Modeling NoC Traffic Locality and Energy Consumption with Rent's Communication Probability Distribution

George Bezerra[1], Stephanie Forrest[1], Melanie Moses[1],
Al Davis[2], Payman Zarkesh-Ha[3]

[1]CS Dept. University of New Mexico
[2]CS Dept. University of Utah
[3]ECE Dept. University of New Mexico

Outline

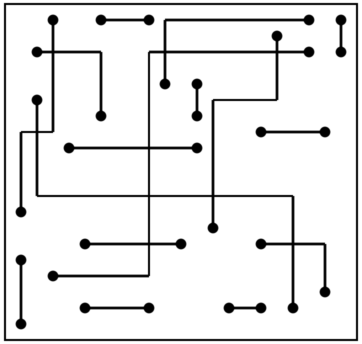
1) Communication Probability Distribution (CPD)

2) Rent’s rule-based synthetic traffic patterns

3) Modeling energy consumption using the CPD

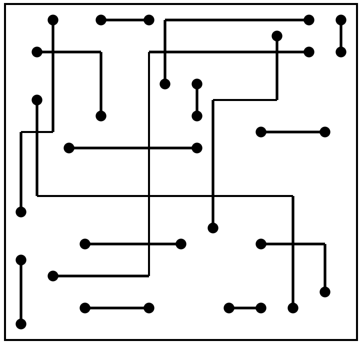
Communication Locality in VLSI

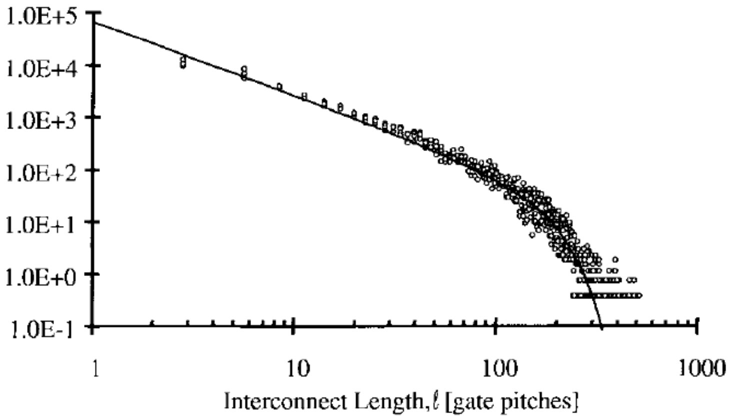
- Conventional VLSI Design
  - Dedicated wires
  - Wires are routed to minimize communication distances (circuit placement)
  - Rent’s rule-based wire length distribution



Communication Locality in VLSI

- Conventional VLSI Design
  - Dedicated wires
  - Wires are routed to minimize communication distances (circuit placement)
  - Rent's rule-based wire length distribution

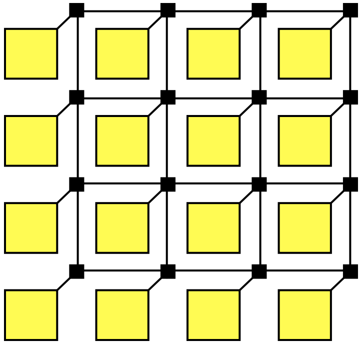




(Davis *et al.*, 1998)

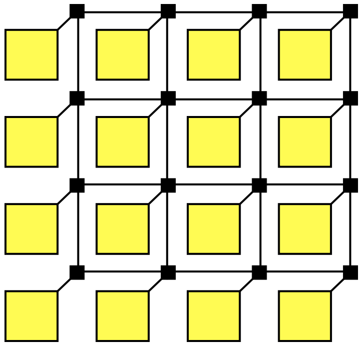
Communication Locality in NoC

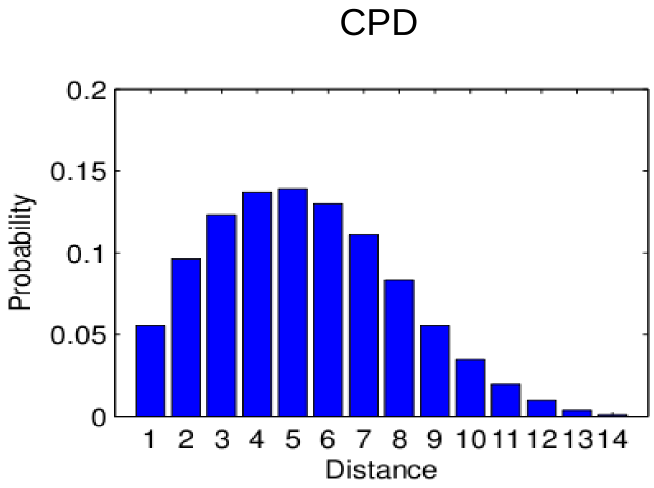
- Network on chip
  - Shared communication medium
  - Packets are routed to minimize communication distances (application mapping)
  - Communication probability distribution (CPD)



Communication Locality in NoC

- Network on chip
  - Shared communication medium
  - Packets are routed to minimize communication distances (application mapping)
  - Communication probability distribution (CPD)





Rent's rule for parallel programs

- Bandwidth version of Rent's rule (Greenfield *et al.*, 2007)

$$B = b N^{p}$$

Rent’s rule for parallel programs

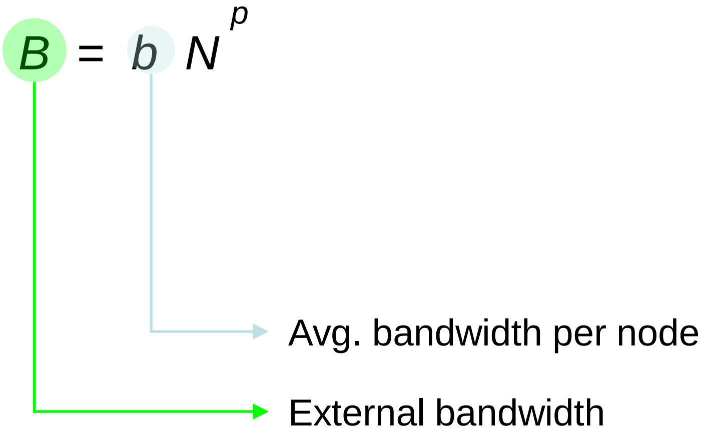
- Bandwidth version of Rent’s rule (Greenfield *et al.*, 2007)

$$B = b\ N^{p}$$

External bandwidth

Rent’s rule for parallel programs

- Bandwidth version of Rent’s rule (Greenfield *et al.*, 2007)



