Complex Dynamics of Microprocessor Performances During Program Execution

Regularity, Chaos, and Others

Hugues BERRY, Daniel GRACIA PÉREZ, Olivier TEMAM

Alchemy, INRIA, Orsay, France
www-rocq.inria.fr/˜berry/

NKS 2006 Conference, Washington, D.C., June 18th, 2006

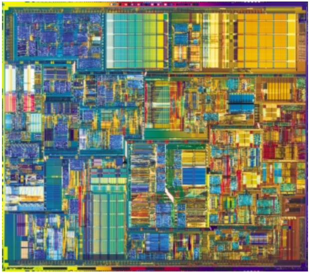
Modern microprocessors are highly complex...

- Moore's Law:
  Exponential increase of the number of transistors/processor

Modern microprocessors are highly complex...

- Moore's Law:
  Exponential increase of the number of transistors/processor

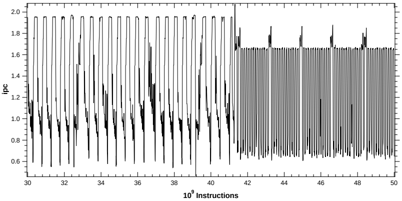
- Pentium 4 (42 million transist.)
- Itanium 2 (410 million transist.)
- A huge quantity of elements, with complex interactions



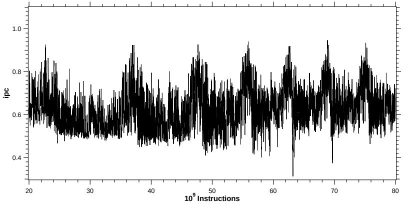
...to increase performance

- Performance = execution speed of a program
- Varies along execution

bzip2



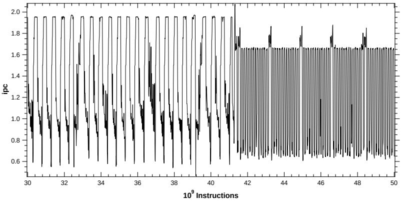
vpr



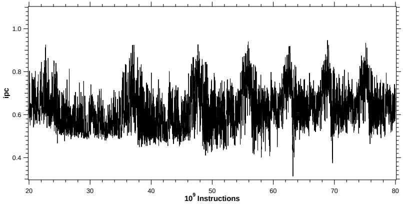
...to increase performance

- Performance = execution speed of a program
- Varies along execution

bzip2



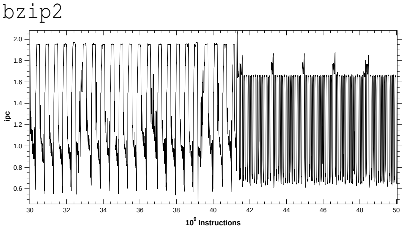
vpr

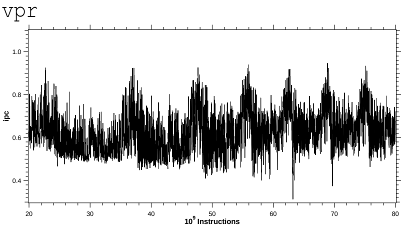


- **How to quantify/characterize this dynamics?** = crucial for understanding/predicting how to increase microprocessor efficiency

...to increase performance

- Performance = execution speed of a program
- Varies along execution





- **How to quantify/characterize this dynamics?** = crucial for understanding/predicting how to increase microprocessor efficiency

Our approach:

use methods from nonlinear time series analysis

Measurements

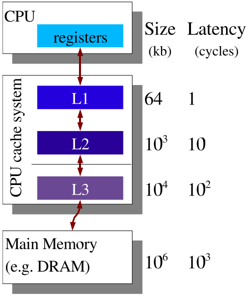
- Execution of typical programs (SPEC benchmarks) by a typical modern superscalar processor (e.g. Pentium 4).

Measurements

- Execution of typical programs (SPEC benchmarks) by a typical modern superscalar processor (e.g. Pentium 4).

Simultaneous measurements during execution:

- number of instructions executed during a clock cycle: **ipc**
- miss fraction in L1 cache: **L1**
- miss fraction in L2 cache: **L2**

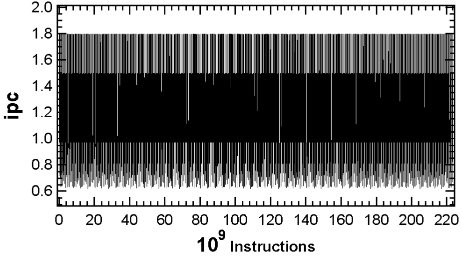


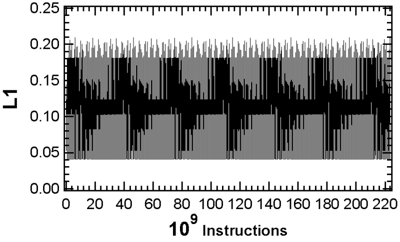
L1 & L2 miss rate = indices for memory usage efficiency (vanishing indices denote highest efficiency)

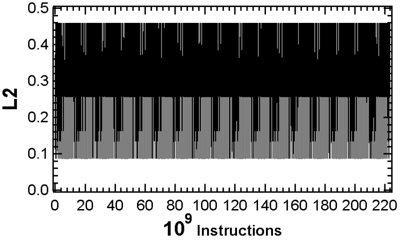
Example 1. applu

(Nonlinear PDEs solver for fluid dyamics)

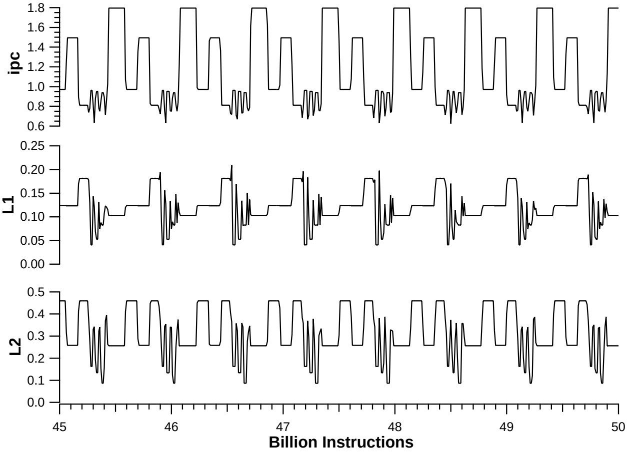
applu: General aspect





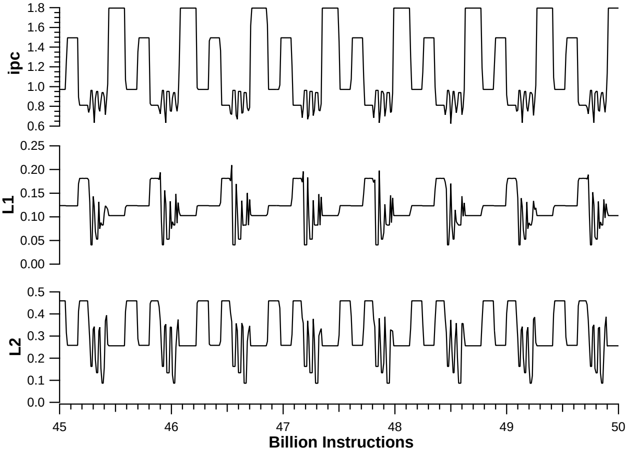


applu: Details



- **REGULAR, PERIODIC DYNAMICS** (limit cycle).

