RUPRECHT-KARLS-
UNIVERSITÄT
HEIDELBERG

Analyzing the Energy (Dis-)Proportionality of Scalable Interconnection Networks

Felix Zahn, Pedro Yebenes, Steffen Lammel,
Pedro J. Garcia, Holger Fröning

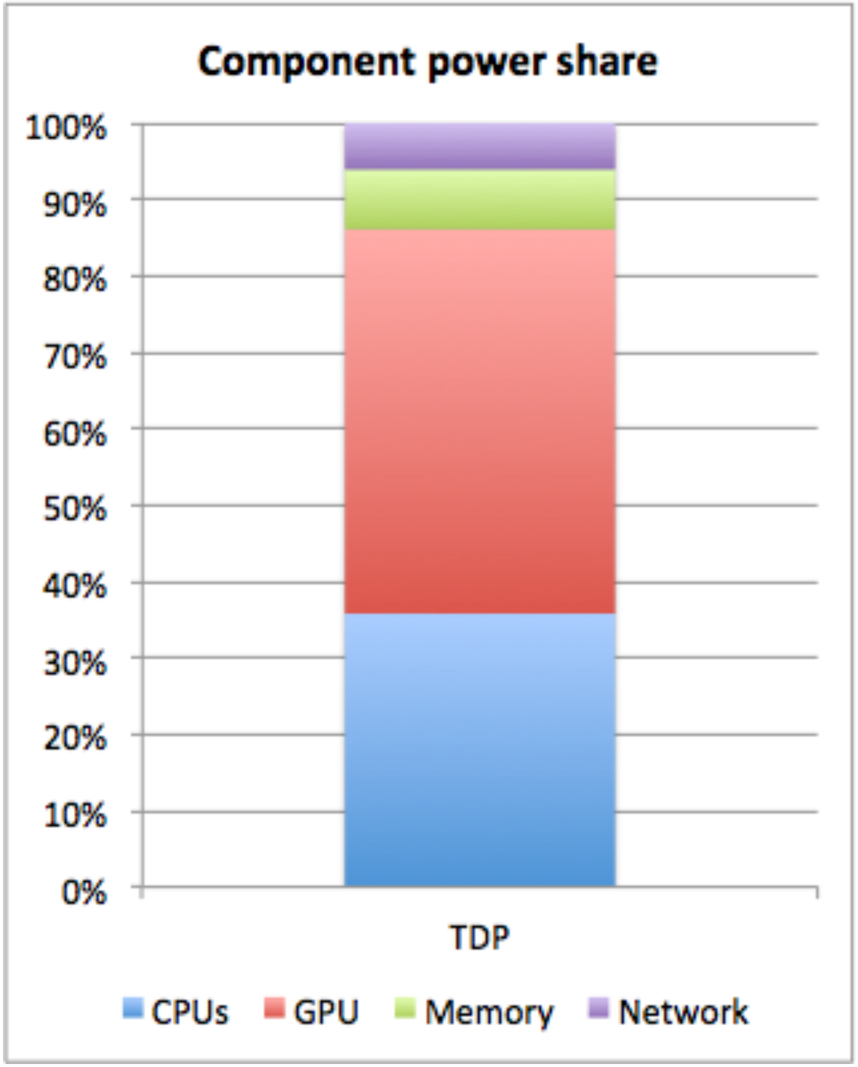
2nd IEEE International Workshop on High-Performance
Interconnection Networks in the Exascale and Big-Data Era

Barcelona, 12th March, 2016

RUPRECHT-KARLS-
UNIVERSITÄT
HEIDELBERG

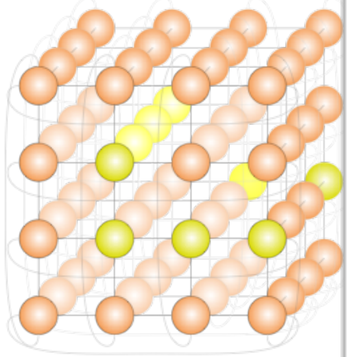
Does network power matter at all?