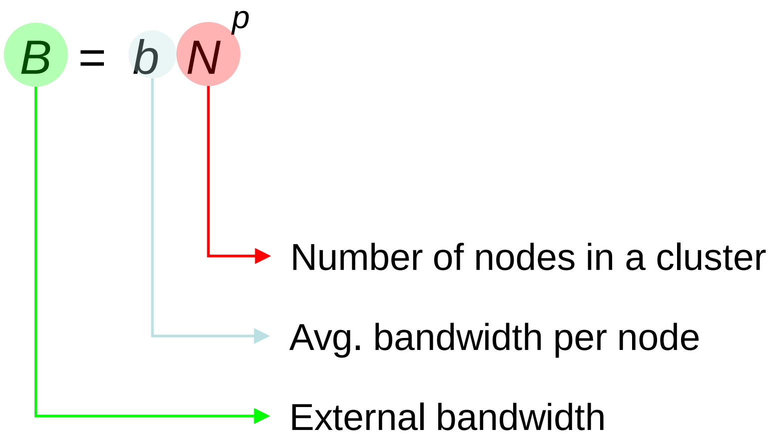
$$B = b\,N^{p}$$

Avg. bandwidth per node

External bandwidth

Rent’s rule for parallel programs

- Bandwidth version of Rent’s rule (Greenfield *et al.*, 2007)



$$B = b N^p$$

Number of nodes in a cluster

Avg. bandwidth per node

External bandwidth

Rent’s rule for parallel programs

- Bandwidth version of Rent’s rule (Greenfield *et al.*, 2007)

$$B = b\,N^{p}$$

Rent’s exponent

Number of nodes in a cluster

Avg. bandwidth per node

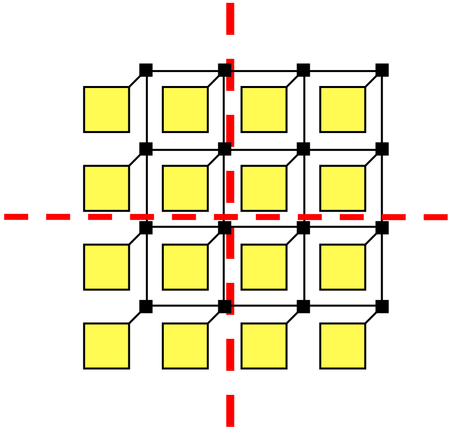
External bandwidth

Rent’s rule for parallel programs

- Measuring Rent’s rule in CMP (Heirman *et al*., 2008)

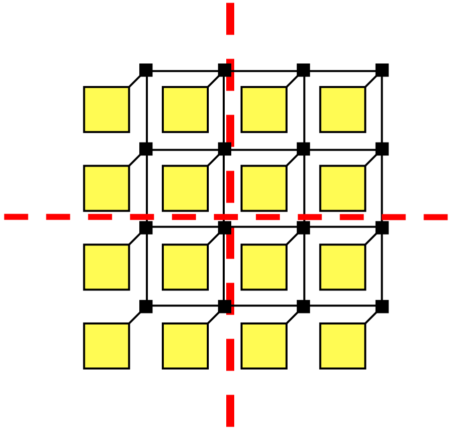
Rent’s rule for parallel programs

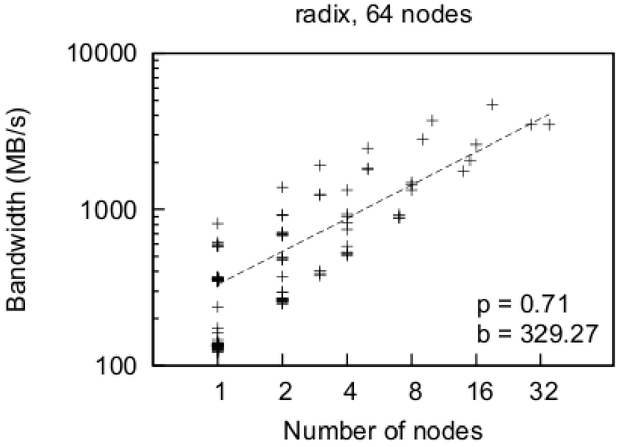
- Measuring Rent’s rule in CMP (Heirman *et al*., 2008)
  - Apply a hierarchical partitioning algorithm



Rent’s rule for parallel programs

- Measuring Rent’s rule in CMP (Heirman *et al*., 2008)
  - Apply a hierarchical partitioning algorithm

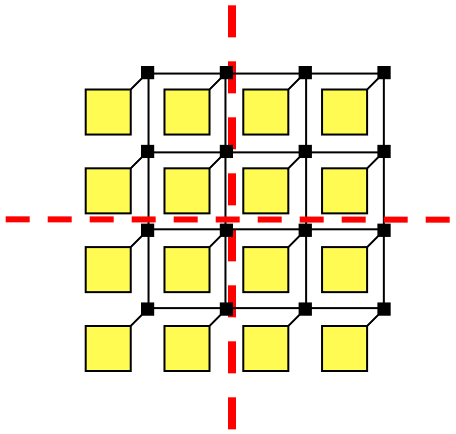


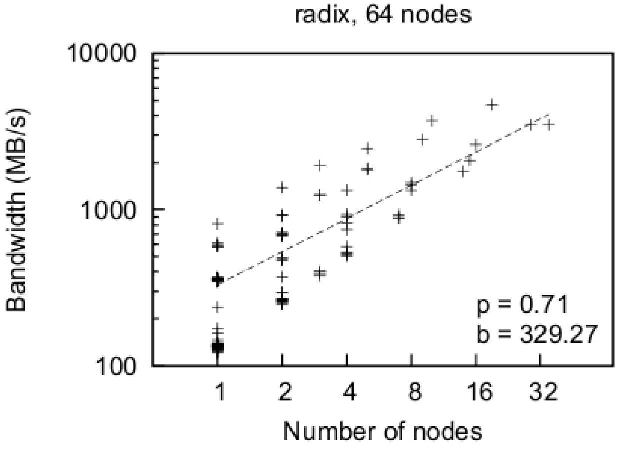


Rent’s rule for parallel programs

- Measuring Rent’s rule in CMP (Heirman *et al*., 2008)
  - Apply a hierarchical partitioning algorithm
  - All 13 benchmark applications followed Rent’s rule with

    $0.55 \leq p \leq 0.75$





Generating synthetic workloads

- Can we generate synthetic traffic patterns that have Rent’s rule properties?
  - Fast and simple way to evaluate an NoC design
  - More accurate representation of communication locality (CPD) than traditional synthetic workloads
  - Simulate hypothetical workload scenarios by varying the Rent’s exponent

Rent’s rule traffic patterns

- Assume that Rent’s rule holds in an unbounded Manhattan grid.