applu: Details



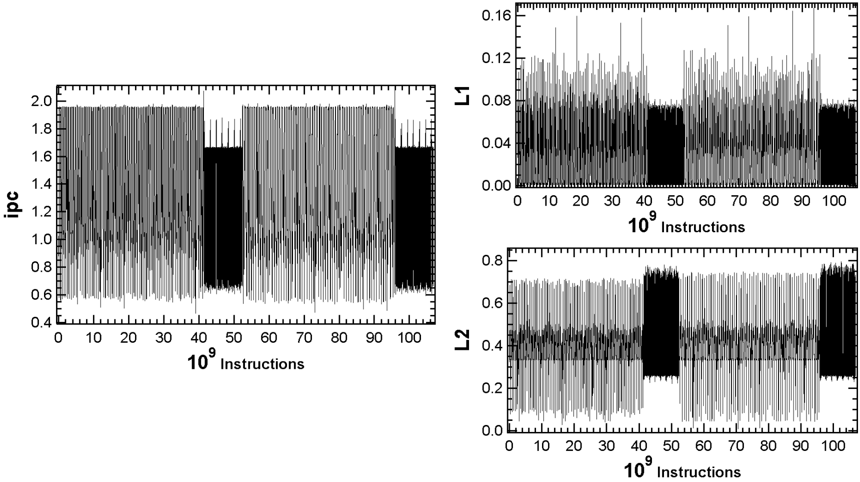
- **REGULAR, PERIODIC DYNAMICS** (limit cycle).

Also found for e.g. apsi (Pollutants air dispersion)

Example 2. bzip2

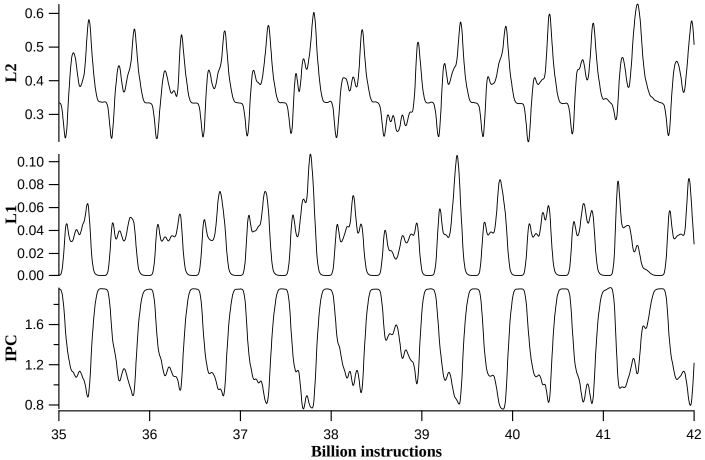
(File compression)

bzip2: General aspect



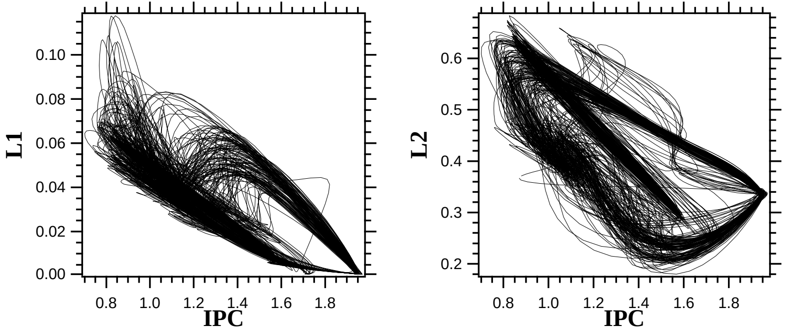
- 2 different phases

bzip2: Details



- Partly regular but much variability / aperiodicity

bzip2: Phase plan projections



- A clear “structure” (attractor?)

Attractor reconstruction: Principle

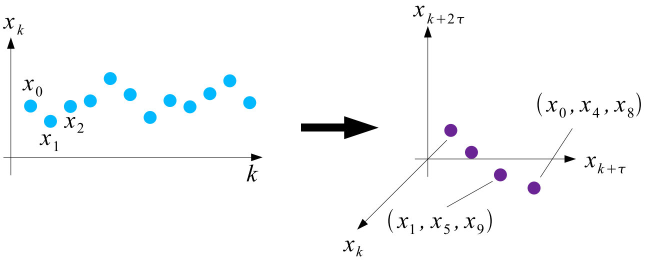
- Aim: construct the attractor underlying the dynamics from a single scalar time series.

Attractor reconstruction: Principle

- Aim: construct the attractor underlying the dynamics from a single scalar time series.
- Delay embedding:

$$x_k \Longrightarrow \mathbf{X_k} = \left(x_k, x_{k+\tau}, \ldots, x_{k+(m-1)\tau}\right)$$

eg. $m = 3$, $\tau = 4$

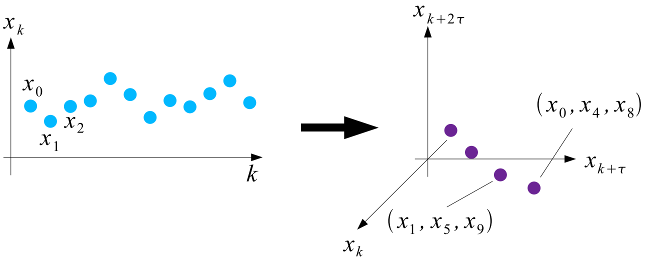


Attractor reconstruction: Principle

- Aim: construct the attractor underlying the dynamics from a single scalar time series.
- Delay embedding:

$$x_k \Longrightarrow \mathbf{X_k} = \left(x_k, x_{k+\tau}, \ldots, x_{k+(m-1)\tau}\right)$$

eg. $m = 3$, $\tau = 4$



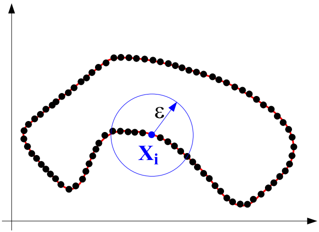
- For adequately chosen m and τ, the reconstructed (embedded) attractor is (topologically) equivalent to the real dynamics attractor [Takens (1981) *Lecture Notes Math.* **898**:366]

Attractor dimension: Principle

[Grassberger & Procaccia (1983) *Physica D* **9**:189]

- Compute “Correlation sums”:

$$C(m,\varepsilon) = \frac{2}{p(p-1)} \sum_{i=1}^{n} \sum_{j>i}^{n} \Theta\left(\varepsilon - \| \mathbf{X_i} - \mathbf{X_j} \|\right)$$



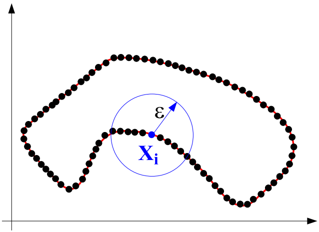
- If a strange attractor is present: $C(m,\varepsilon) \propto \varepsilon^{D_2} \quad \text{for } m >> D_2.$
  - D_2 = (fractal) correlation dimension
  - $\Rightarrow$ scaling of $C(m,\varepsilon)$ = independent of m

Attractor dimension: Principle

[Grassberger & Procaccia (1983) *Physica D* **9**:189]