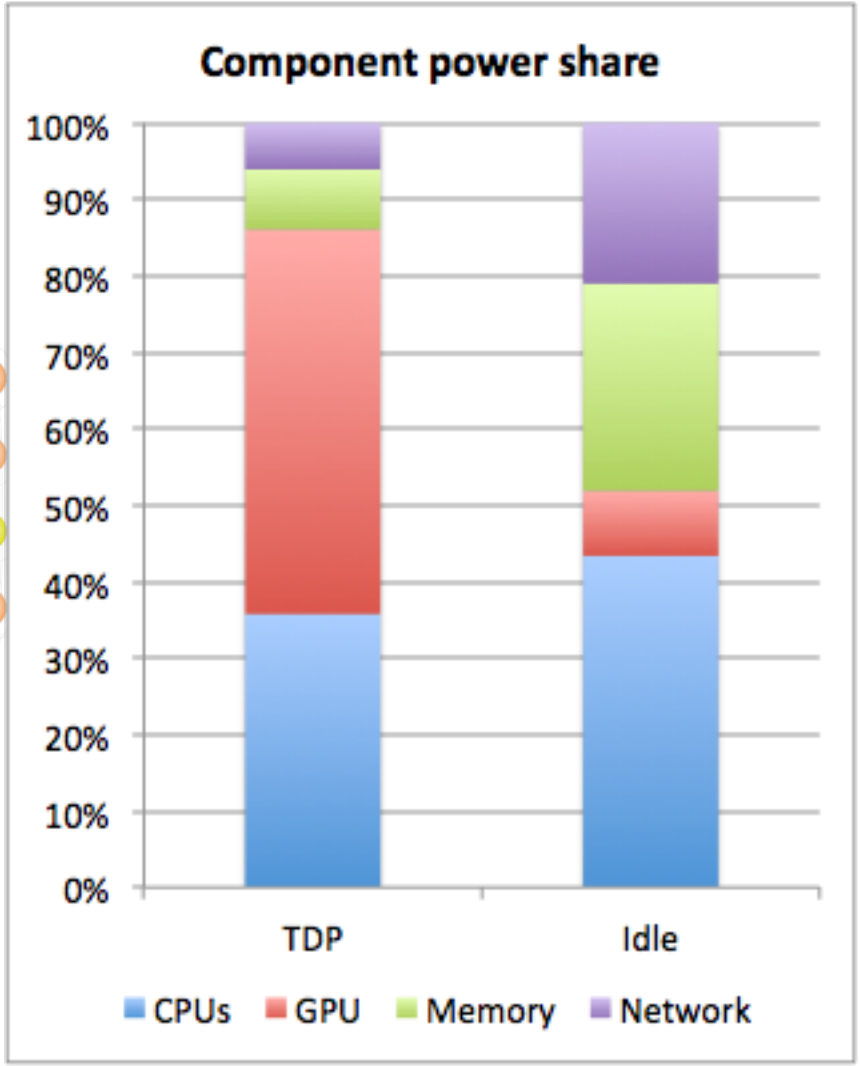
- Pitfall: don’t make assumptions based on maximum power ratings
  - At TDP, processors outshine anything
  - But are processors always operating at 100% load?
  - **Energy-proportional**: at x% load, a component consumes only x% power



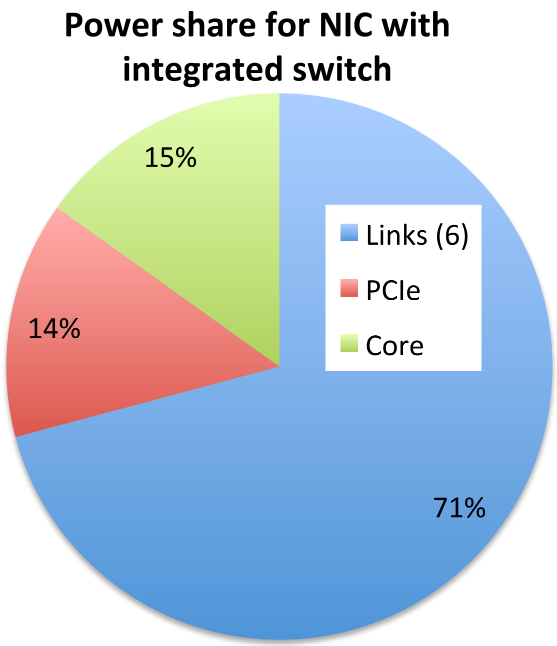
It does!

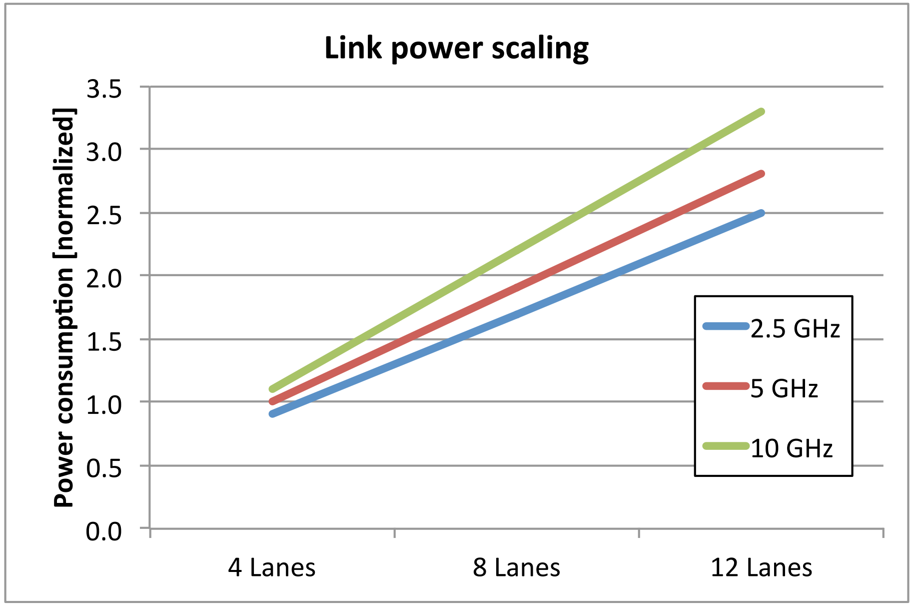
- System power
  - Scalable energy-efficient network
  - Direct network, integrated switches
- Dynamic range of components
- Many memory-bound applications
  - E.g., emerging integer applications (R. Murphy, Sandia) & graph computations
    - DFS & BFS
    - Connected Components
    - Isomorphism
    - Shortest Path
    - Graph Partitioning
    - BLAST (alignment search)
    - zChaff (satisfiability)
- Exception: compute-bound applications with perfect overlap