Rent's rule traffic patterns

- Assume that Rent's rule holds in an unbounded Manhattan grid.
- The probability of communication between two nodes with distance *d* apart is? (Adapted from Davis *et al*., 1998)

$$\begin{aligned} P(d) &= \frac{1}{4d}\left[(1+d(d-1))^p - (d(d-1))^p\right. \\ &\quad \left. + (d(d+1))^p - (1+d(d+1))^p\right], \end{aligned}$$

Rent’s rule traffic patterns

- Assume that Rent’s rule holds in an unbounded Manhattan grid.
- The probability of communication between two nodes with distance *d* apart is? (Adapted from Davis *et al*., 1998)

$$\begin{aligned} P(d) &= \frac{1}{4d}\left[(1+d(d-1))^p - (d(d-1))^p\right. \\ &\quad \left. + (d(d+1))^p - (1+d(d+1))^p\right], \end{aligned}$$

```
For each source node $i$:
    For each destination $j$:
        generate a packet from $i$ to $j$ with probability $P(d_{i,j})$
```

Method Validation

1) Generate Rentian traffic using our method

Method Validation

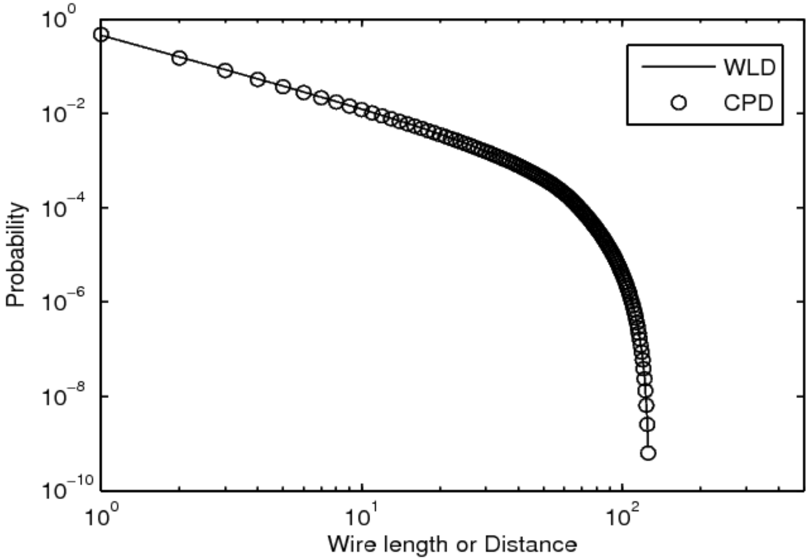
1) Generate Rentian traffic using our method
2) Measure the resulting CPD

Method Validation

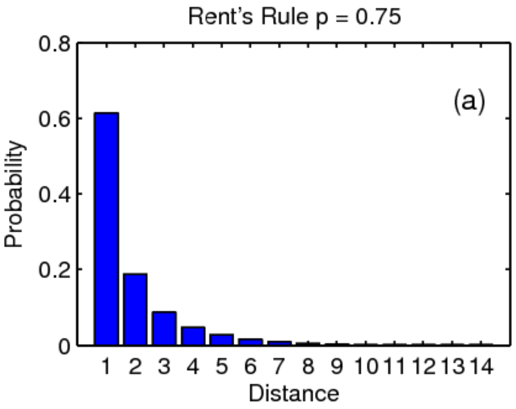
1) Generate Rentian traffic using our method
2) Measure the resulting CPD
3) Compare CPD with the wire length distribution in VLSI (*Davis et al*., 1998)

Method Validation

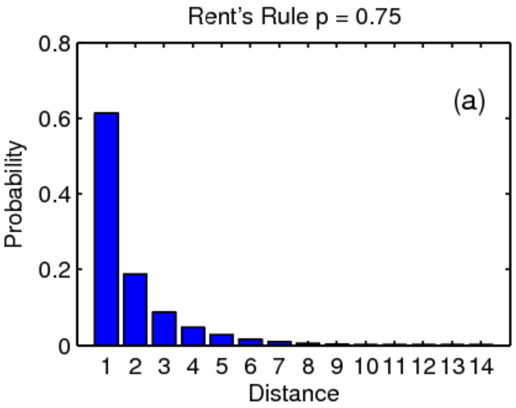
1) Generate Rentian traffic using our method
2) Measure the resulting CPD
3) Compare CPD with the wire length distribution in VLSI (*Davis et al*., 1998)