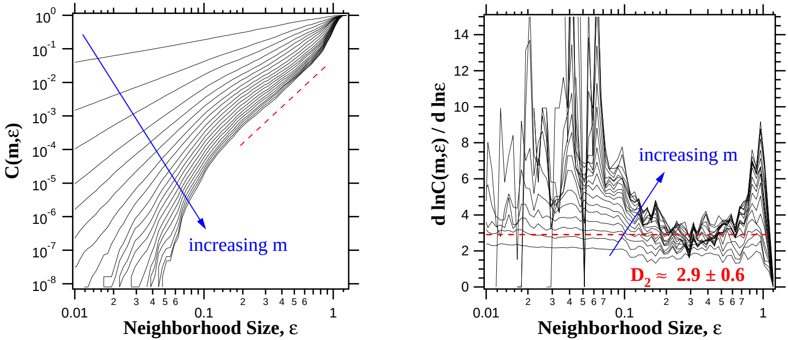
- Compute "Correlation sums":

$$C(m,\varepsilon) = \frac{2}{p(p-1)} \sum_{i=1}^{n} \sum_{j>i}^{n} \Theta\left(\varepsilon - \parallel \mathbf{X_i} - \mathbf{X_j} \parallel\right)$$



- If a strange attractor is present: $C(m,\varepsilon) \propto \varepsilon^{D_2}$ for $m >> D_2$.
  - D_2 = (fractal) correlation dimension
  - $\Rightarrow$ scaling of $C(m,\varepsilon)$ = independent of m
- Whereas, for a stochastic (random) time series: $C(m,\varepsilon) \propto \varepsilon^{m}$
  - $\Rightarrow$ scaling of $C(m,\varepsilon)$ = depends on m

bzip2: Attractor dimension

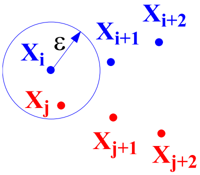


- Presence of a clear scaling zone (where $C(m,\varepsilon) \propto \varepsilon^{\text{Const}}$)
- Indicates the presence of a strange attractor (i.e. low dimensional deterministic chaos).

bzip2: Sensitivity to initial conditions

Compute "stretching factors" [Kantz (1994) *Phys. Lett. A* **185**:177]

$$S(\varepsilon, m, t) = \left\langle \ln \left(\frac{1}{p_i} \sum_{\mathbf{X_j} \in \mathcal{U}_i} \| \mathbf{X_{i+t}} - \mathbf{X_{j+t}} \| \right) \right\rangle$$

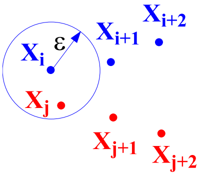


$$\| \mathbf{X_{i+t}} - \mathbf{X_{j+t}} \| \propto \exp(\lambda_{max} t) \Rightarrow S(\varepsilon, m, t) \propto \lambda_{max} t$$

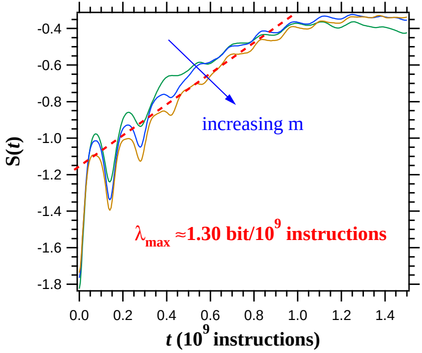
bzip2: Sensitivity to initial conditions

Compute "stretching factors" [Kantz (1994) *Phys. Lett. A* **185**:177]

$$S(\varepsilon, m, t) = \left\langle \ln \left(\frac{1}{p_i} \sum_{\mathbf{X_j} \in \mathcal{U}_i} \| \mathbf{X_{i+t}} - \mathbf{X_{j+t}} \| \right) \right\rangle$$



$$\| \mathbf{X_{i+t}} - \mathbf{X_{j+t}} \| \propto \exp(\lambda_{max} t) \Rightarrow S(\varepsilon, m, t) \propto \lambda_{max} t$$



- Presence of sensitivity to initial conditions ($\lambda_{max} > 0$)
- ***Another strong element in favor of a chaotic dynamics for bzip2***

Conclusion

bzip2:

- Seems to result from chaotic dynamics

Conclusion

bzip2:

- Seems to result from chaotic dynamics
- $\lambda_{max} \approx 0.65$ bit/average “orbit”
- Comparable to textbook chaotic models (e.g. Rössler, $\lambda_{max} = 0.78$ bits/orbit).

Conclusion

bzip2:

- Seems to result from chaotic dynamics
- $\lambda_{max} \approx 0.65$ bit/average "orbit"
- Comparable to textbook chaotic models (e.g. Rössler, $\lambda_{max} = 0.78$ bits/orbit).
- Time series very difficult to predict.

Conclusion

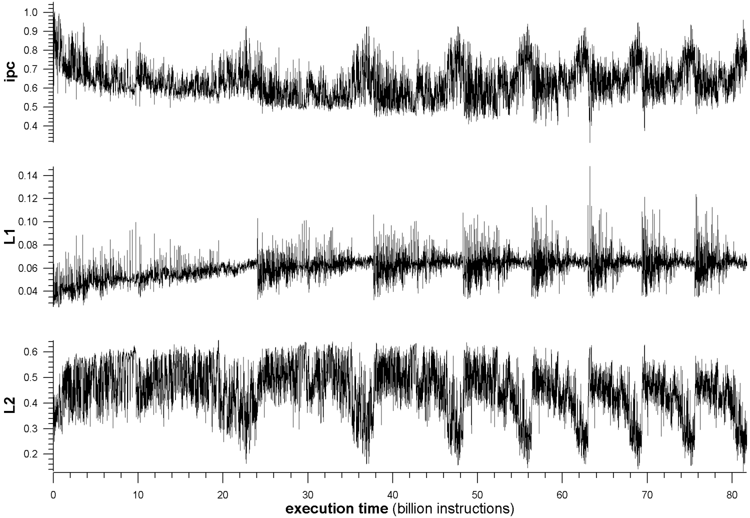
bzip2:

- Seems to result from chaotic dynamics
- $\lambda_{max} \approx 0.65$ bit/average “orbit”
- Comparable to textbook chaotic models (e.g. Rössler, $\lambda_{max} = 0.78$ bits/orbit).
- Time series very difficult to predict.
- Similar behavior observed for `galgel` (Fluid dynamics) or `fma3d` (Finite elements for mechanics)

