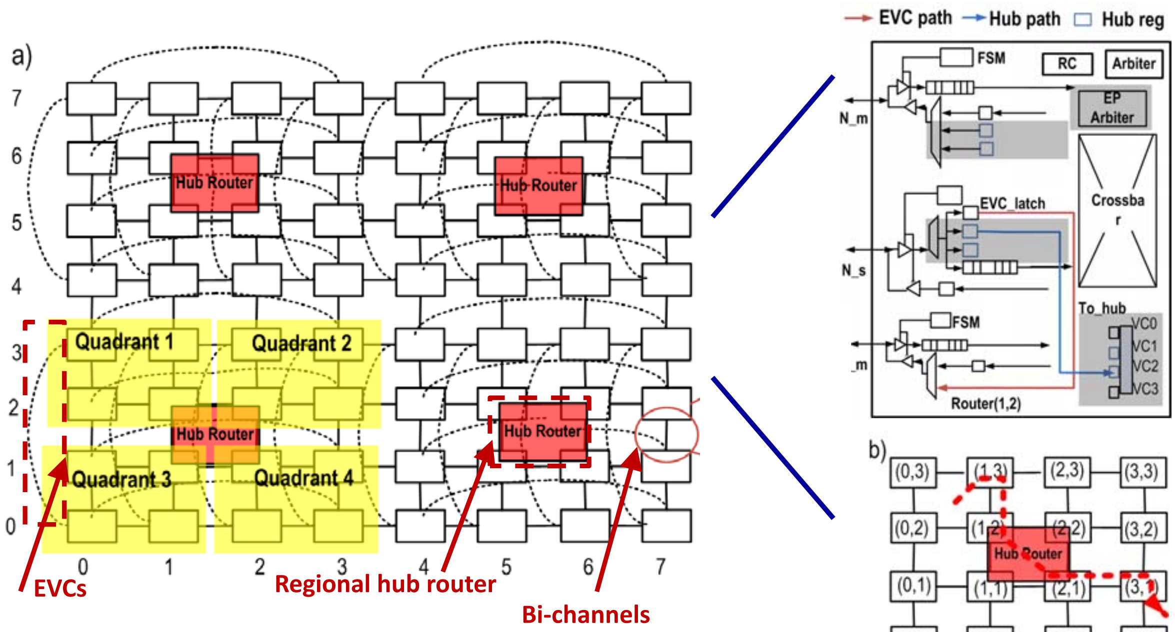


[H. Matsutani et al. ASPDAC 2013]



Wired and wireless NoCs can be used intra-chip, while inter-chip communication is based on wireless inductive-coupling.

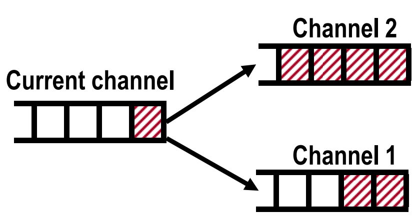
Flow-control mechanisms, reconfigurability, adaptive routing have all been used to improve performance



[Q. Zhiliang et al. CODES-ISSS 2012]

One way or another, they all exploit the small world effects…

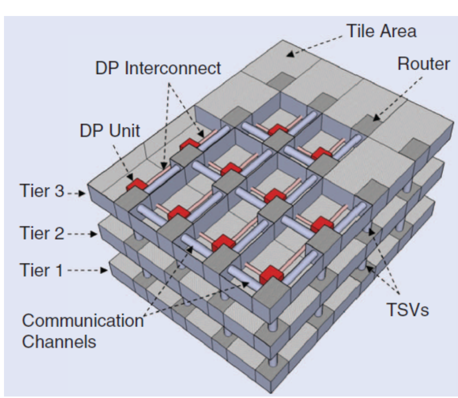
Optimum path-seeking behavior may be goal-oriented. Need to collect fitness information locally, at routing nodes

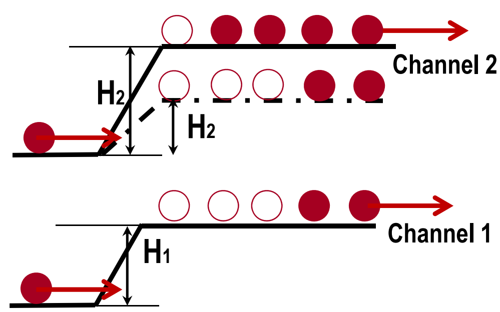


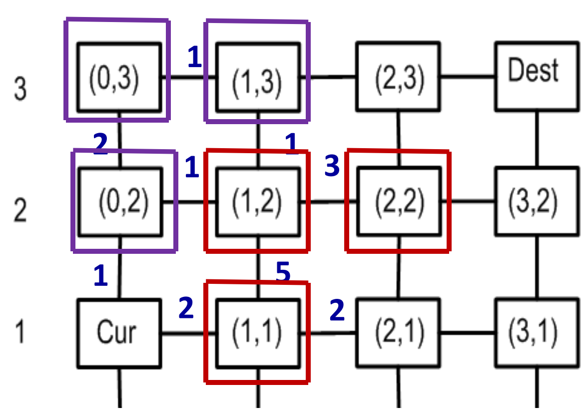
$$fitness_i = \alpha \cdot m_{in} \cdot w_{in}^{-1}$$

m = avg. number of free slots

w = avg. waiting time in channel



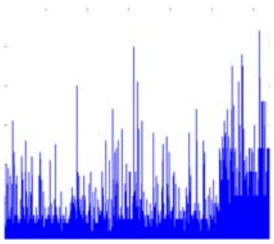




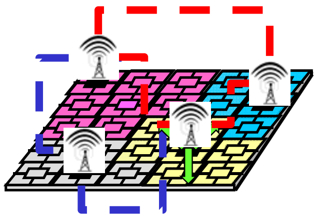
[T. Mak et al. IEEE Cir&Sys Mag.2011]

This presentation focuses on the new CPS paradigm and its impact on future integrated systems

Structure
Architecture and small world effects

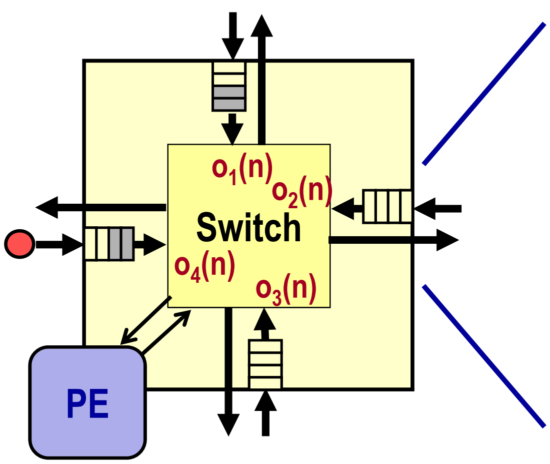


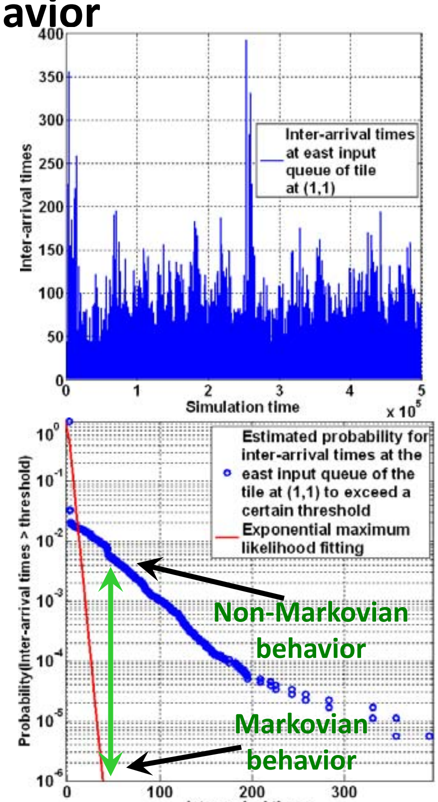
Dynamics
Workloads and multiscale behavior