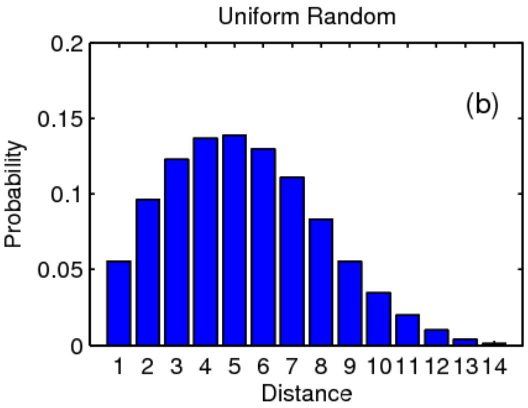


Comparison with other synthetic workloads

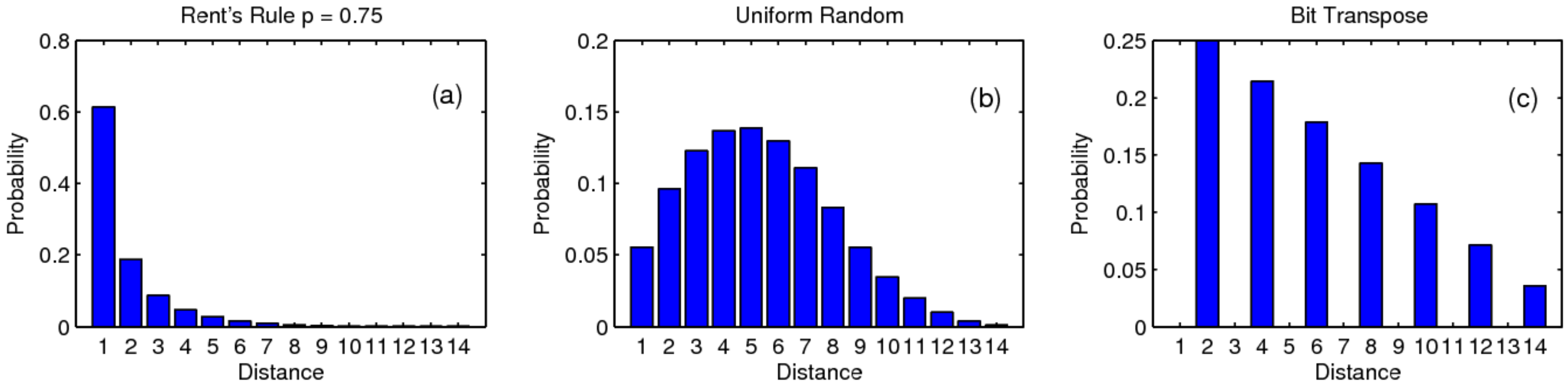


Comparison with other synthetic workloads

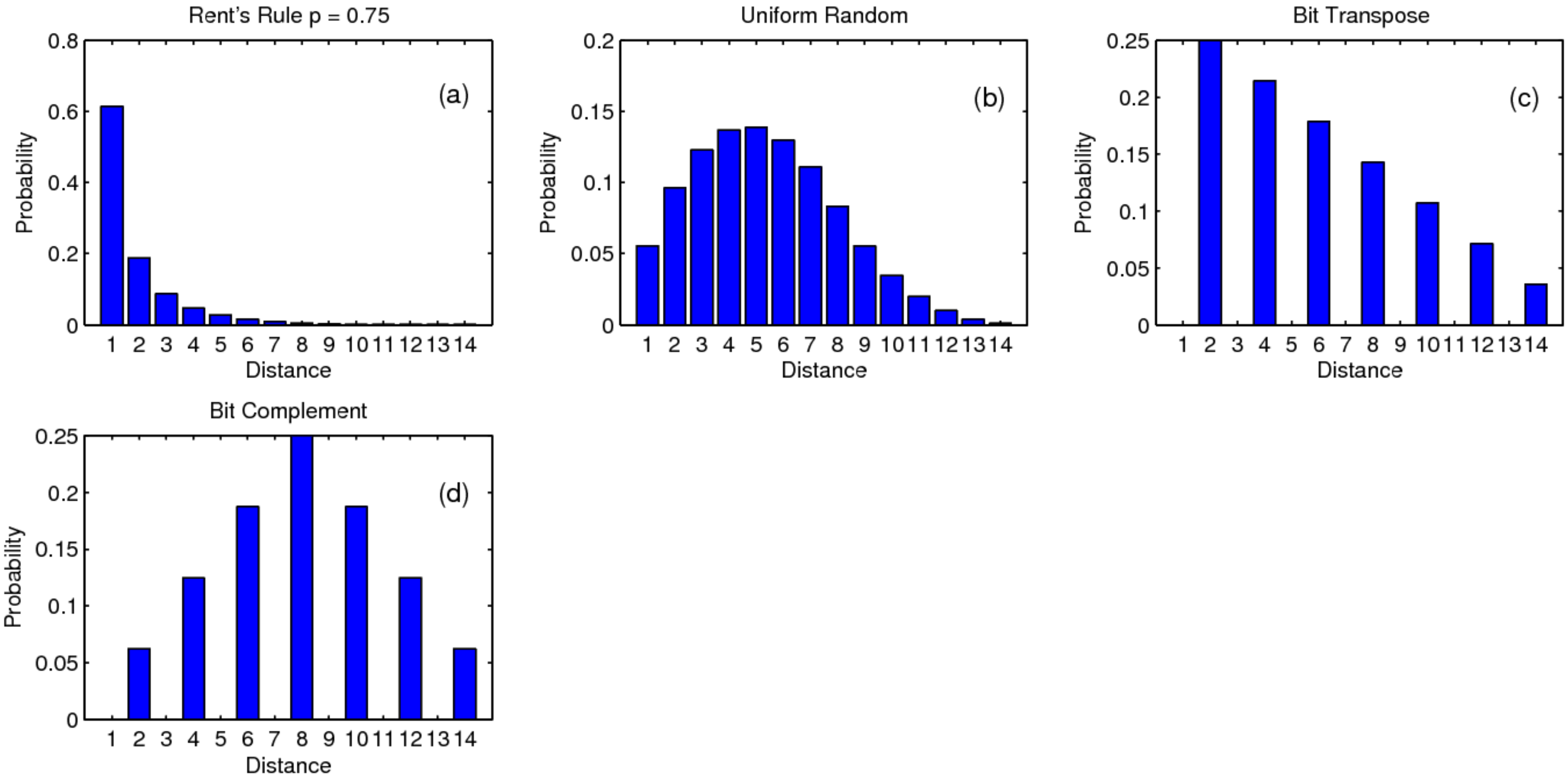




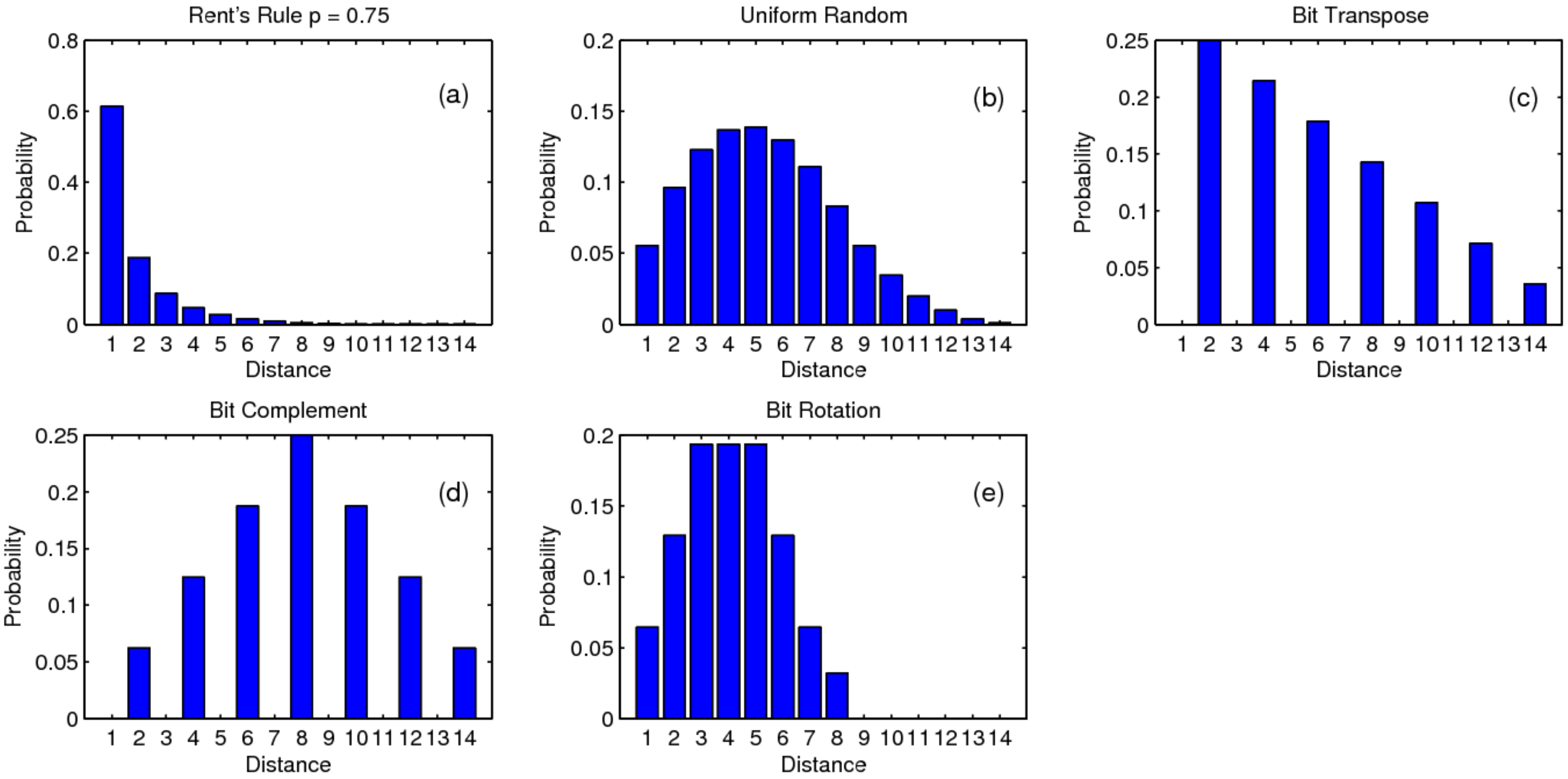
Comparison with other synthetic workloads