Example 3. vpr

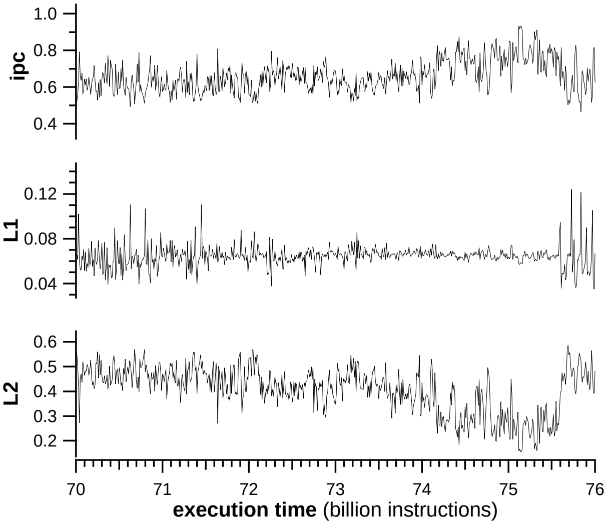
(Node placements and routing in networks)

vpr: General aspect



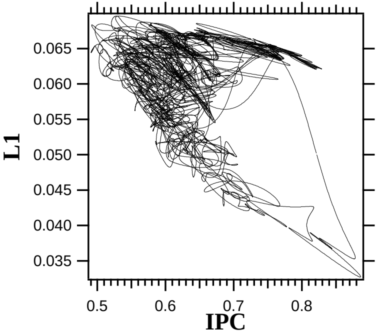
- Very irregular

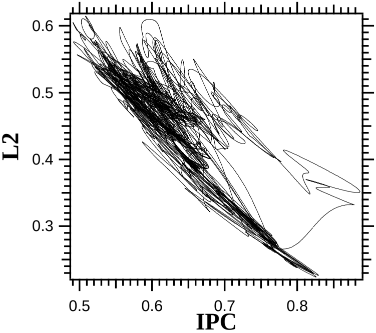
vpr: Details



- Irregularity confirmed

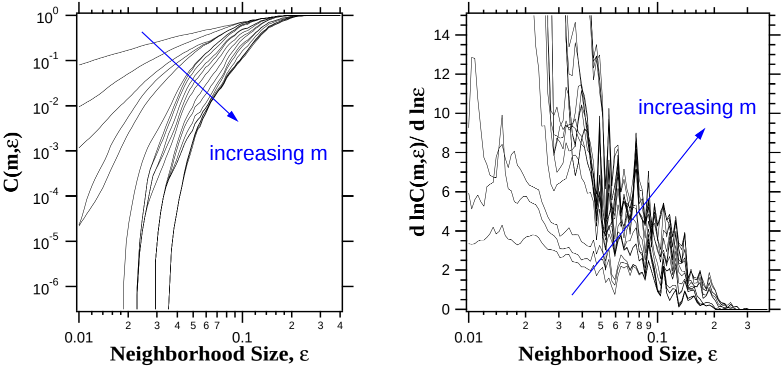
vpr: Phase plan projection





- Hardly structured

vpr: Attractor dimension



- No clear scaling zone
- No evidence for a (low dimensional) attractor
- Stochastic signal??

Conclusion

vpr:

- Seems to originate from a non deterministic time series

Conclusion

vpr:

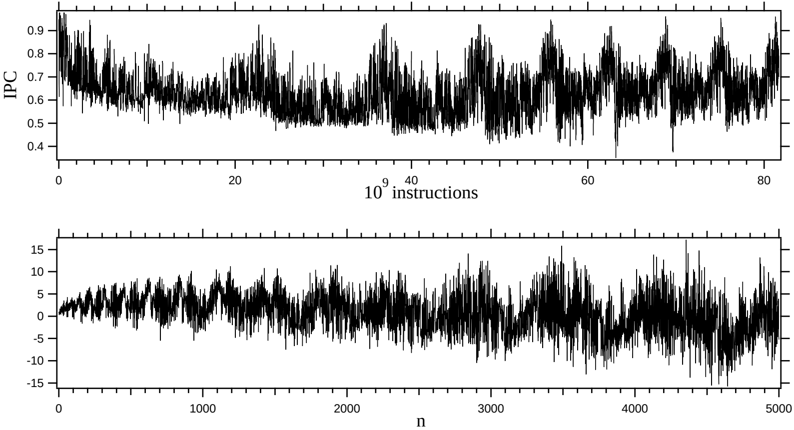
- Seems to originate from a non deterministic time series
- However, the underlying processes in the microprocessor are ***fundamentally*** deterministic
- How to discriminate stochastic/deterministic with long repetition period?

Conclusion

vpr:

- Seems to originate from a non deterministic time series
- However, the underlying processes in the microprocessor are ***fundamentally*** deterministic
- How to discriminate stochastic/deterministic with long repetition period?
- Similar behaviors observed for `art` (Neural networks) or `crafty` (Chess game)

Example (from NKS p.129)



Simple “nested” recursion:

$$f(n) = f(f(n-1)) + f(n-2f(n-1)+1), \quad f(1) = f(2) = 1$$

Shown are fluctuations around the average trend $0.42n^{0.816}$

- Easy to generate complex (stochastic-like) series with simple deterministic processes

To conclude

Three dynamics types observed

- Regular periodicity

Three dynamics types observed

- Regular periodicity
- Deterministic chaos

Three dynamics types observed

- Regular periodicity
- Deterministic chaos
- Non chaotic, “stochastic-like” behaviors

Three dynamics types observed

- Regular periodicity
- Deterministic chaos
- Non chaotic, “stochastic-like” behaviors
- Dynamics type of a program $\neq$ correlated to simple characteristics (e.g. floating-point *vs* integer-based codes...)
- but rather to ***how microprocessor architecture (memory hierarchy, branch predictors...) is used by the program***.

Three dynamics types observed

- Regular periodicity
- Deterministic chaos
- Non chaotic, “stochastic-like” behaviors
- Dynamics type of a program $\neq$ correlated to simple characteristics (e.g. floating-point *vs* integer-based codes...)
- but rather to ***how microprocessor architecture (memory hierarchy, branch predictors...) is used by the program***.

For further information/analysis:

- Berry, Gracia Pérez & Temam (2006) *CHAOS* **16**:013110 (arXiv:nlin.AO/0506030)
- www-rocq.inria.fr/˜berry

Why this complexity?

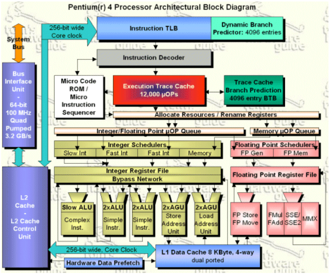
Hiding memory latency

- The increase rate of the clock frequency is much larger than that of memory accesses
- The latency for memory access is thus always larger (currently, **hundreds** of cycles for RAM access)
- A myriad of mechanisms has been developed to “hide” this caveat and increase performance:

Hiding memory latency

- The increase rate of the clock frequency is much larger than that of memory accesses
- The latency for memory access is thus always larger (currently, **hundreds** of cycles for RAM access)
- A myriad of mechanisms has been developed to “hide” this caveat and increase performance:

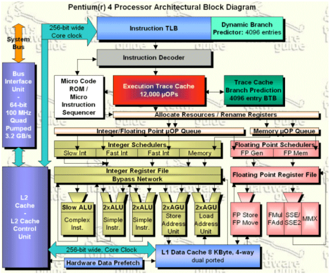
- Cache memory hierarchies (data and instructions)
- Parallelization at various levels
- Pipelining
- Speculative execution
- ...



Hiding memory latency

- The increase rate of the clock frequency is much larger than that of memory accesses
- The latency for memory access is thus always larger (currently, **hundreds** of cycles for RAM access)
- A myriad of mechanisms has been developed to “hide” this caveat and increase performance:

- Cache memory hierarchies (data and instructions)
- Parallelization at various levels
- Pipelining
- Speculative execution
- ...