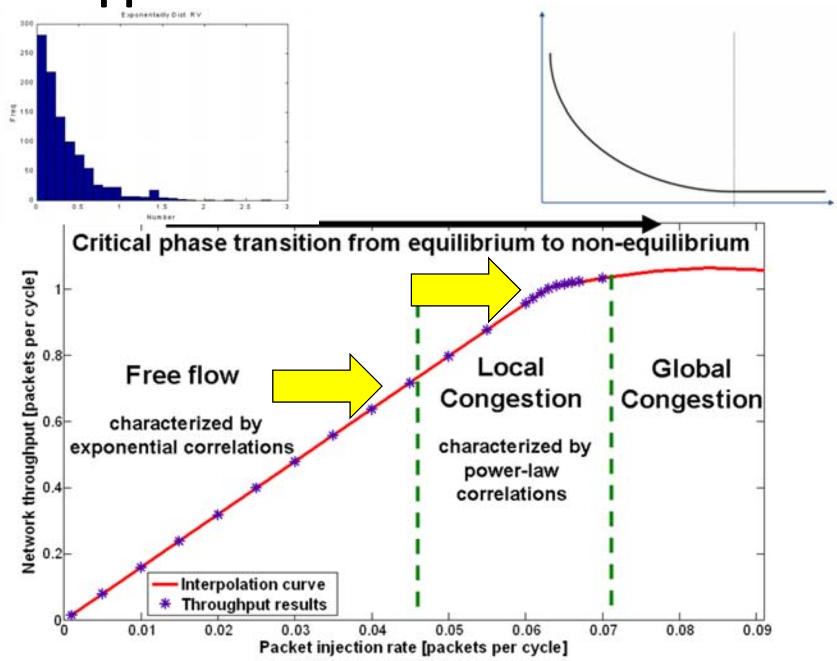
Control
Power and resource management

Packet inter-arrival times at interface queues play a fundamental part in network behavior





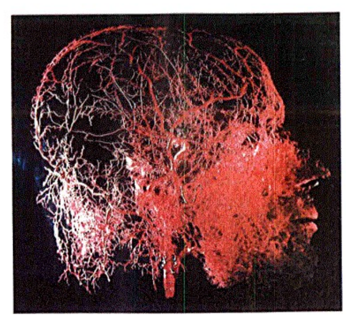
Network behavior at high injection rates needs a non-equilibrium approach that accounts for fractal behavior



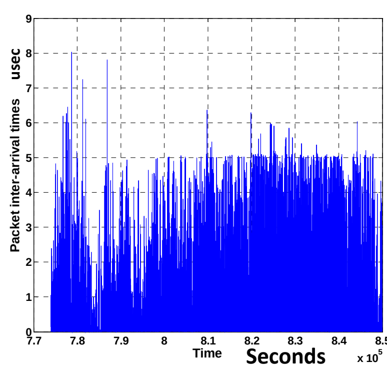
High injection rates cause inter-arrival times deviate from exponential distribution and exhibit power law correlations.

Fractals are geometrical objects or stochastic processes displaying self-similar behavior over multiple scales

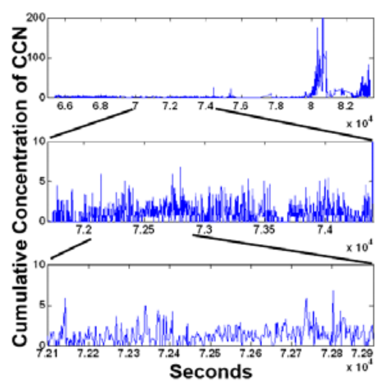
Space



Time

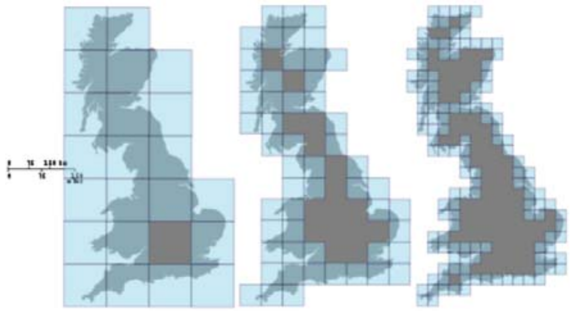


Bellcore traffic



How long is the coast of Britain? Answering this question involves statistical self-similarity and fractal dimension

Fractal dimension can be computed via box counting



$$N(R) \approx R^{-D}$$

Choose an increasing set *R* of edge lengths

For each size r_i in *R*

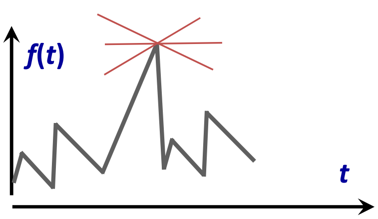
- Super-impose a series of distinct squares (boxes) over the *data*
- Count the minimum # of boxes needed to cover data and store it in vector *N*

Compute fractal dimension by fitting the equation between *R* and *N*

Box counting method can be applied to determine the fractal dimension of 1D, 2D , or 3D vectors

Real world processes are not always smooth. Fractional dynamics needs a formalism stronger than integer calculus

Fractional behavior

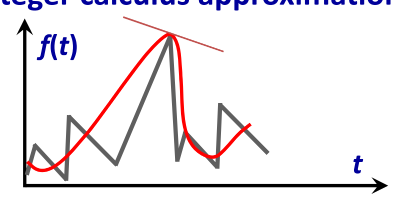


Mandelbrot (1975)

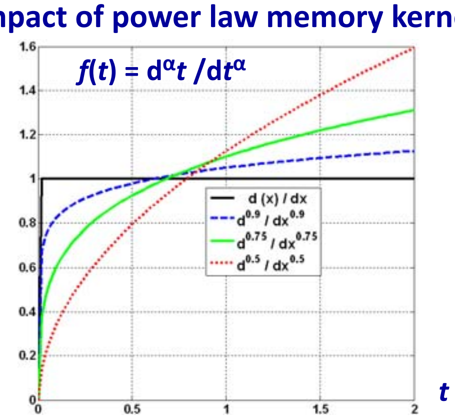
$$\langle \Delta x \rangle \propto \Delta t^{H}, 0 < H < 1$$

$$\frac{d^{\alpha} x(t)}{dt^{\alpha}} = \lim_{\Delta t \to 0} \frac{1}{\Delta t^{\alpha}} \sum_{j=0}^{[t/\Delta t]} (-1)^{j} \binom{\alpha}{j} x(t - j\Delta t)$$

Integer calculus approximation



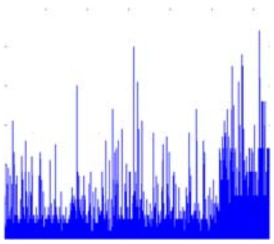
Impact of power law memory kernel



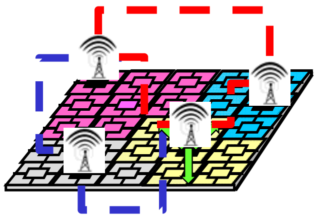
For CPS, time is an intrinsic component of the programming model. Consequently, system dynamics becomes essential.