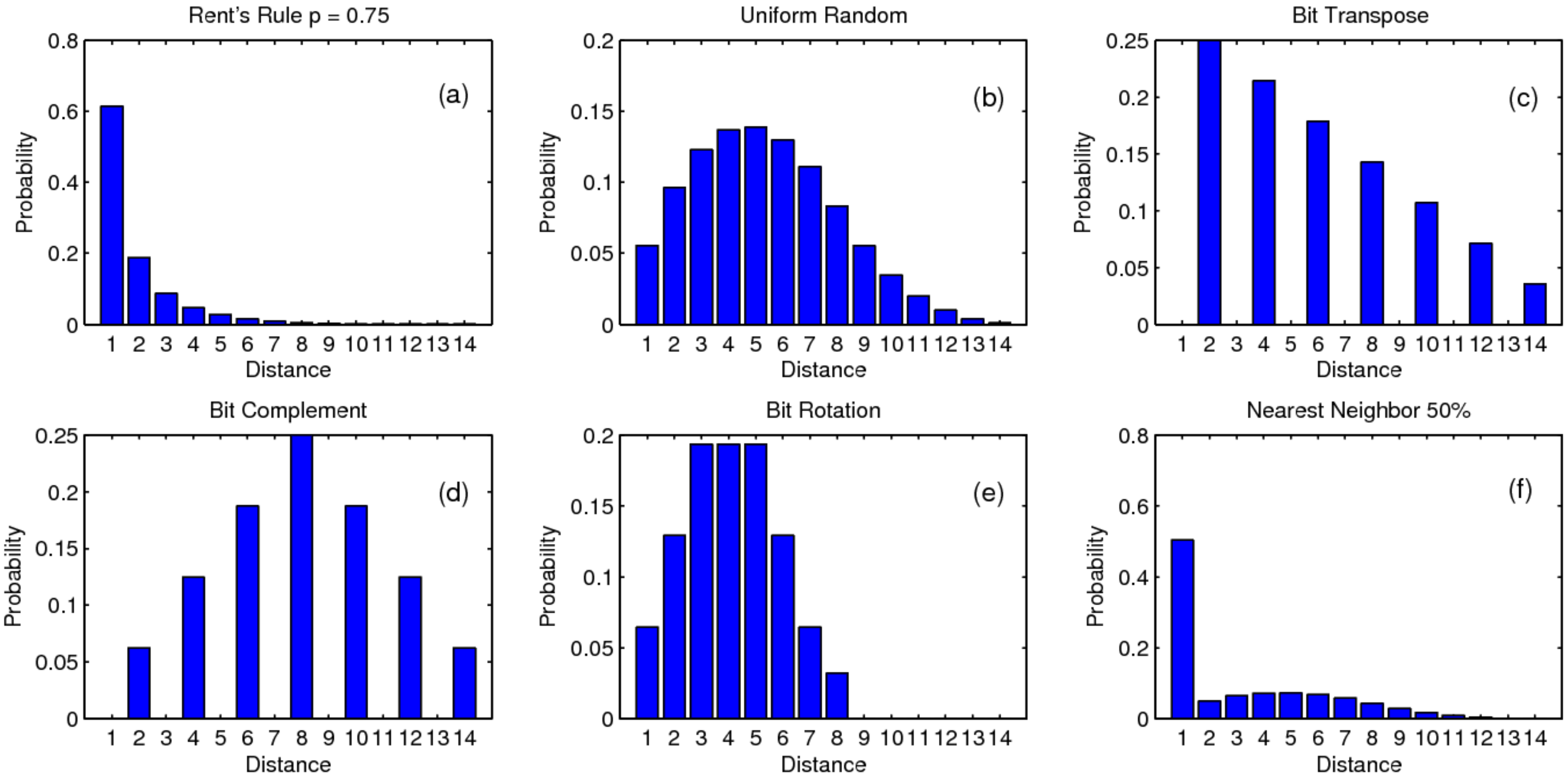
Comparison with other synthetic workloads



Comparison with other synthetic workloads



Comparison with other synthetic workloads



Modeling energy consumption using the CPD

- Can we use the CPD to predict the energy consumption of an application?
    - Energy is roughly proportional to the distance traveled by a packet
    - Fast and simple way of assessing NoC energy consumption
    - Simple way to evaluate application mapping techniques
    - Aid in the design of energy-efficient applications

Energy Model

1) Assume that energy is linear with distance:

$$E_{flit}(d) = d \cdot E_{link} + (d+1) \cdot E_{router},$$

Energy Model

1) Assume that energy is linear with distance:

$$E_{flit}(d) = d \cdot E_{link} + (d+1) \cdot E_{router},$$

2) Energy density:

$$E_{density} = \sum_{d=1}^{max} E_{flit}(d) \cdot CPD(d)$$

Energy Model

1) Assume that energy is linear with distance:

$$E_{flit}(d) = d \cdot E_{link} + (d+1) \cdot E_{router},$$

2) Energy density:

$$E_{density} = \sum_{d=1}^{max} E_{flit}(d) \cdot CPD(d)$$

3) Total energy:

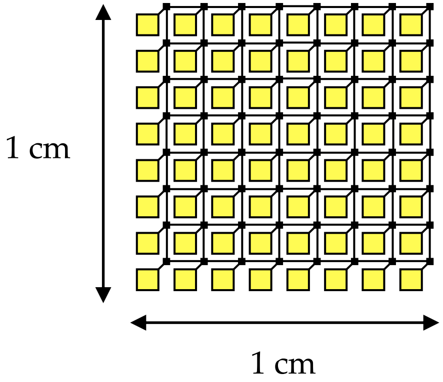
$$E_{total} = N_{packets} \cdot N_{flits} \cdot \sum_{d=1}^{max} E_{flit}(d) \cdot CPD(d)$$

Experiments

- Make energy predictions using our model
- Compare with simulations using Orion
- 8x8 (65nm) and 10x10 (45nm) NoCs
- 7 synthetic traffic patterns