Increase of the complexity

Example: speculative execution

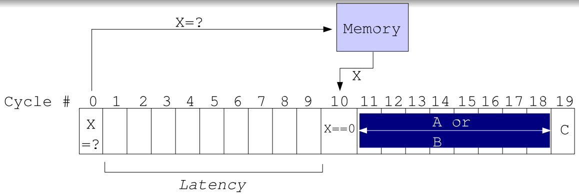
Example

```
if (X==0)
    { A }
else
    { B }
C
```

Example: speculative execution

Example

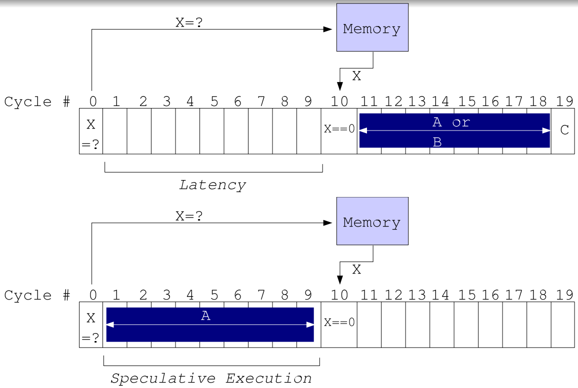
```
if (X==0)
    { A }
else
    { B }
C
```



Example: speculative execution

Example

```
if (X==0)
    { A }
else
    { B }
C
```

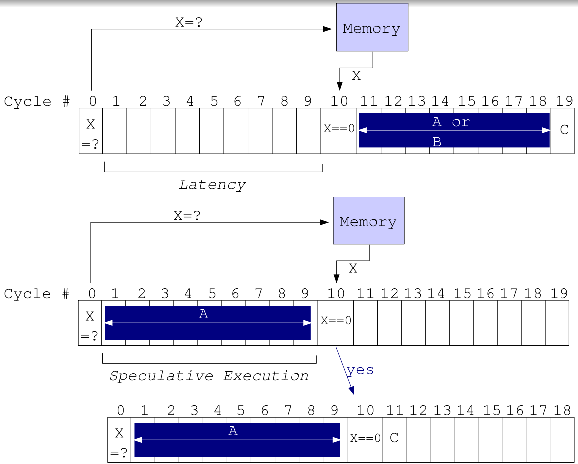


- speculative execution (branch predictor).

Example: speculative execution

Example

```
if (X==0)
    { A }
else
    { B }
C
```



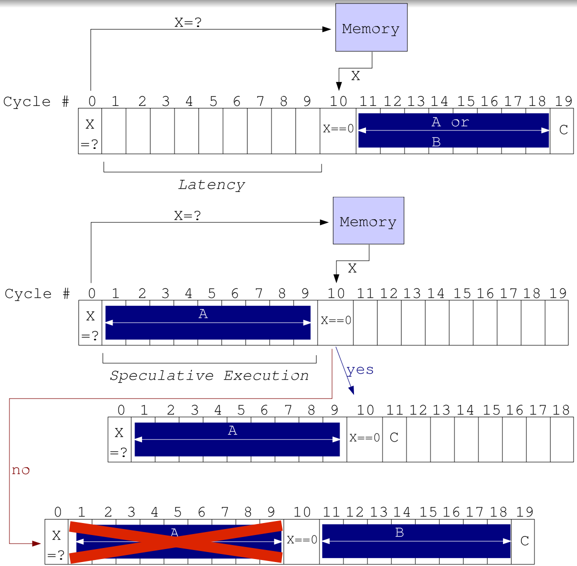
- speculative execution (branch predictor).
- If prediction = true: the memory latency time has been skipped

Example: speculative execution

Example

```
if (X==0)
     { A }
else
     { B }
C
```

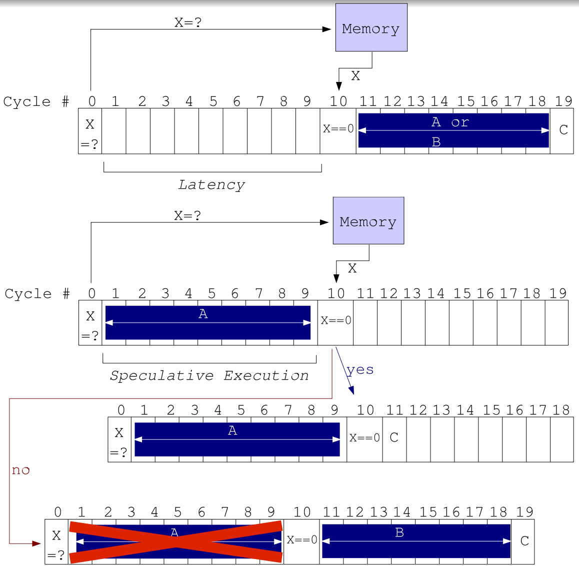
- speculative execution (branch predictor).
- If prediction = true: the memory latency time has been skipped
- If not: forget speculative execution results



Example: speculative execution

Example

```
if (X==0)
    { A }
else
    { B }
C
```



- speculative execution (branch predictor).
- If prediction = true: the memory latency time has been skipped
- If not: forget speculative execution results

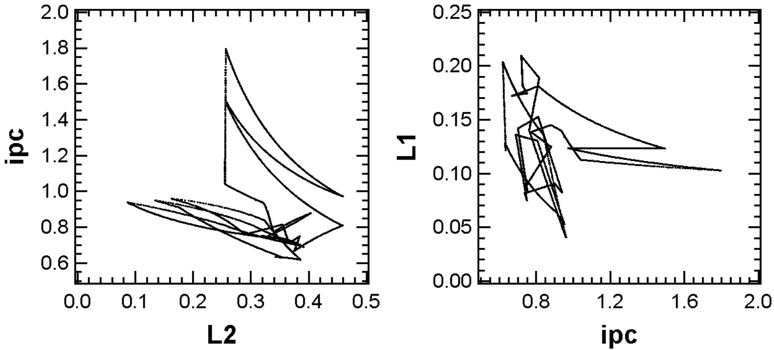
Performance can be history-dependent

So that:

- Performance *(number of instructions executed per time units)* at a given point depends on a huge quantity of architectural mechanisms, that interact in a nonlinear fashion.
- The state of each of these mechanisms at a given point cannot be known precisely.
- This property has been exploited to build random number generators (Seznec & Sandrier, 2003).

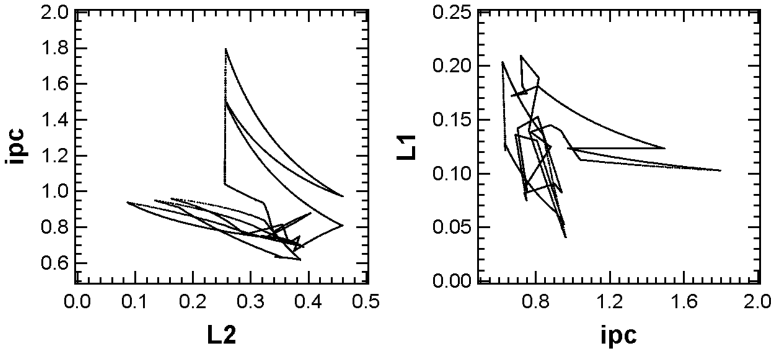
Supplementary results for applu

applu: Phase plan projection



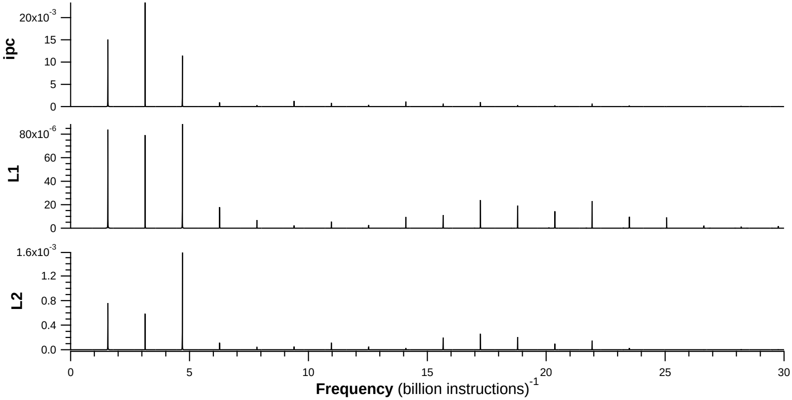
- Clear regular periodicity (limit cycle)

applu: Phase plan projection



- Clear regular periodicity (limit cycle)
- **PERIODICAL PERFORMANCE OSCILLATIONS.**