This presentation focuses on the new CPS paradigm and its impact on future integrated systems

Structure
Architecture and small world effects

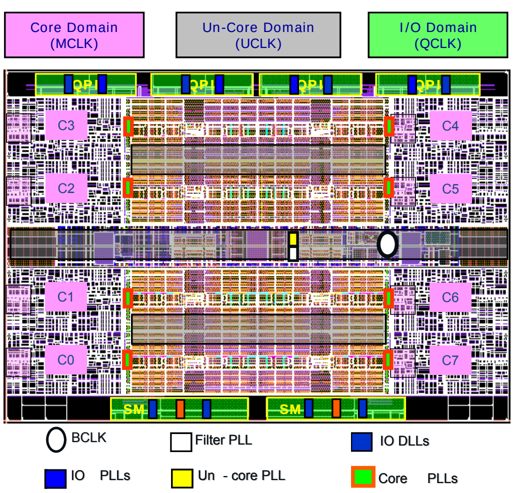


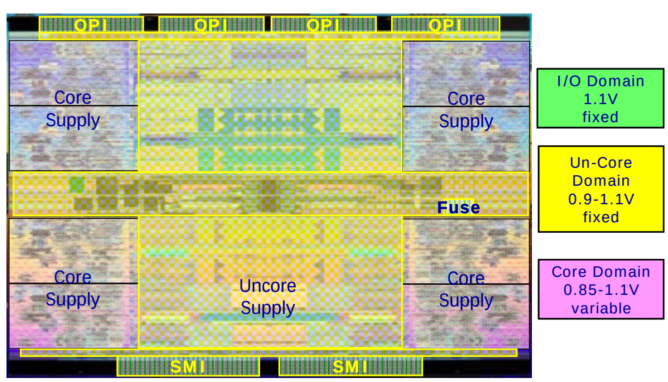
Dynamics
Workloads and multiscale behavior



Control
Power and resource management

8-Core Xeon® Processor has three clock domains and three voltage domains that help minimizing power.

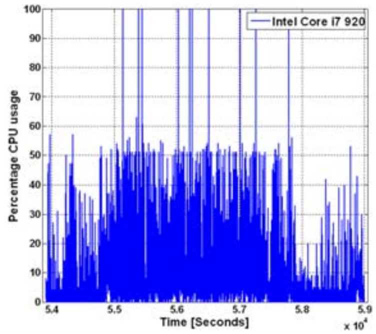


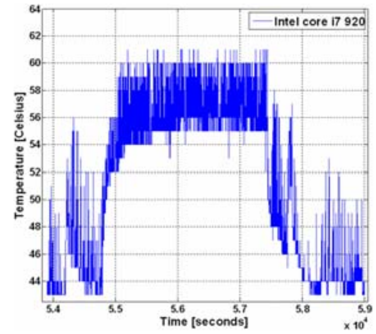


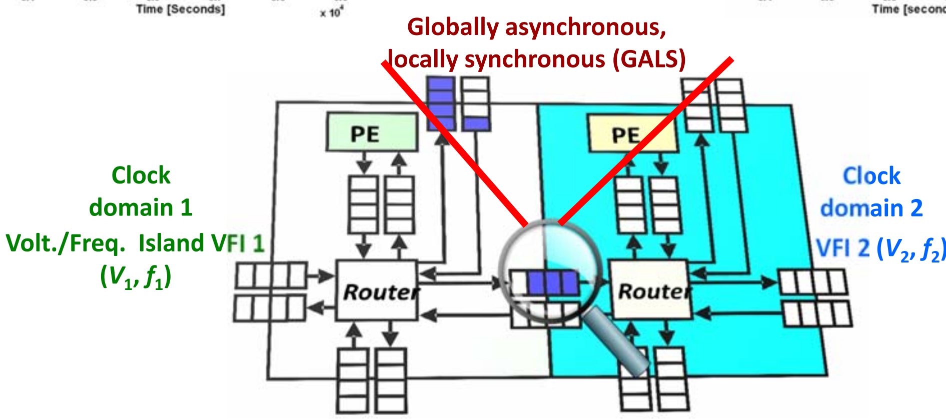
[Rusu, ISSCC 2009]

Device count reported 2.3B transistors. The chip has 9 temperature sensors, one in each core hot spot and one in the die center.

Fine-grain power management becomes possible by exploiting workload variations





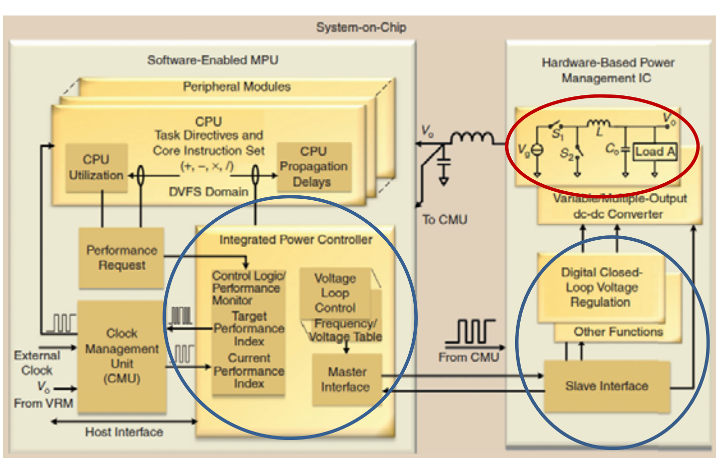


Minimizing energy consumption

Total energy consumption to be minimized

$$\underbrace{E_{Total} = E_{App}}_{\text{Application (useful) energy consumption (comp+comm)}} + \sum_{i=1}^{n \;(\#\text{VFIs})} \underbrace{E_{VFI}(i)}_{\text{Overhead of } i^{th} \text{ VFI}}$$

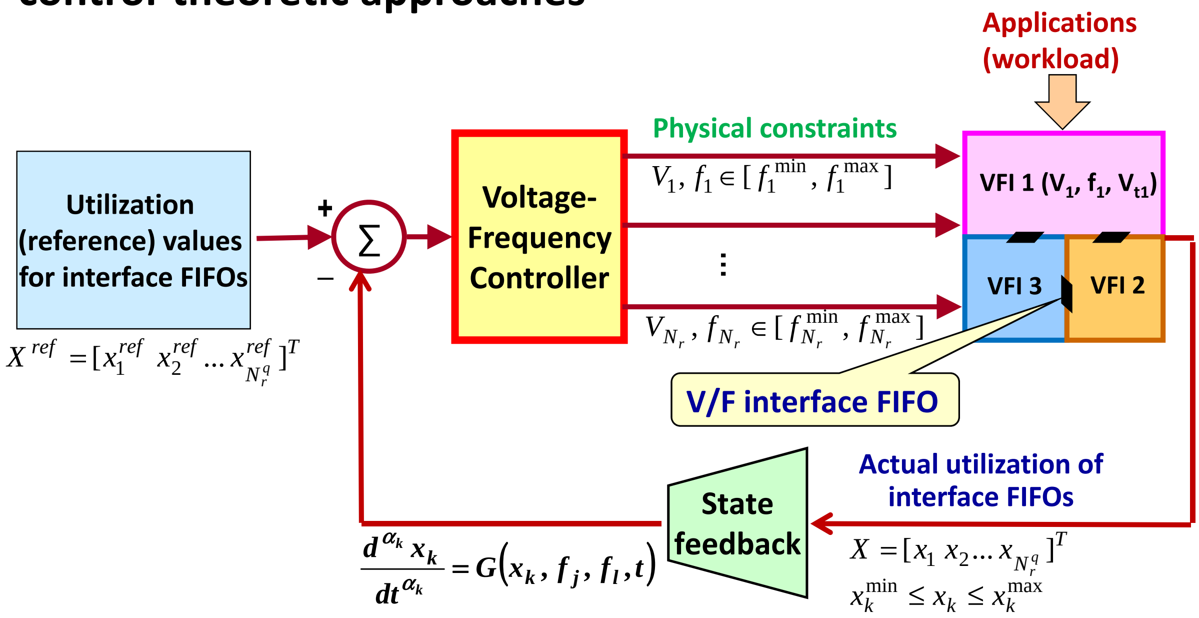
$$\boxed{E_{VFI} = E_{ClkGen} + E_{Vconv} + E_{MixClkFifo}}$$



- Integrated power controller – independent of CPUs
  - Single inductor multiple output (SIMO) converters
  - Adaptive-output converters
  - Voltage regulator controller (VRC) on Intel SCC

[D. Ma, Cir&Syst Mag, 2010]

Fine-grain power management can be implemented via control-theoretic approaches



Inter-arrival times dictate the fractal exponent (α_k) of state equations; this is used to characterize the system dynamics.