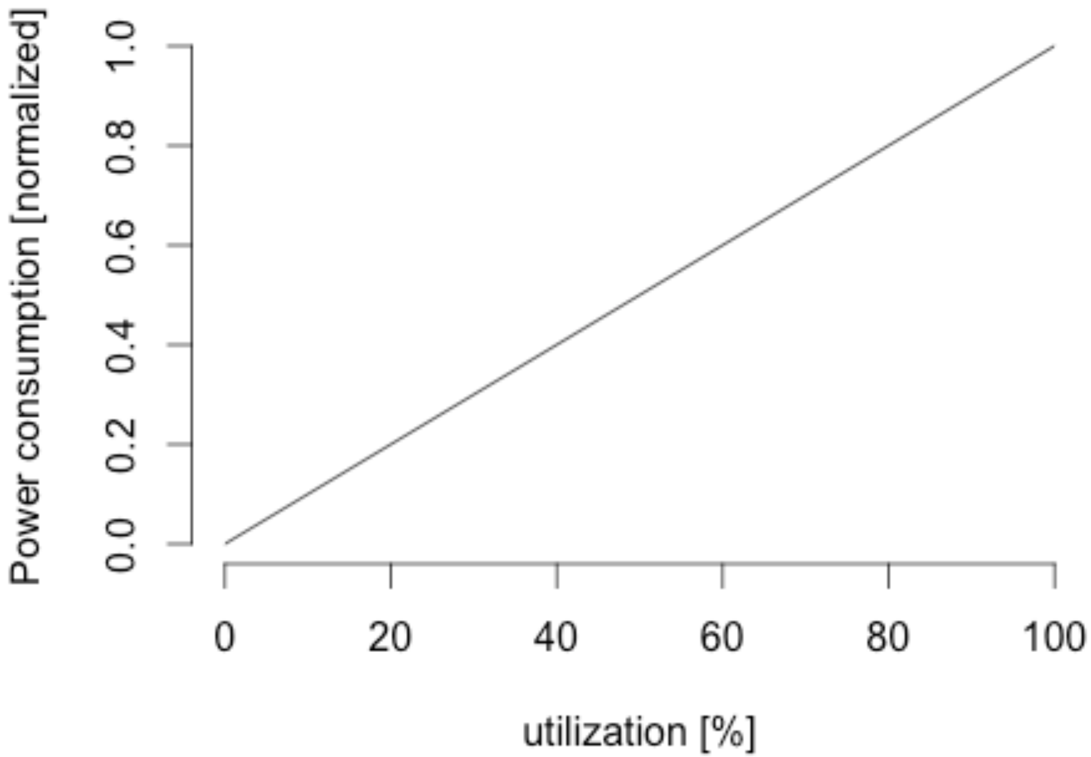
Motivation



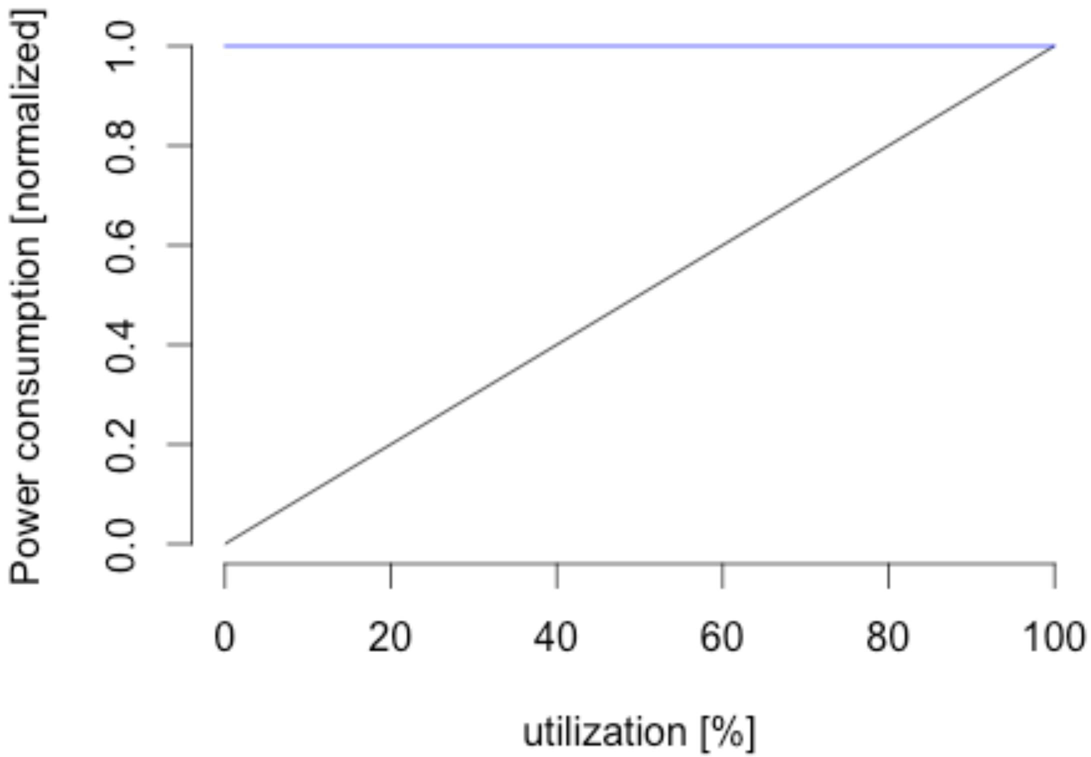


- Serialization technology dominates power consumption
  - Clock recovery, high frequency, equalization, pre-emphasis, …
- It is link width that matters, not frequency
  - CML = **Current** Mode Logic
  - Linear scaling for 10GHz case
  - Frequency dependent part is CMOS only

Energy-proportionality?