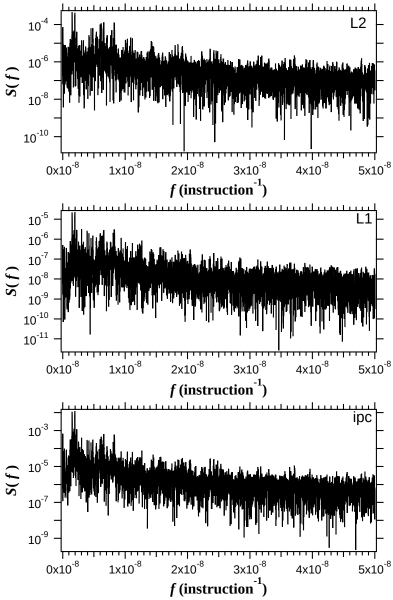
applu: Spectral analysis



- **Clear periodic behavior.**

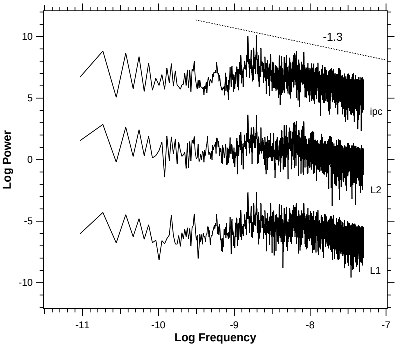
Supplementary results for bzip2

bzip2: Spectral Analysis

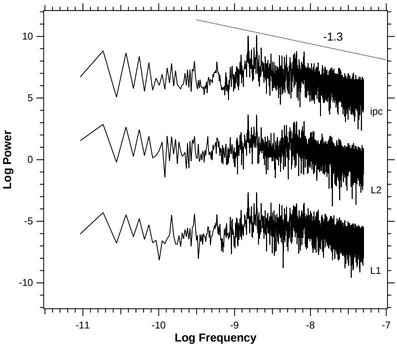


- Some peaks, but a very "dense" structure, typical of chaotic/stochastic signals

bzip2: Spectral Analysis (2)

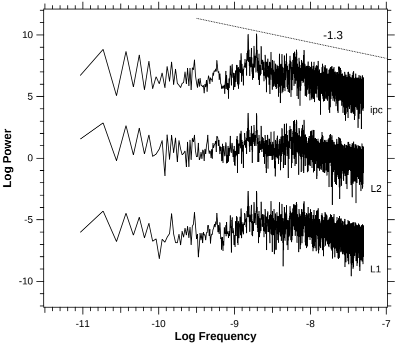


bzip2: Spectral Analysis (2)



- $S(f) \propto f^{-\beta}$
  with $\beta \approx 1.3 \Longrightarrow \approx$ "$1/f$" spectrum

bzip2: Spectral Analysis (2)

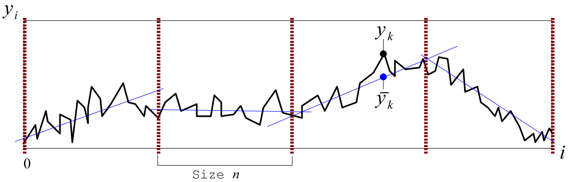


- $S(f) \propto f^{-\beta}$ with $\beta \approx 1.3 \Longrightarrow \approx$ "$1/f$" spectrum
- Fractal series with long term correlations

Detrended Fluctuation Analysis: Principle

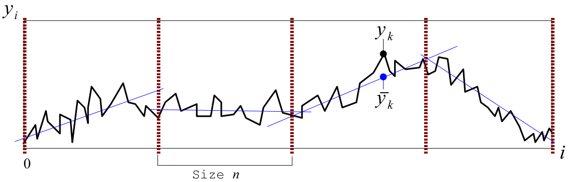
- Principle [Peng *et al.* (1995) *CHAOS* **5**:82]:
  - Centering and integration: $y_k = \sum_{i=1}^{k} [x_i - \langle x \rangle]$

- Principle [Peng *et al.* (1995) *CHAOS* **5**:82]:
  - Centering and integration: $y_k = \sum_{i=1}^{k} [x_i - \langle x \rangle]$
  - The series is then cut in segments of size n instructions



Detrended Fluctuation Analysis: Principle

- Principle [Peng *et al.* (1995) *CHAOS* **5**:82]:
  - Centering and integration: $y_k = \sum_{i=1}^{k} [x_i - \langle x \rangle]$
  - The series is then cut in segments of size n instructions

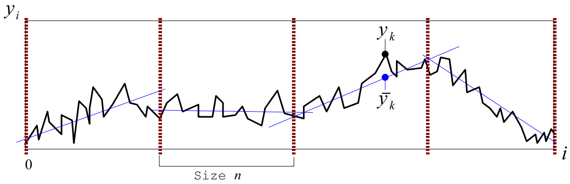


  - Fluctuations around the linear tendency:

$$F(n) = \sqrt{\frac{1}{N} \sum_{k=1}^{N} [y_k - \bar{y}_k]^2}$$

Detrended Fluctuation Analysis: Principle

- Principle [Peng *et al.* (1995) *CHAOS* **5**:82]:
  - Centering and integration: $y_k = \sum_{i=1}^{k} [x_i - \langle x \rangle]$
  - The series is then cut in segments of size n instructions



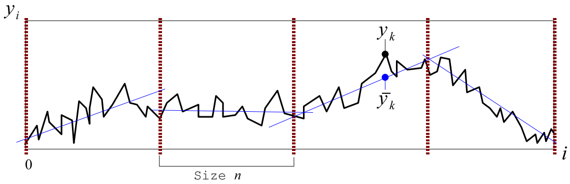
  - Fluctuations around the linear tendency:

$$F(n) = \sqrt{\frac{1}{N} \sum_{k=1}^{N} [y_k - \bar{y}_k]^2}$$

- For "fractal" time series: $F(n) \propto n^{\alpha}$

Detrended Fluctuation Analysis: Principle

- Principle [Peng *et al.* (1995) *CHAOS* **5**:82]:
  - Centering and integration: $y_k = \sum_{i=1}^{k} [x_i - \langle x \rangle]$
  - The series is then cut in segments of size n instructions