V/F controller selects the minimum operating frequencies s.t. the queues reference values are satisfied

Computational workload

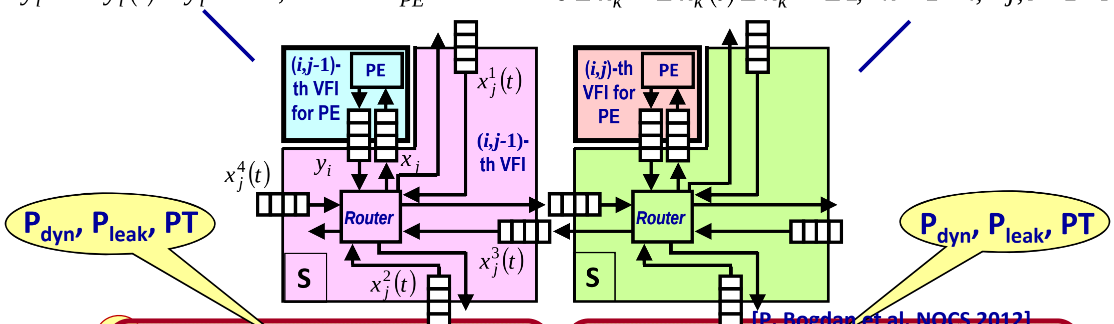
$$\frac{d^{\alpha_i} y_i(t)}{dt^{\alpha_i}} = a_i(t) y_i(t) + b_i(t) f_i(t) - c_i(t) f_j(t),$$

$$0 \le y_i^{\min} \le y_i(t) \le y_i^{\max} \le 1, \quad i = 1 \div N_{PE}$$

Traffic variability **Write/Read**

$$\frac{d^{\alpha_k} x_k(t)}{dt^{\alpha_k}} = a_k(t) x_k(t) + b_k(t) f_j(t) - c_k(t) f_l(t),$$

$$0 \le x_k^{\min} \le x_k(t) \le x_k^{\max} \le 1, \quad k = 1 \div 4, \quad j, l = 1 \div N_r$$



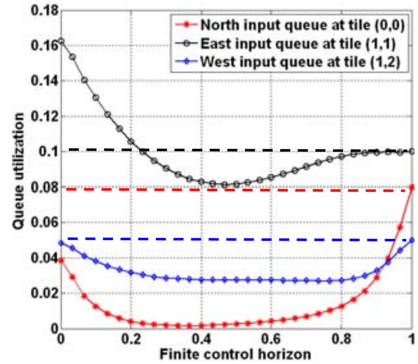
[P. Bogdan et al. NOCS 2012]

$$\min \int_{t_i}^{t_f} \left\{ \sum_{i=1}^{N_{PE}} \left[\frac{1}{2} w_i \left(y_i(t) - y_i^{ref} \right)^2 + \frac{1}{2} z_i f_i(t)^2 \right] + \sum_{j=1}^{N_r} \left[\frac{1}{2} r_j f_j(t)^2 + \sum_{k=1}^{4} \left[\frac{1}{2} q_k \left(x_j^k(t) - x_k^{ref} \right)^2 \right] \right] \right\} dt$$

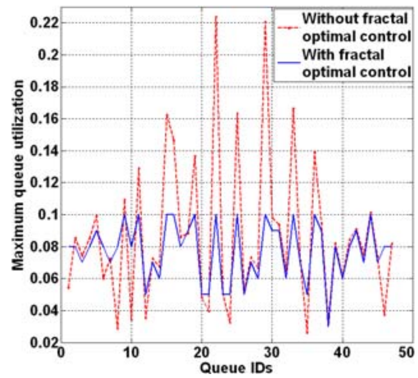
$$f_i^{\min} \le f_i(t) \le f_i^{\max}, \quad i = 1 \div N_{PE}, \quad and \quad f_j^{\min} \le f_j(t) \le f_j^{\max}, \quad j = 1 \div N_r$$

Accurate mathematical modeling and rigorous optimization can enable cross-layer power management

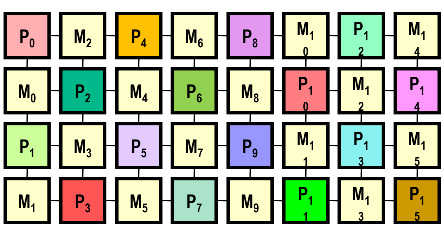
Queues utilization at tiles (0,0), (1,1) and (1,2) for a 4×4 mesh NoC running Apache HTTP webserver application

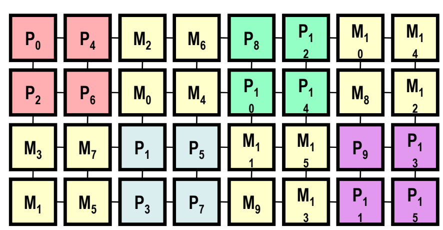


FOC keeps the utilization of all queues below 0.1 by adjusting the operating frequencies of all PEs and routers. FOC consumes less power compared to a LQR controller

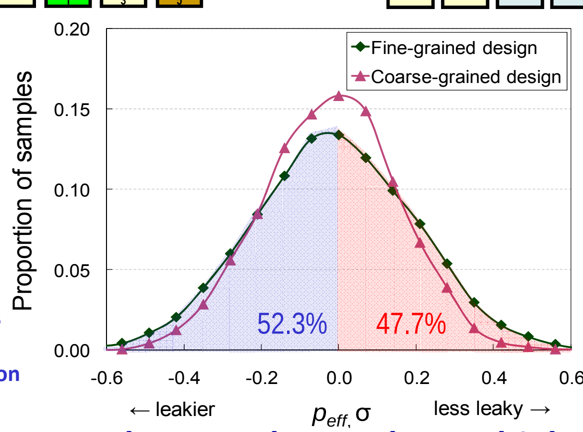


How about core-to-core variability? Use frequency asymmetry to reduce performance loss





Bias towards lower V/F levels
Greater power reduction
Greater performance reduction

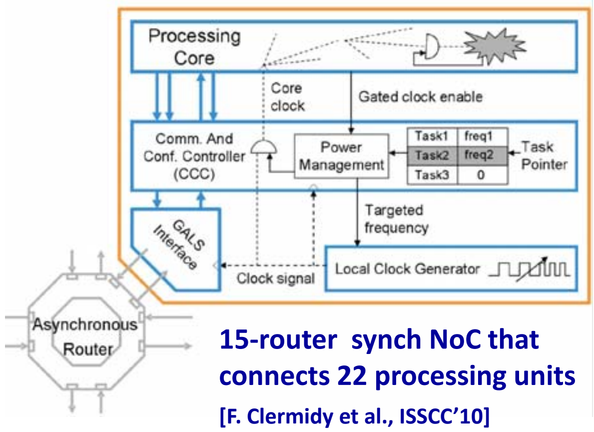


Bias towards higher V/F levels
Smaller power increase
Smaller performance increase

Run leakier cores at lower voltages, but at higher than normal frequencies for those voltages.