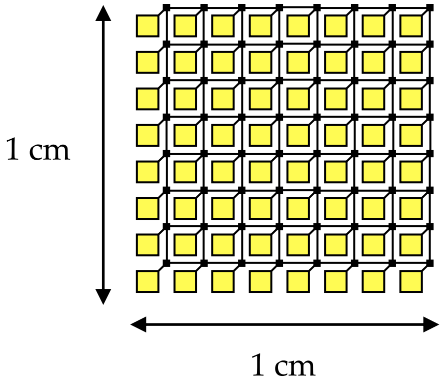
Results

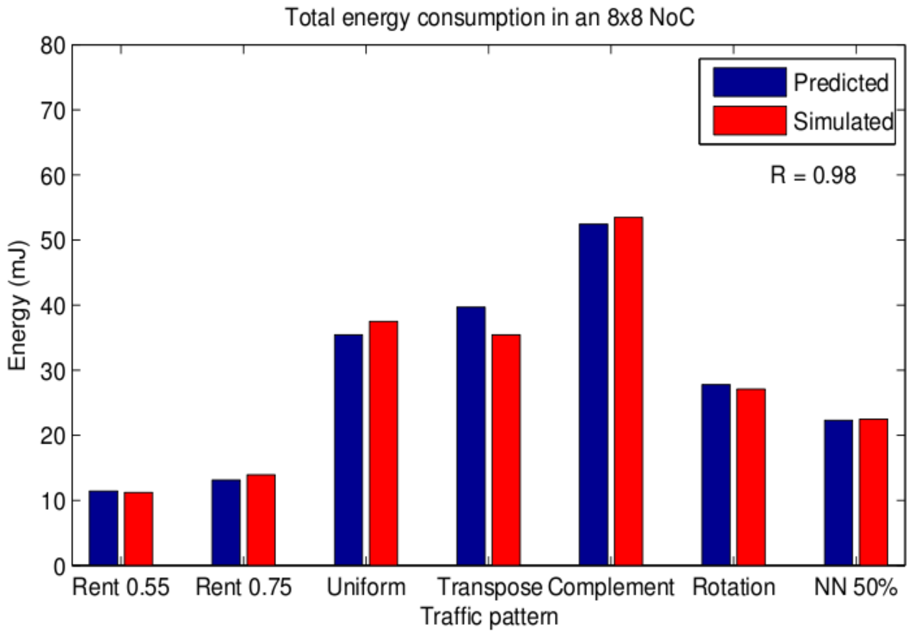
64 cores, 65nm, 1GHz



Results

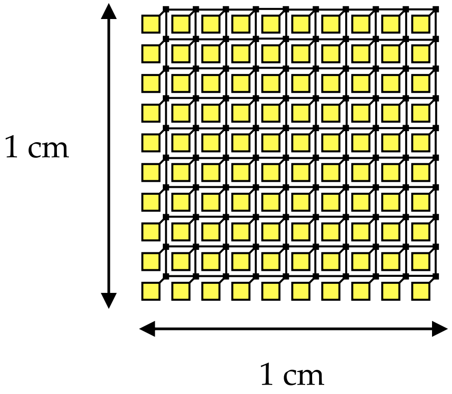
64 cores, 65nm, 1GHz





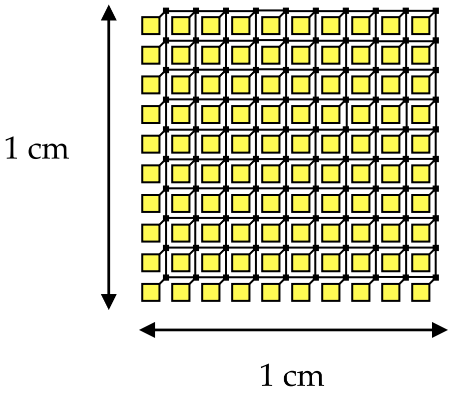
Results

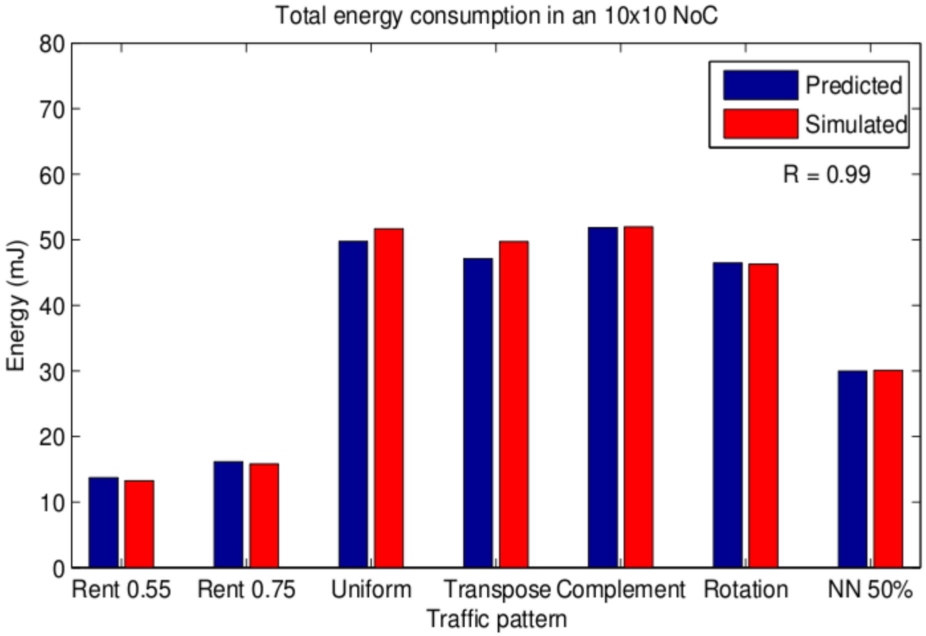
100 cores, 45nm, 3GHz



Results

100 cores, 45nm, 3GHz





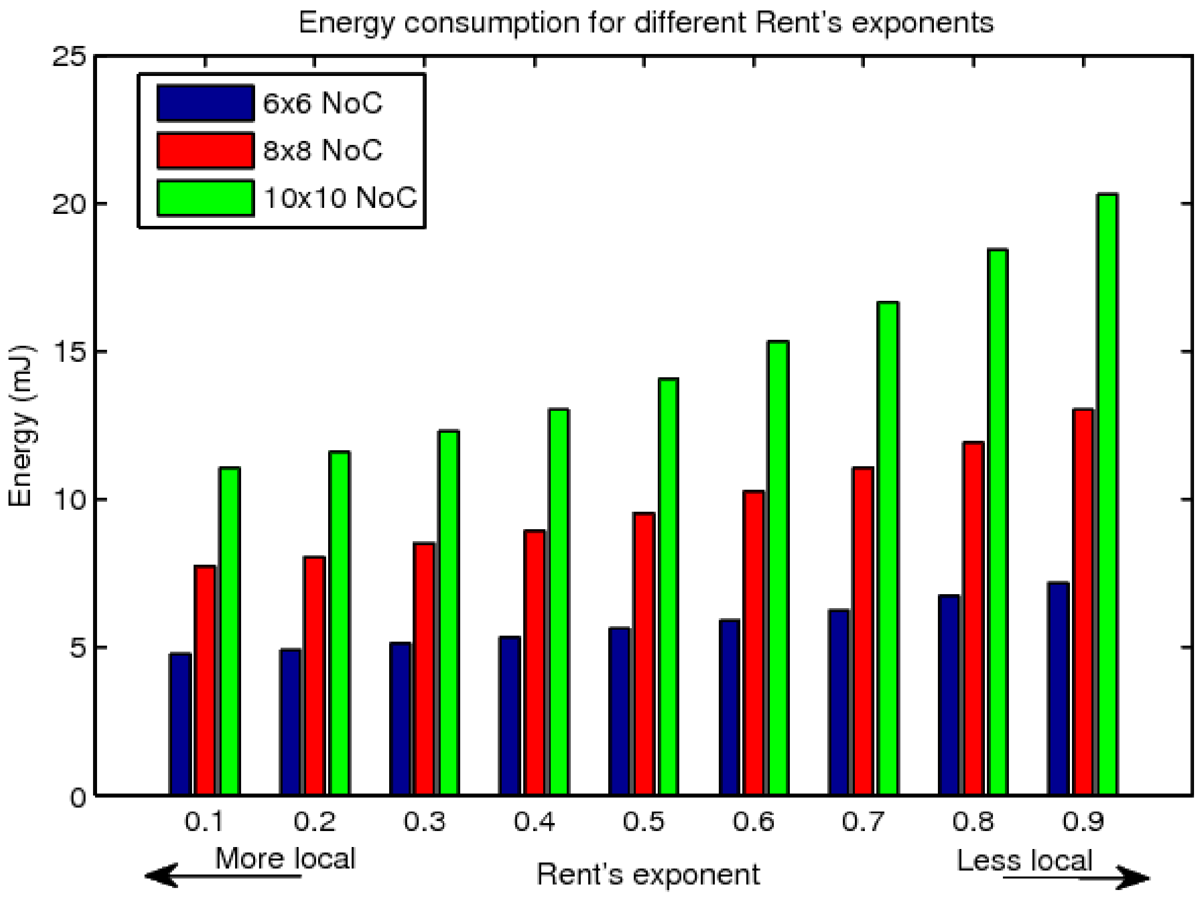
Varying the Rent's exponent

- What is the impact of the Rent's exponent on energy consumption?
  - Our Rentian traffic method can be used to represent a continuum of application complexity scenarios.
  - We will look at NoC energy consumption as a function of the Rent's exponent and size of the system.