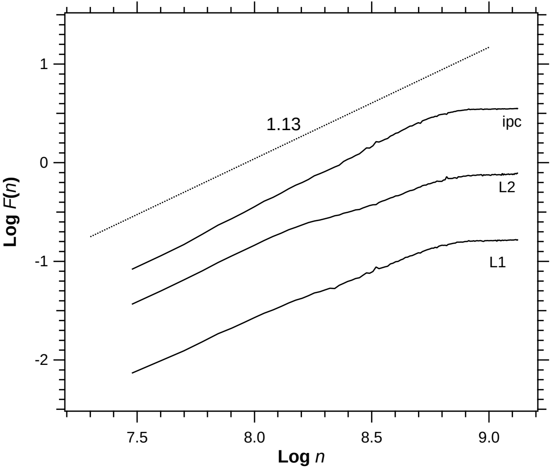


  - Fluctuations around the linear tendency:

$$F(n) = \sqrt{\frac{1}{N} \sum_{k=1}^{N} [y_k - \bar{y}_k]^2}$$

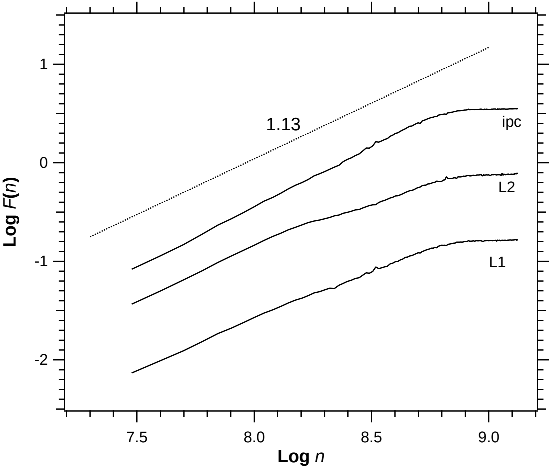
- For “fractal” time series: $F(n) \propto n^{\alpha}$
- $\alpha = 0.5$: no correlation;
- $\alpha > 0.5$: “Fractal” time series with long term correlations;
- Theoretically, $\alpha = (1+\beta)/2$ [Rangarajan & Ding (2000) *PRE* **61**:4991].

bzip2: DFA



- $\alpha \approx 1.13$ (compare to $(1+\beta)/2 = 1.15$)

bzip2: DFA



- $\alpha \approx 1.13$ (compare to $(1+\beta)/2 = 1.15$)
- bzip2 = "**fractal**" series with along term correlations
- Correlations = persistent

  (large values are more likely to occur after a large value)

Recurrence plots (RPs): Principle

- Thresholded RPs: [Eckmann *et al.* (1987) *Europhysics Lett.* **5**:973]
  - Qualitative tests for the presence of patterns and nonlinearity in time series
  - Build the distance matrix between each pair of points in the embedded attractor, then threshold the distance:

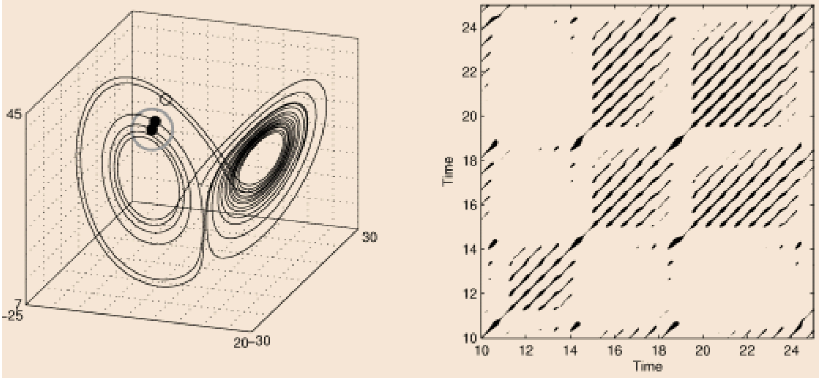
$$\mathbf{R}_{i,j} = \Theta\left(\xi - \parallel \mathbf{X_i} - \mathbf{X_j} \parallel\right) \qquad i,j = 1,\ldots,p$$

where $\Theta(\cdots)$: Heaviside step function

Qualitative graphical interpretation:

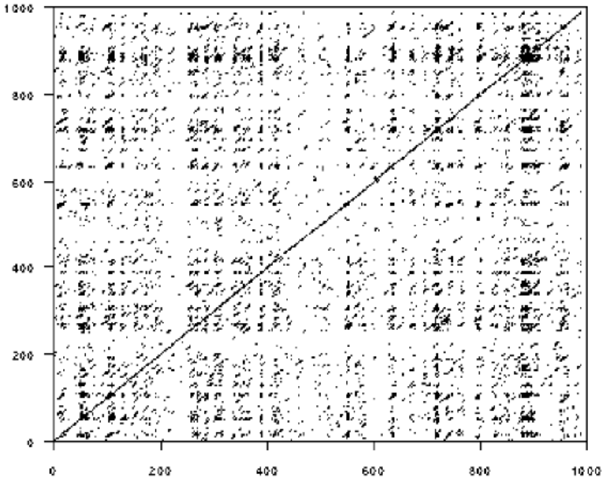
- Diagonals: determinism
- Isolated points: stochasticity
- Interrupted diagonals + isolated points: chaos

Lorenz attractor



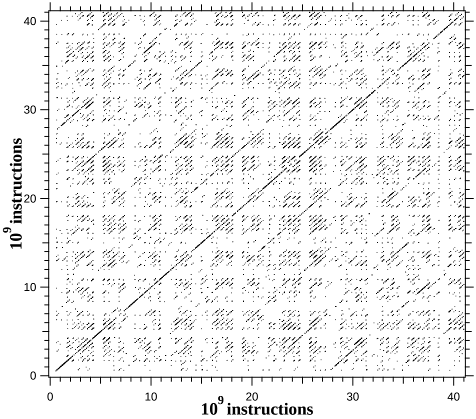
Examples of RPs

Gaussian (white) noise



bzip2: RPs

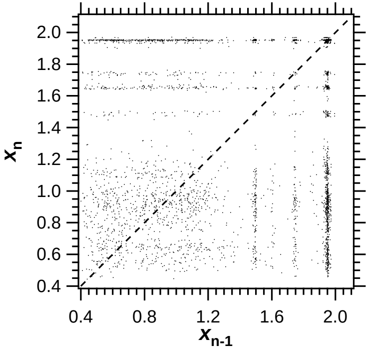
- Scalar series: ipc; Embedding w/ $m = 5$ and $\tau = 0.14 \times 10^9$ instructions.



chaotic time series?

bzip2: Poincaré sections

Poincaré sections (at minima):

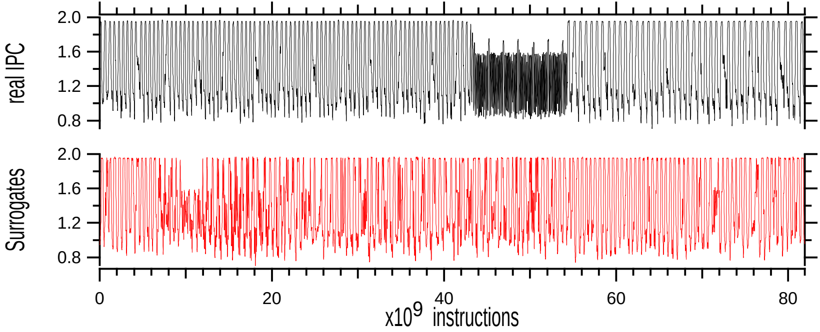


- Structured map
- Mainly mono-dimensional
- Coherent w/ a strange attractor

bzip2: Surrogate data tests

Surrogates have same Fourrier amplitudes and value distribution as real data.