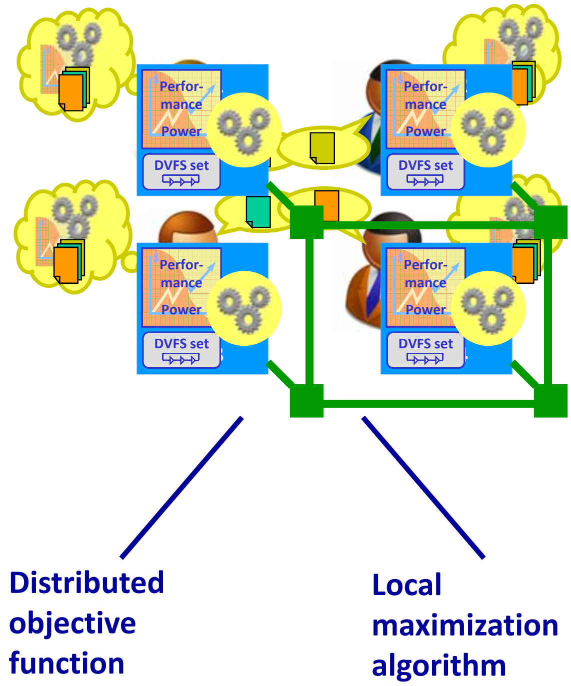
[Herbert et al, IEEE TVLSI 2012]

For thousand core systems distributed approaches for power management are of crucial importance



15-router synch NoC that connects 22 processing units

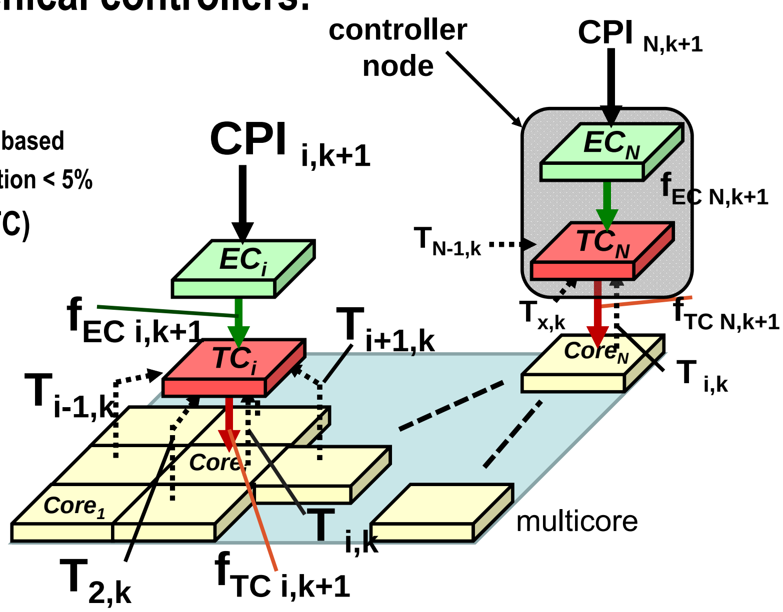
[F. Clermidy et al., ISSCC'10]



Not only a power issue…

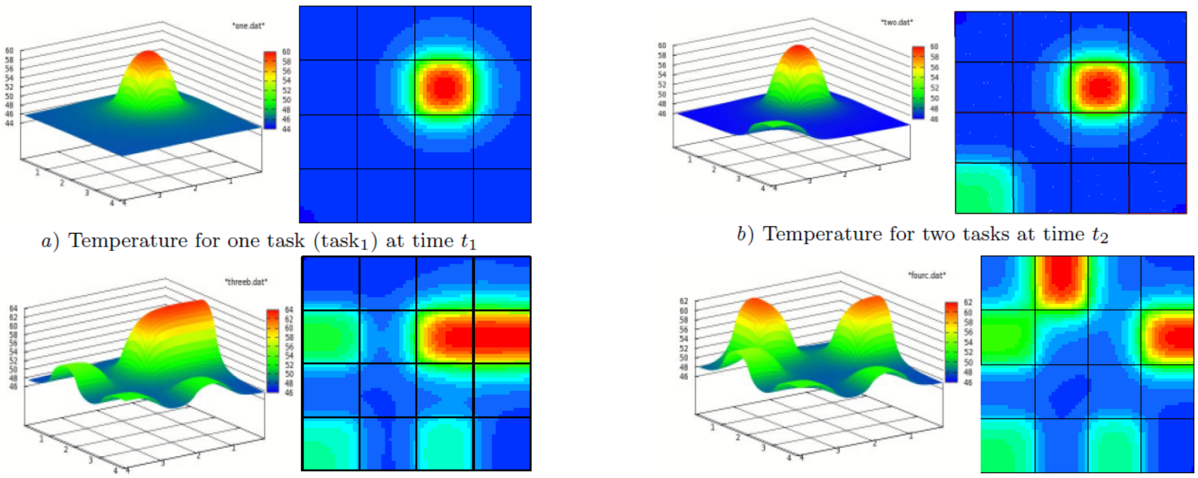
Distributed and hierarchical controllers:

- Energy Controller (EC)
  - Output Frequency f_{EC}
    - Minimize power – CPI based
    - Performance degradation < 5%
- Temperature Controller (TC)
  - Distributed MPC
  - Inputs:
    - f_{EC}, T_{CORE}, $T_{NEIGHBOURS}$
  - Output
    - Core frequency (f_{TC})



[Slide courtesy of L. Benini, Bologna U.]

Peak temperature is important. DTM cannot be achieved by considering power alone. Physical context is crucial…



a) Temperature for one task (task_1) at time t_1

b) Temperature for two tasks at time t_2

c) Five tasks at time t_i

d) Task_1 (re-)mapped at time t_{i+1}

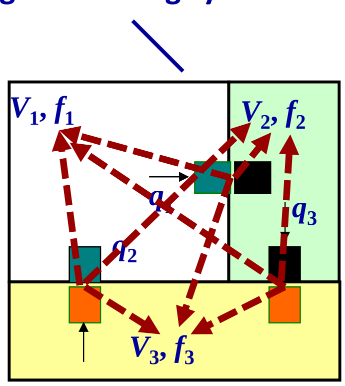
[T. Ebi et al. ICCAD 2009]

$$(sell_n - buy_n) - (sell_i - buy_i) > \tau_n$$

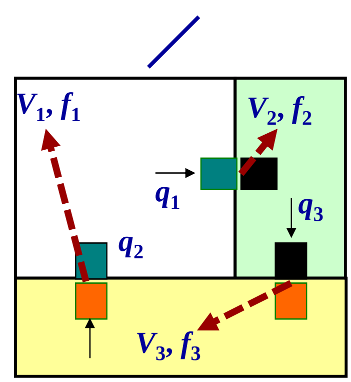
Agent-based systems can help system components (re)configure and optimize their resource usage independently

Both centralized and distributed approaches have their own limitations

Scalability issues due to cost and latency of long wires. Long synchronization times

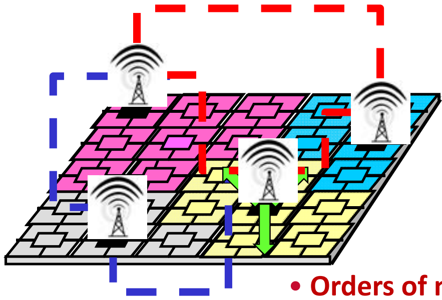


Highly scalable but potential problems in control performance



Need 'best-of-both-worlds' between fully-centralized and fully-distributed solutions. This is true for thermal management too.

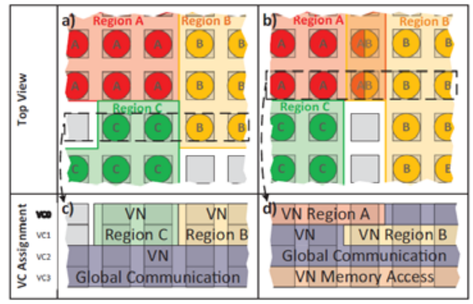
An hierarchy of globally distributed locally centralized control may help the system self-organize

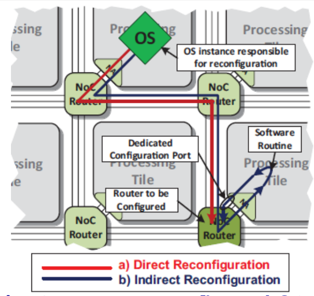


• Orders of magnitude!
• Gets better with size

System Size	Flat Mesh (nJ)	WiNoC (nJ)	Factor
128	1319	22.57	58x
256	2936	24.02	122x
512	4992	37.48	133x

[Ganguly et al. IEEE Trans. Comp., 2010]



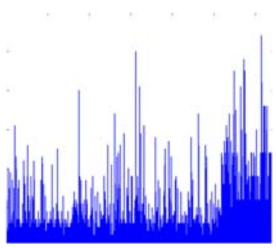


[Heisswolf et al. ACM Trans. Reconfig. Tech&Syst, 2013]

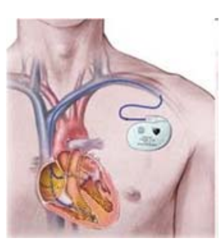
Local control w/ full state information, global control w/ partial information. Small world effects help convergence