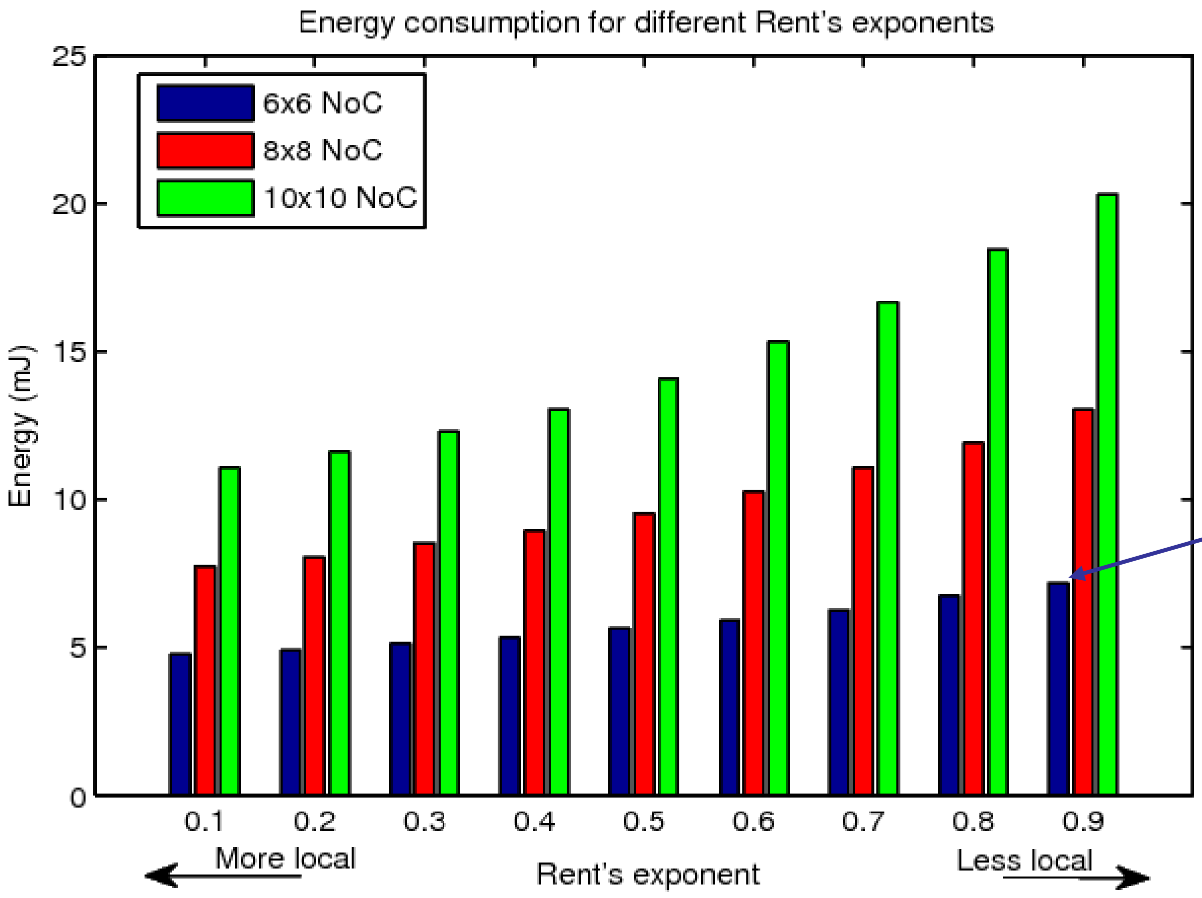
Experiments

- Generate Rentian traffic for $0.1 \leq p \leq 0.9$
- Simulate and measure energy consumption for 6x6, 8x8, and 10x10 NoCs