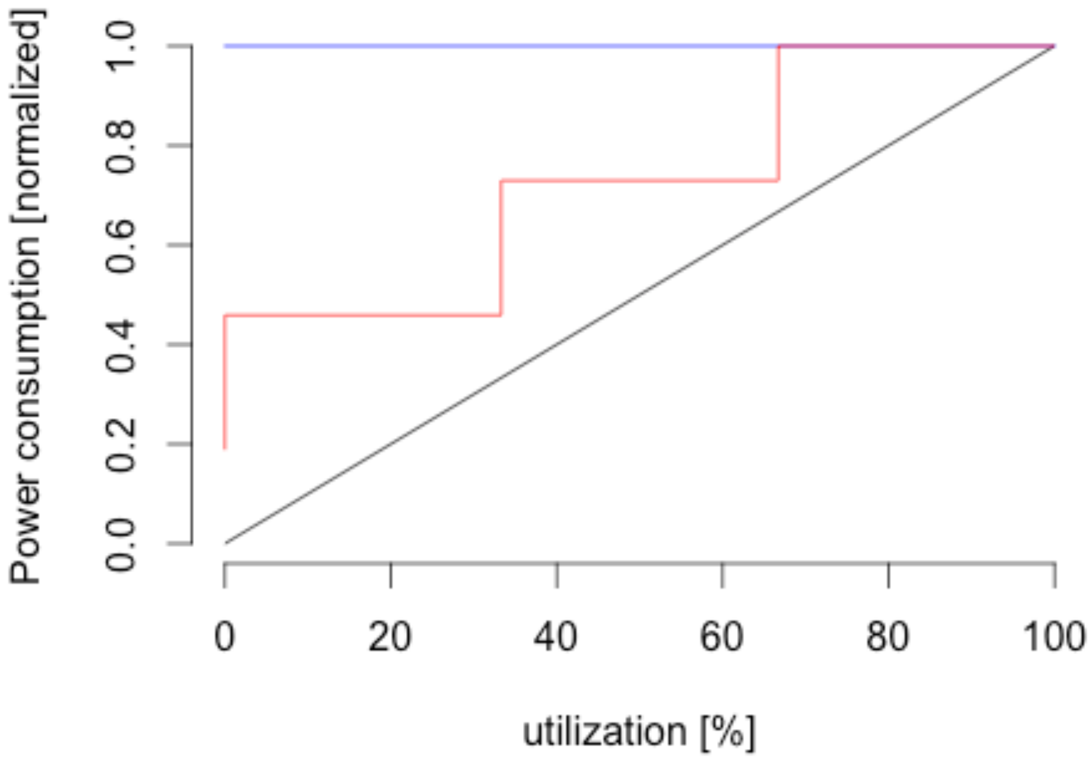


Energy-proportionality vs. today's interconnects



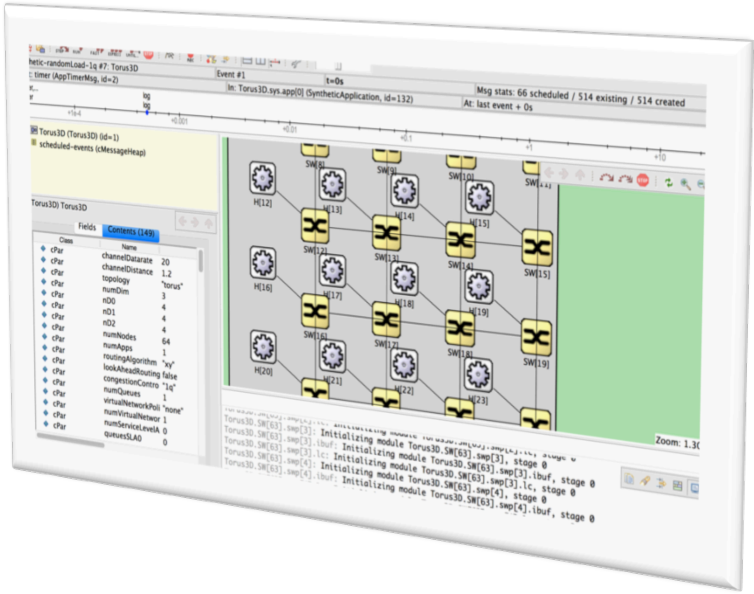
Energy-proportionality – today's possibilities

- Energy/data: 34.4 pJ/bit (65nm TSMC-produced Serializers only)