Nonlinearity tested using a simple nonlinearity predictor and time reversal assymetry statistics.



- Possibility that the IPC trace is due to a stationary, possibly rescaled, linear Gaussian random process is be rejected at the 95 % level of significance
- Same conclusion raised when applied to isolated bzip2 regions

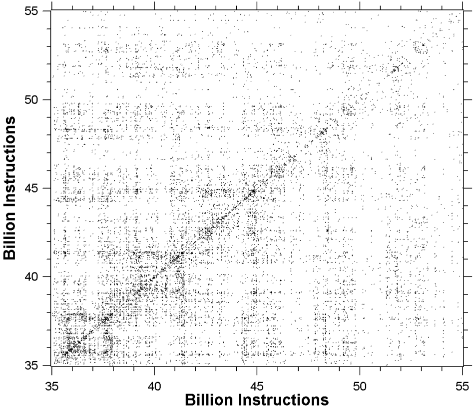
Supplementary results for vpr

vpr: RPs

- Series: ipc; Embedding w/ $m = 4$ and $\tau = 4.16 \times 10^9$ instructions.

vpr: RPs

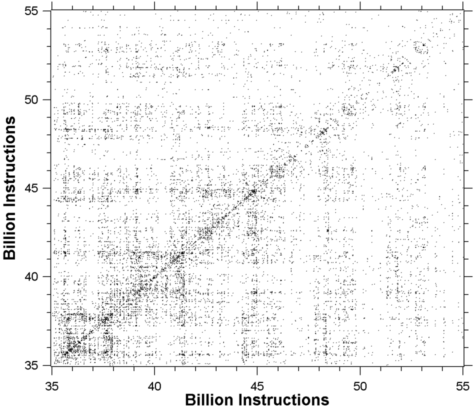
- Series: ipc; Embedding w/ $m = 4$ and $\tau = 4.16 \times 10^9$ instructions.



- Low determinism (lowly structured)
- Close to what is expected for a white noise

vpr: RPs

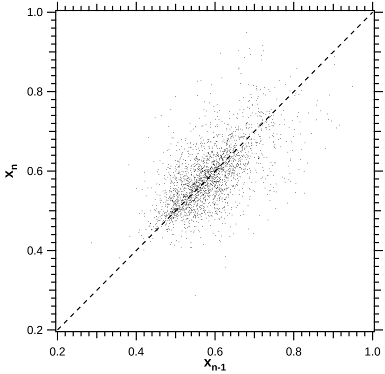
- Series: ipc; Embedding w/ $m = 4$ and $\tau = 4.16 \times 10^9$ instructions.



Non chaotic series. But stochastic?

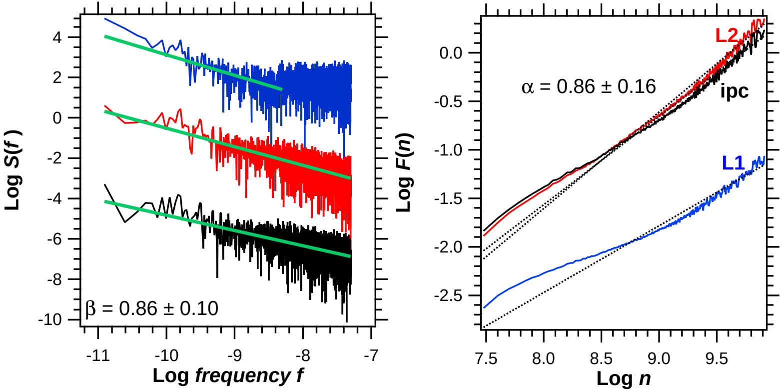
- Low determinism (lowly structured)
- Close to what is expected for a white noise

Poincaré sections (at minima):



- Hardly structured
- Stochasticity?

vpr: Spectral Analysis + DFA

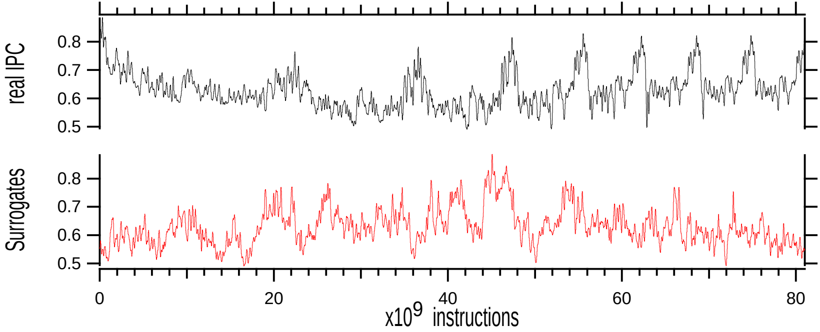


- $S(f) \propto f^{-\beta}$ with $\beta \approx 0.86$
- Bad linear regression, but $\alpha \approx 0.86$ (compare w/ $(1+\beta)/2 = 0.93$)
- vpr could also be "fractal"

vpr: Surrogate data tests

Surrogates have same Fourrier amplitudes and value distribution as real data.

Nonlinearity tested using a simple nonlinearity predictor and time reversal assymetry statistics.

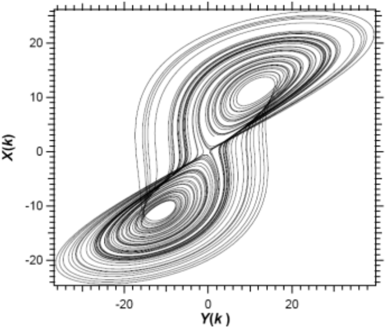


- The null hypothesis that the IPC trace is due to a stationary, possibly rescaled, linear Gaussian random process could not be rejected (95 % level of significance)
- Another point in favor of a stochastic process

Delay embedding of strange attractors

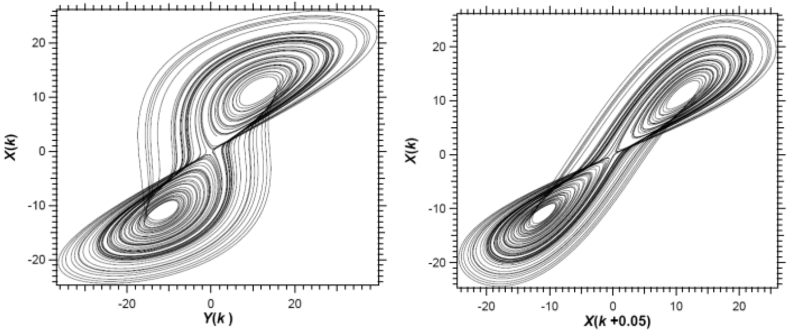
An example of attractor embedding: Lorenz

$\dot{X} = \sigma(Y - X)$ $\dot{Y} = -XZ + rX - Y$ $\dot{Z} = XY - bZ$



An example of attractor embedding: Lorenz

$\dot{X} = \sigma(Y - X)$ $\dot{Y} = -XZ + rX - Y$ $\dot{Z} = XY - bZ$



- Topology conserved w/ $m = 3$, $\tau = 0.05$