This presentation focuses on the new CPS paradigm and its impact on future integrated systems

Structure
Architecture and small world effects

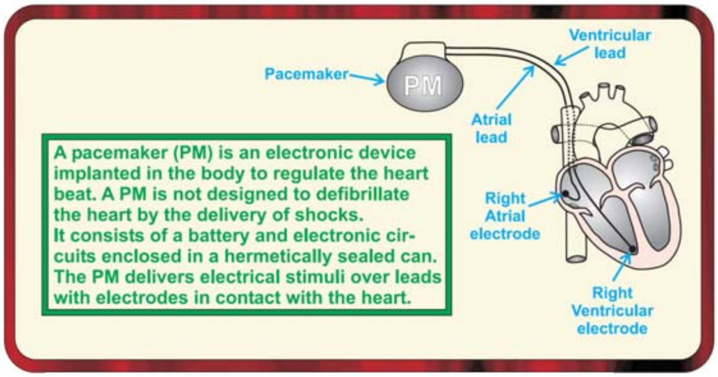


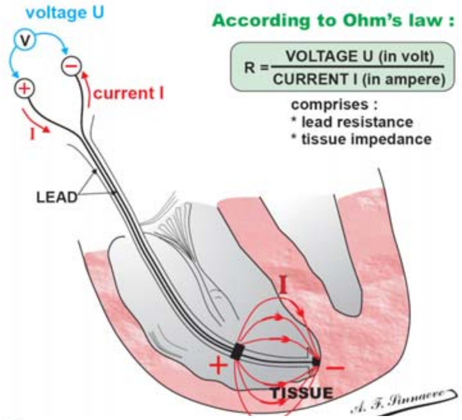
Dynamics
Workloads and multiscale behavior

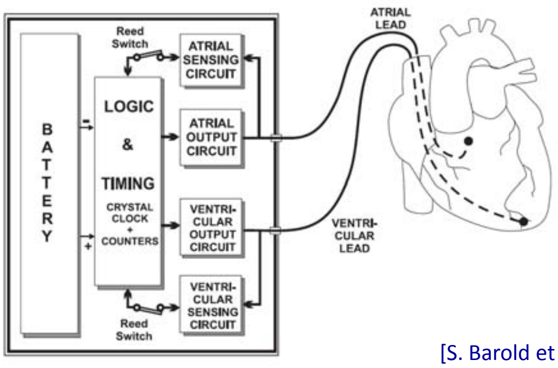


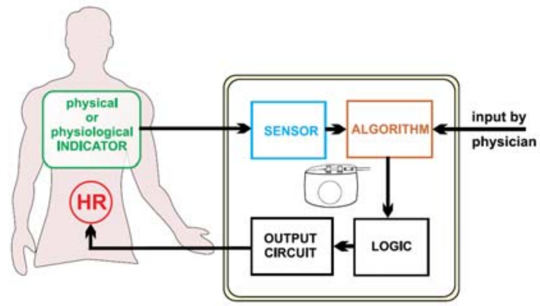
Control
Power and resource management

A pacemaker analyzes the function of the heart; if necessary, sends signals to correct certain abnormalities



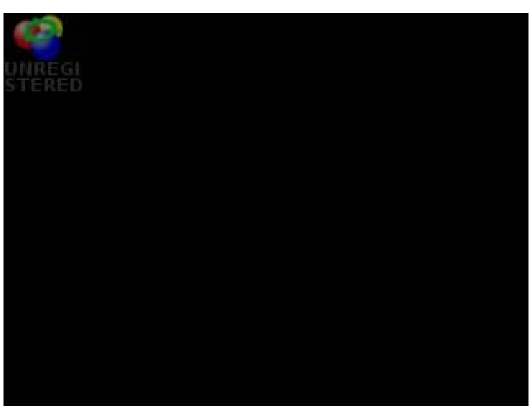


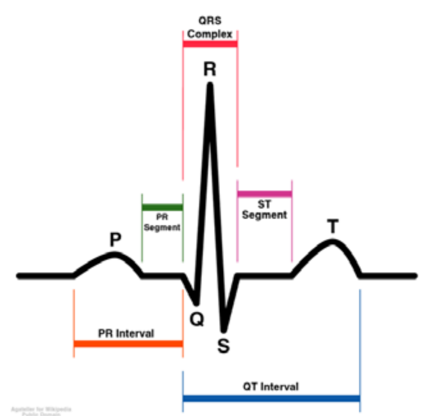




[S. Barold et al. 2004]

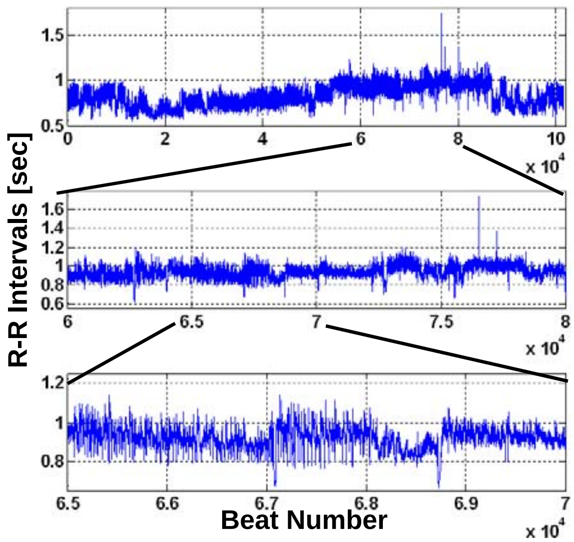
Typical ECG tracing of the cardiac cycle (heartbeat) consists of a P wave, a QRS complex, a T wave, and a U wave

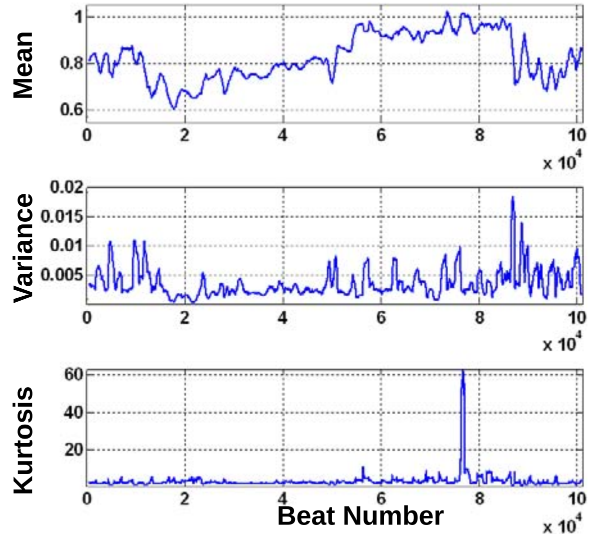




R-R interval: Time between an R wave and the next R wave.
Normal resting heart rate is between 60-100 bpm (0.6-1.0sec)

Heart rate is non-stationary. This means that its distribution and its higher order moments vary with shifts in time

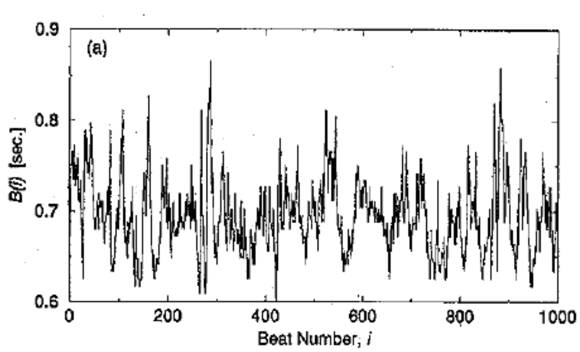


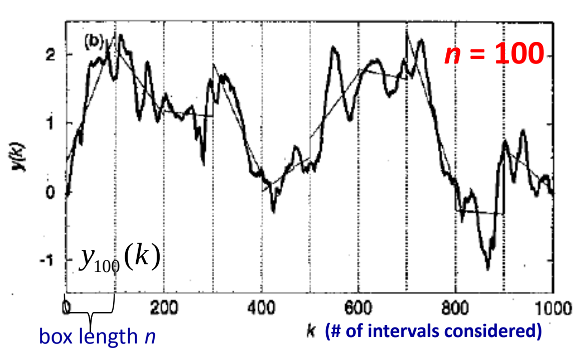


Statistics of R-R intervals vary with time as a function of various activities (sleep, running, etc.)