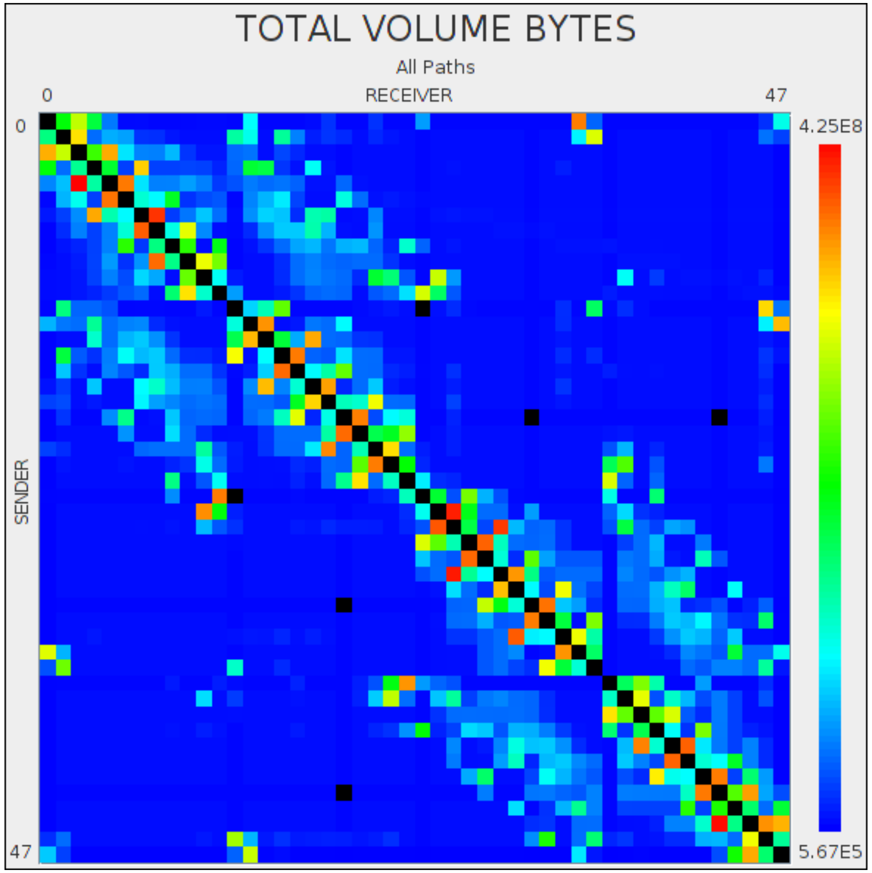
Concept

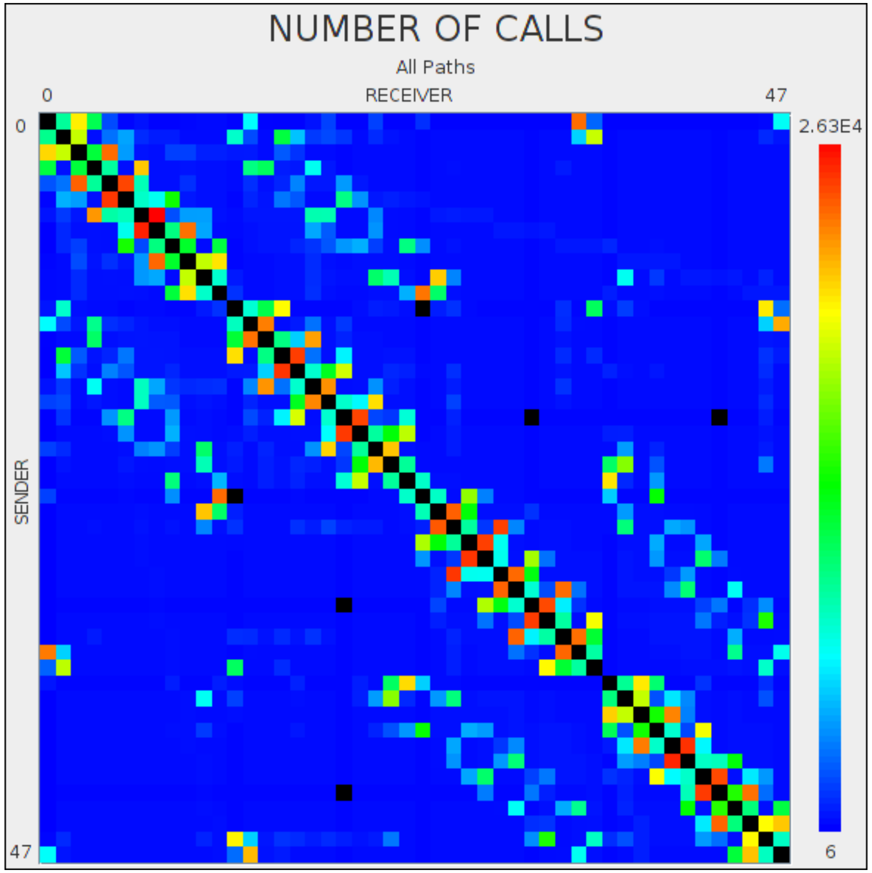
- Application-dependent potential for energy saving
  - Best case: no transition time, links switched on/off depending on whether they are busy or idling