Results

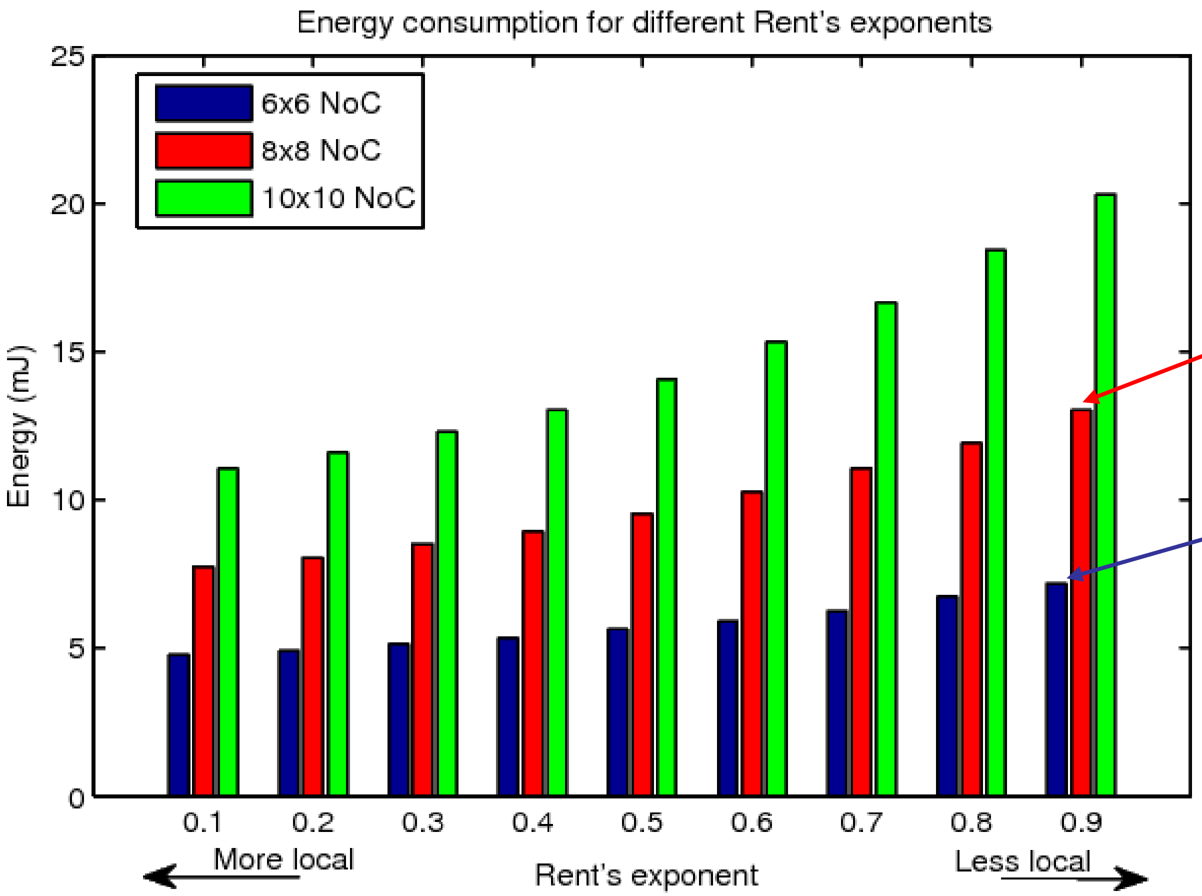

Energy consumption for different Rent's exponents
Energy (mJ)
25
20
15
10
5
0
6x6 NoC
8x8 NoC
10x10 NoC
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
More local
Rent's exponent
Less local

Results


Energy consumption for different Rent's exponents
6x6 NoC
8x8 NoC
10x10 NoC
Energy (mJ)
0
5
10
15
20
25
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
More local
Rent's exponent
Less local

51%

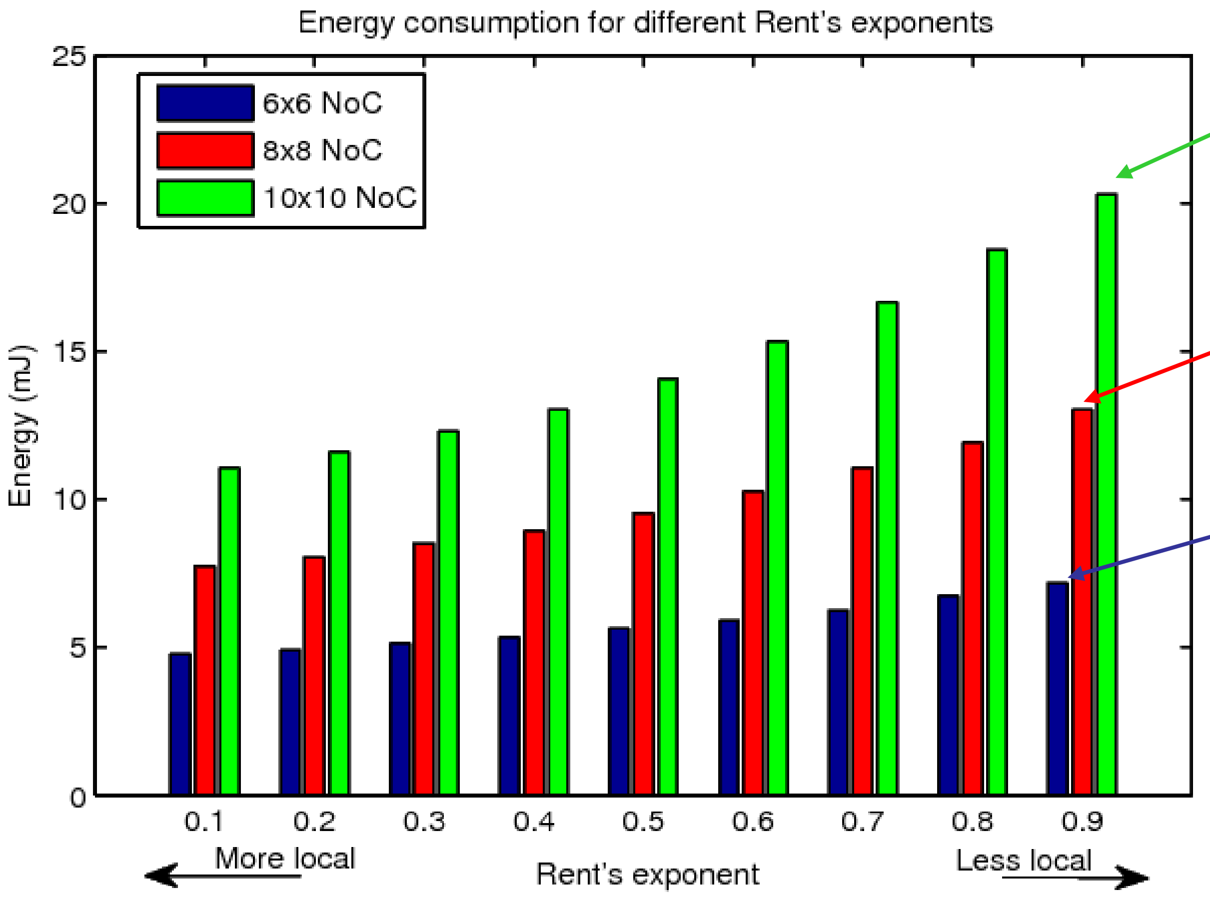
Results



68%

51%

Results



83%

68%

51%

Conclusions

- The CPD is a simple model of communication locality with many potential applications in NoC analysis and design.

Conclusions

- The CPD is a simple model of communication locality with many potential applications in NoC analysis and design.
- We proposed a method based on Rent’s rule that produces synthetic traffic with more realistic CPD.

Conclusions

- The CPD is a simple model of communication locality with many potential applications in NoC analysis and design.
- We proposed a method based on Rent's rule that produces synthetic traffic with more realistic CPD.
- We proposed an energy prediction model based on the CPD with excellent results. <u>For traffic that follows Rent's rule, energy can be estimated directly from the Rent's exponent</u>!

Conclusions

- The CPD is a simple model of communication locality with many potential applications in NoC analysis and design.
- We proposed a method based on Rent’s rule that produces synthetic traffic with more realistic CPD.
- We proposed an energy prediction model based on the CPD with excellent results. For traffic that follows Rent’s rule, energy can be estimated directly from the Rent’s exponent!
- Using our Rent’s rule traffic model, we verified that communication locality has a large (non-linear) impact on energy consumption. This impact will be higher for larger systems.

References

- **Davis *et al*. (1998)** Stochastic wire length distribution for Gigascale Integration (GSI) – Part I: Derivation and validation, *IEEE Transactions on Electron Devices*, VOL. 45, NO. 3, MARCH 1998.
- **Greenfield *et al*. (2007)** Implications of Rent’s rule for NoC design and its fault-tolerance, *Proceedings of the First International Symposium on Networks-on-Chip* (NOCS'07).
- **Heirman *et al*. (2008)**, Rent's rule and parallel programs: characterizing network traffic behavior, *Proceedings of the 2008 international workshop on System level interconnect prediction* (SLIP’08).

Acknowledgements

- National Science Foundation, grants CCF 0621900, CCR 0331580, and SHF 0905236
- Air Force Office of Scientific Research MURI grant FA9550-07-1-0532
- US Department of Energy, Office of Science, grant DE-SC0002113
- Santa Fe Institute
- Program of Interdisciplinary Biological and Biomedical Sciences (PiBBS)