Heart rate datasets available at: http://www.physionet.org/

DFA can reveal the fractal dimension of R-R intervals





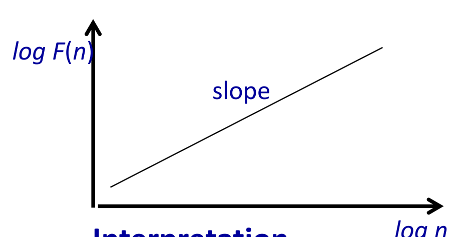
[C.-K. Peng et al.]

i^th inter-beat interval

avg. inter-beat interval

$$y(k)=\sum_{i=1}^{k}\left[B(i)-B_{ave}\right]$$

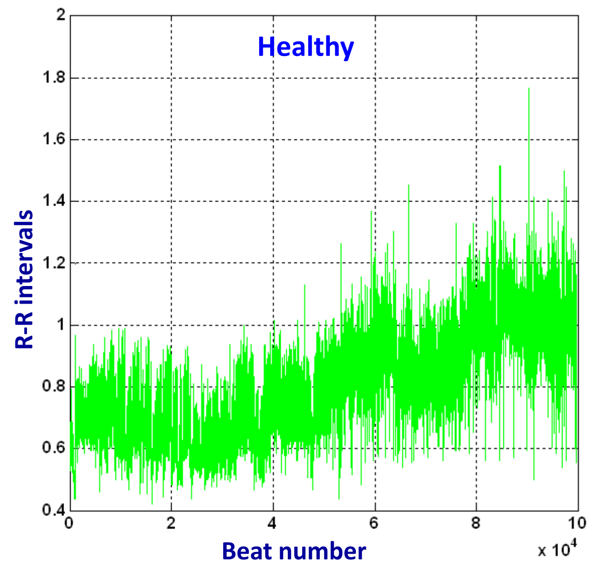
$$F(n)=\sqrt{\frac{1}{N}\sum_{k=1}^{N}\left[y(k)-y_n(k)\right]^2}$$

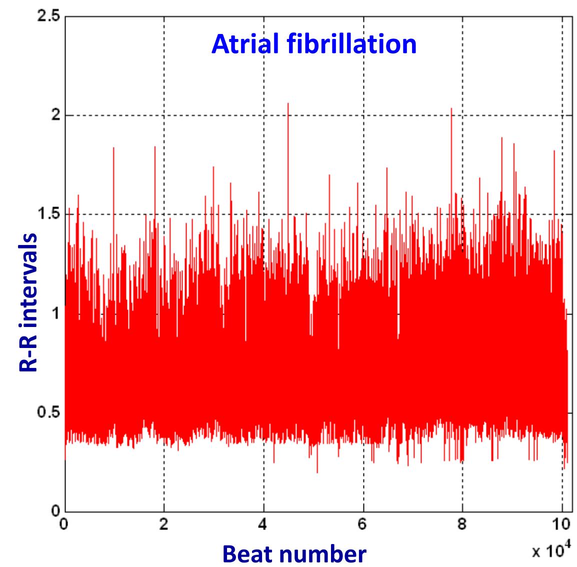


Interpretation

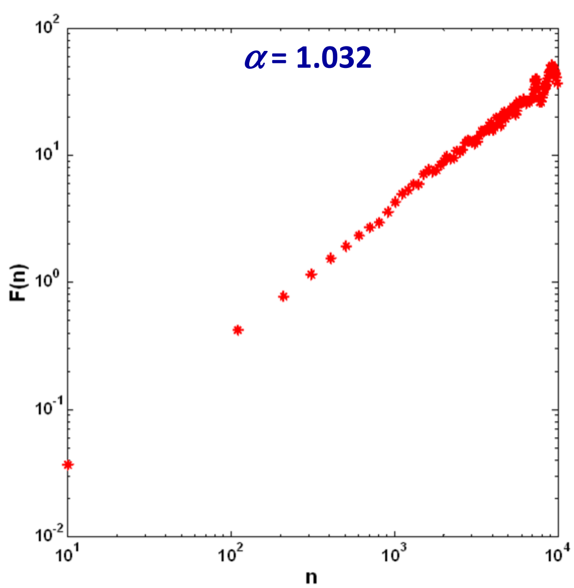
- slope ~ 0.5 white noise
- **slope ~ 1 LRD**
- slope >> 1 Brownian noise

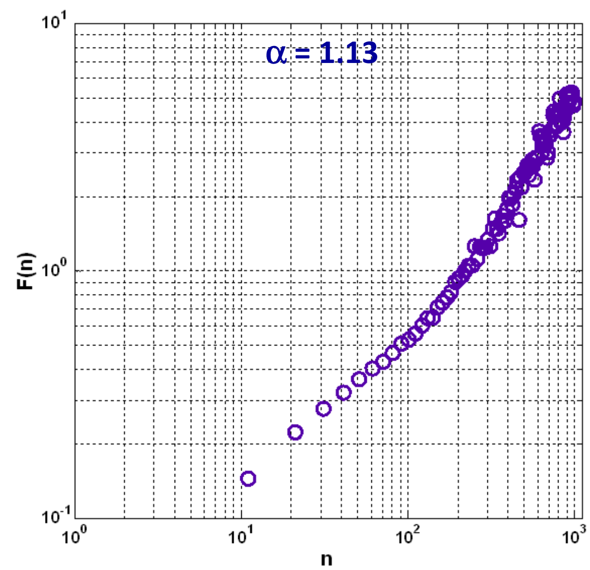
Breakdown of fractal physiological control mechanism can lead to highly periodic output (single scale) or randomness





DFA can be applied to heart rate variability. Degradation of fractal correlations can be used as a quick diagnostic tool.





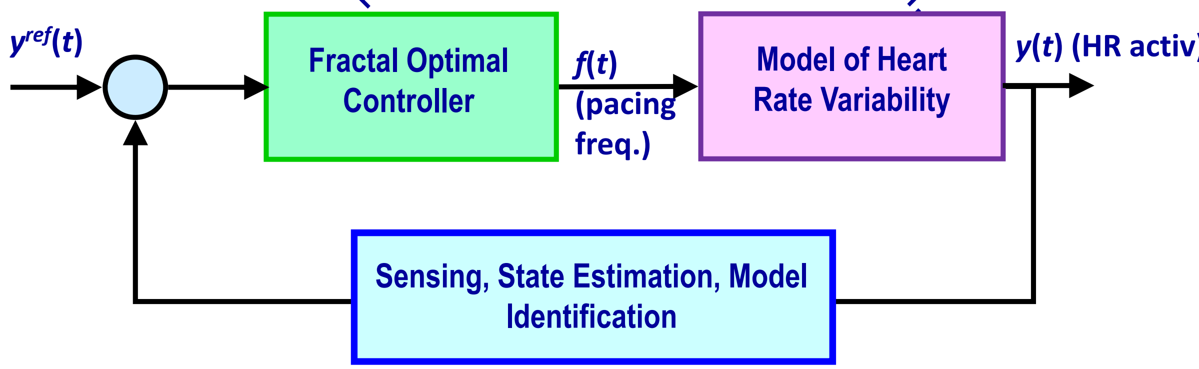
$$F(n) \approx n^{\alpha}$$

Controller finds the pacing frequency which minimizes the deviation of ISE of heart rate from the reference value

$$\min \int_{t_i}^{t_j} \left\{ \frac{1}{2} w \left(y(t) - y^{ref} \right)^2 + \frac{1}{2} z f(t)^2 \right\} dt$$