  - Worst case (common case today): all links run with full power
- Setup: OMNeT++-based simulator
  - 3D Torus topology
  - 64 nodes
  - XYZ-dimension-order routing



Workloads - NAMD

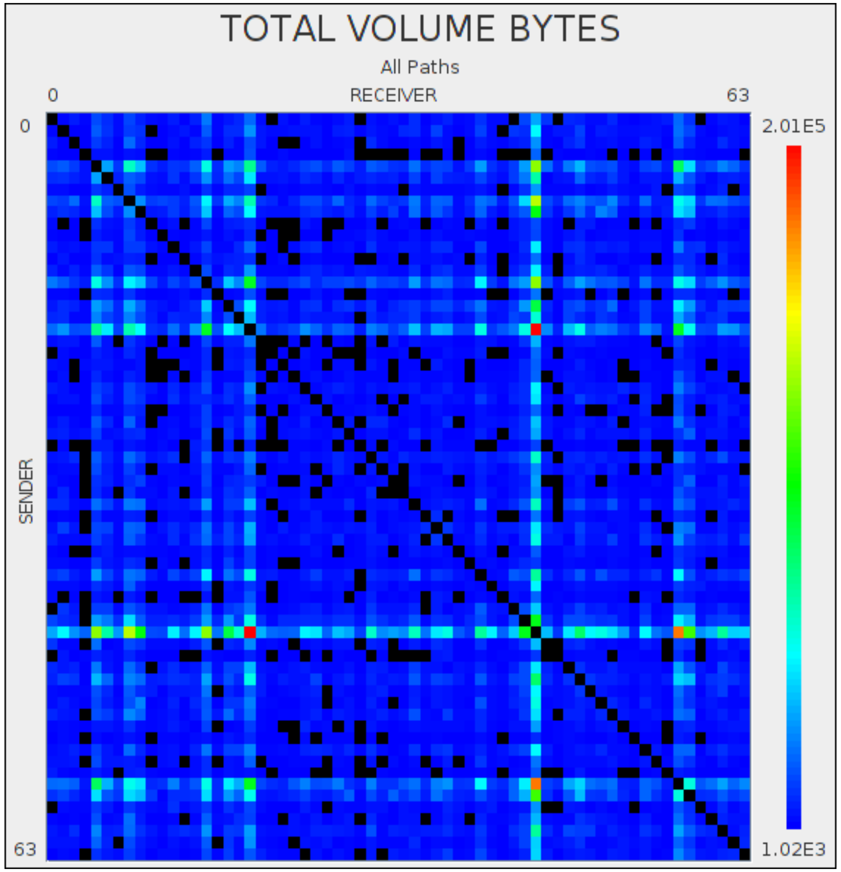
- Satellite Tobacco Mosaic Virus (STMV), 64 MPI tasks

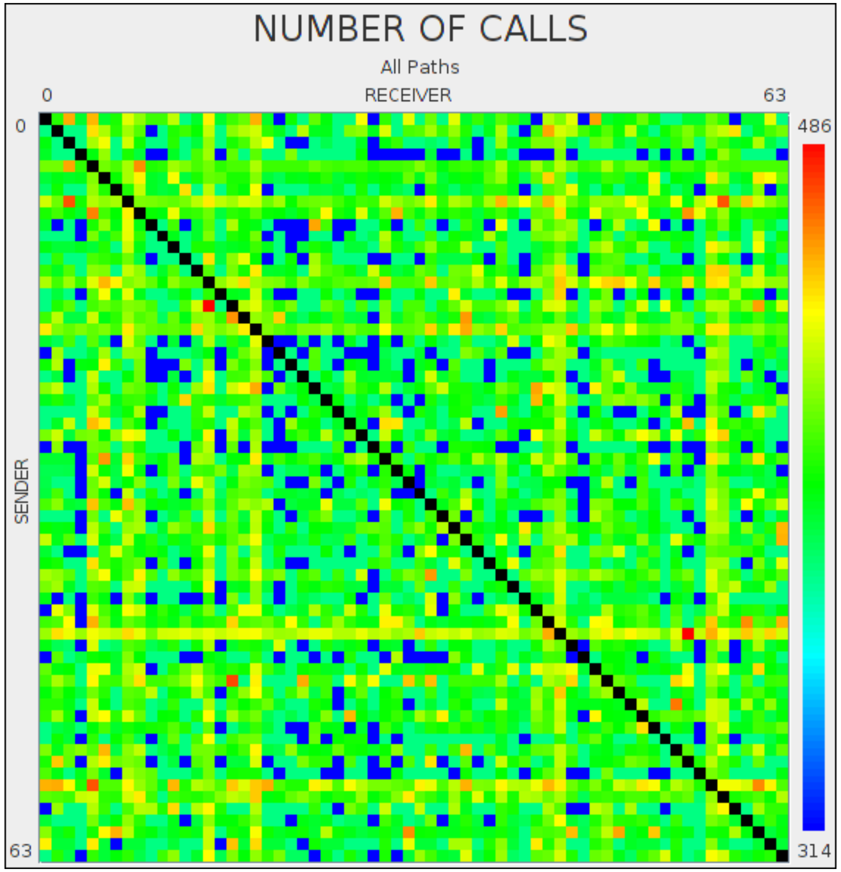




Workloads – Graph500

- scale factor = 12, edge factor = 16, 64 MPI tasks

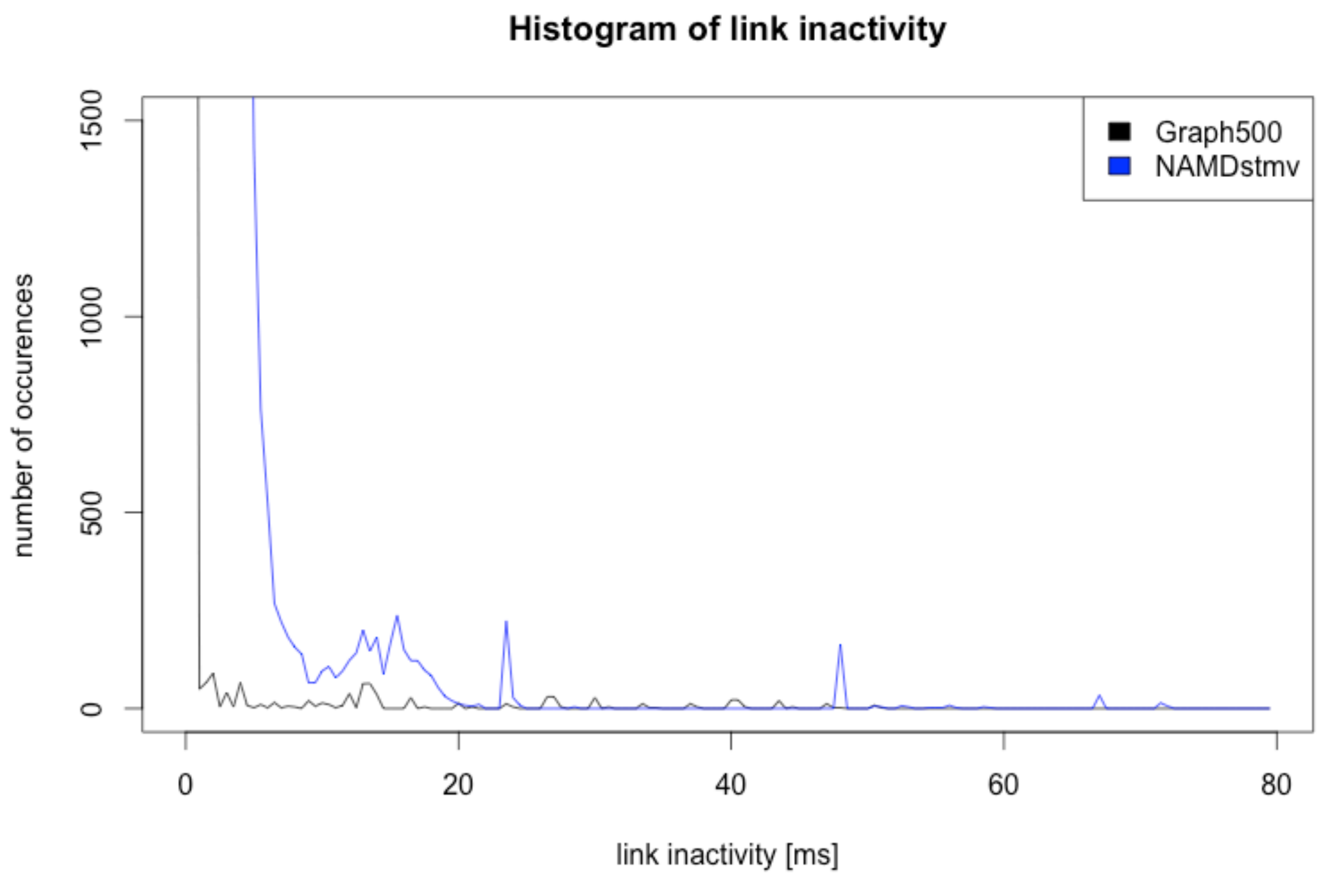