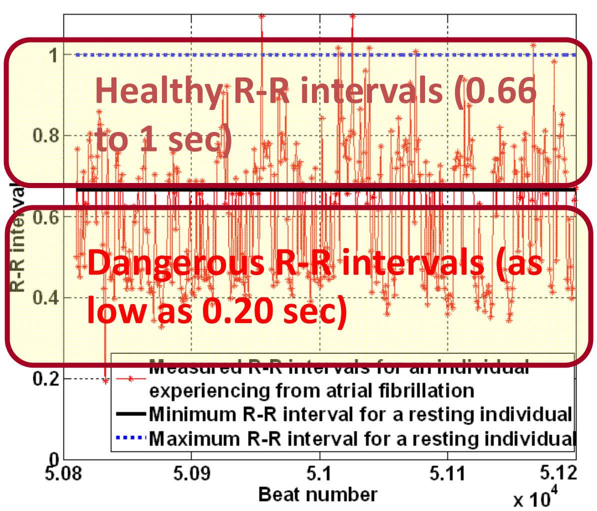
$$y(t_i) = y_0 \quad \text{and} \quad y(t_f) = y_1$$

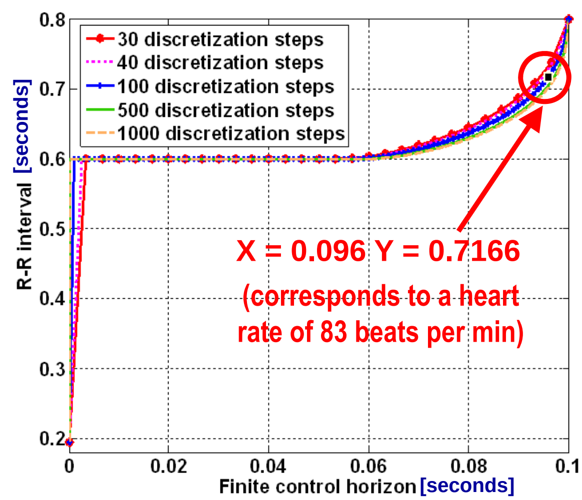
$$f^{\min} \le f(t) \le f^{\max}$$



$$\frac{d^{\alpha} y(t)}{dt^{\alpha}} = a(t) y(t) + b(t) f(t), \qquad y^{\min} \le y(t) \le y^{\max}$$

Atrial fibrillation is characterized by short R-R intervals. If left untreated, it can lead to congestive heart failure





FOC brings the R-R interval from 0.40 to 0.80 secs (i.e., a healthy heart rate of about 75 beats per minute)

Workload analysis should not be an afterthought. In real CPS, network traffic is neither Poisson, nor stationary

Classical dynamics: Linear Dependence & Exponential Inter-Event Distribution

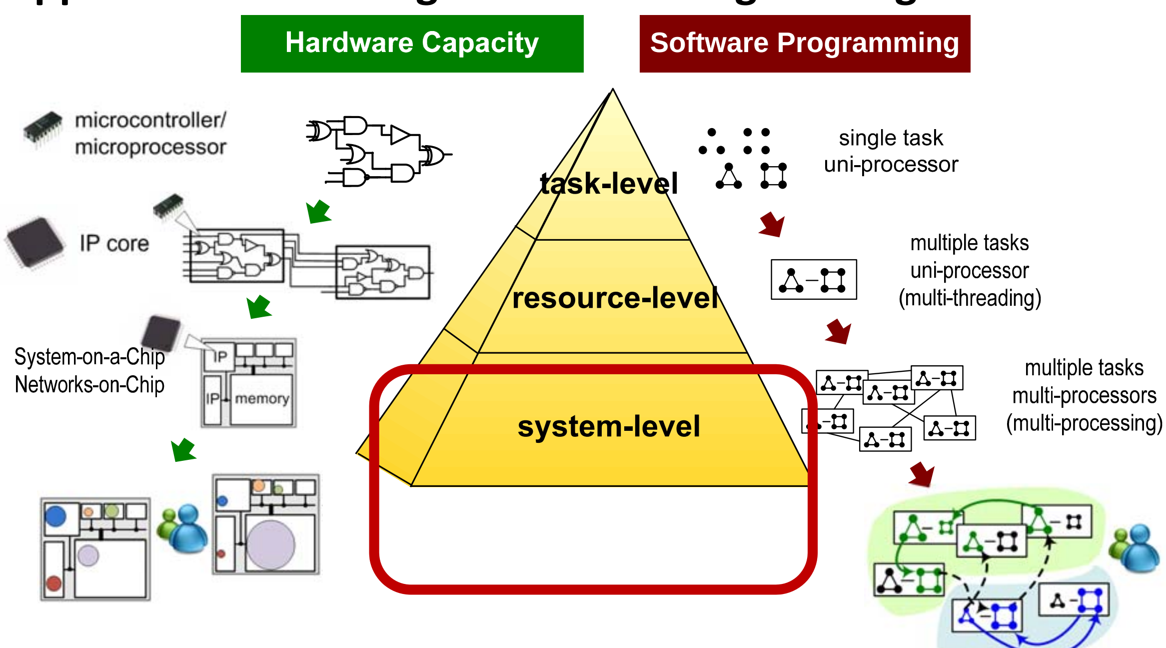
$$\frac{dP(a,t)}{dt} \propto P(a,t)$$
$$\frac{dM_1(t)}{dt} \propto M_1(t)$$

Fractal dynamics: Linear Dependence & Power-Law Inter-Event Distribution

$$\frac{d^{\alpha}P(a,t)}{dt^{\alpha}} \propto P(a,t)$$
$$\frac{d^{\alpha}M_1(t)}{dt^{\alpha}} \propto M_1(t)$$

Statistical properties of the workload have deep implications in resource allocation, architectural design, RT scheduling, etc.

In summary, thousand core systems offer ample opportunities to bring science and engineering even closer



User/application/OS optimization and resource management is the next frontier to conquer.

Finally…

Contributors

Paul Bogdan (Univ. of Southern Calif.), Umit Y. Ogras (Intel), Siddharth Garg (Univ. of Waterloo), Diana Marculescu (Carnegie Mellon Univ), Chen-Ling Chou (Qualcomm), Qian Zhiliang (Hong Kong Univ Sci&Tech), Chi-Ying Tsui (Hong Kong Univ Sci&Tech), Hiroki Matsutani (Keio Univ).

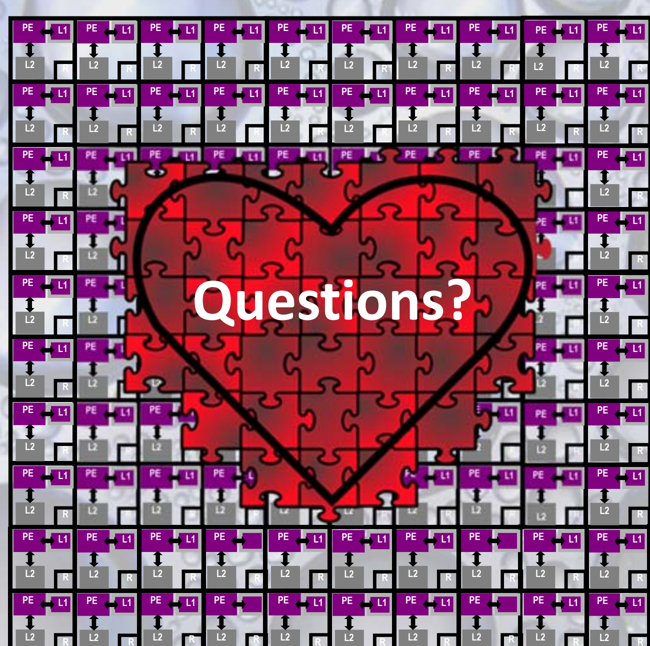
Relevant papers - www.ece.cmu.edu/~sld

Sponsors

National Science Foundation
WHERE DISCOVERIES BEGIN

SRC

Intel Corporation



Thank you!