RUPRECHT-KARLS-
UNIVERSITÄT
HEIDELBERG

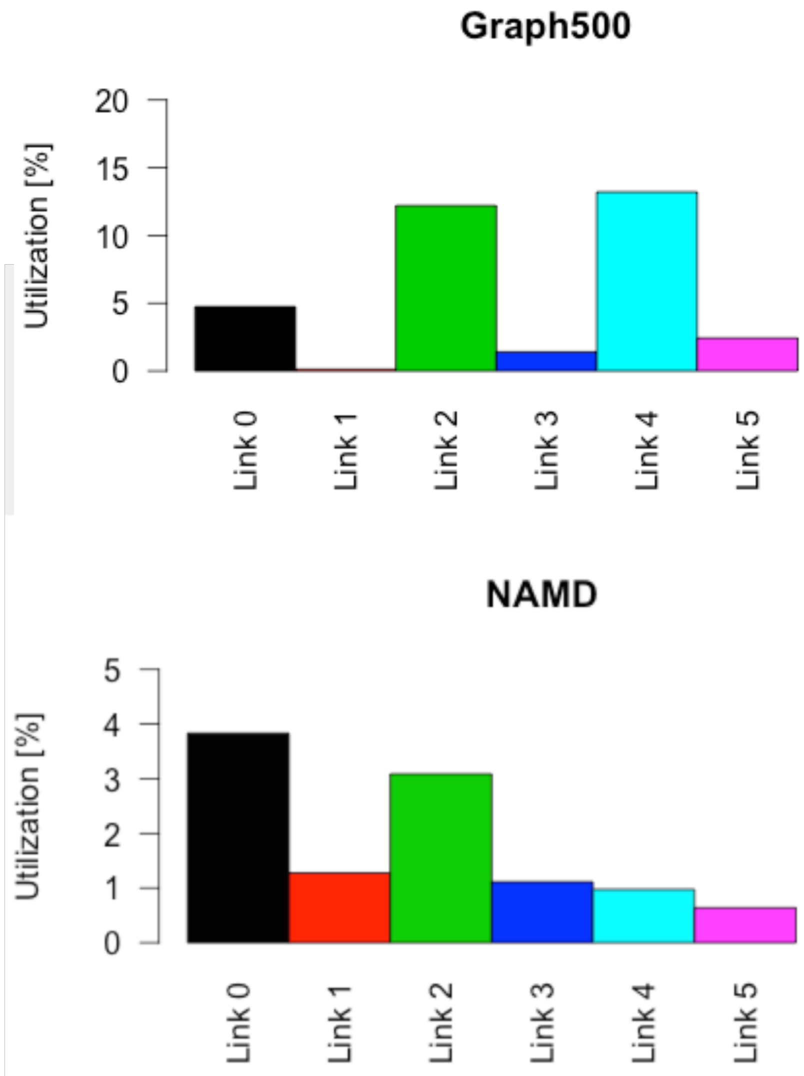
Results – Link inactivity

- Execution time: NAMD 3449 ms, Graph500 95.2 ms

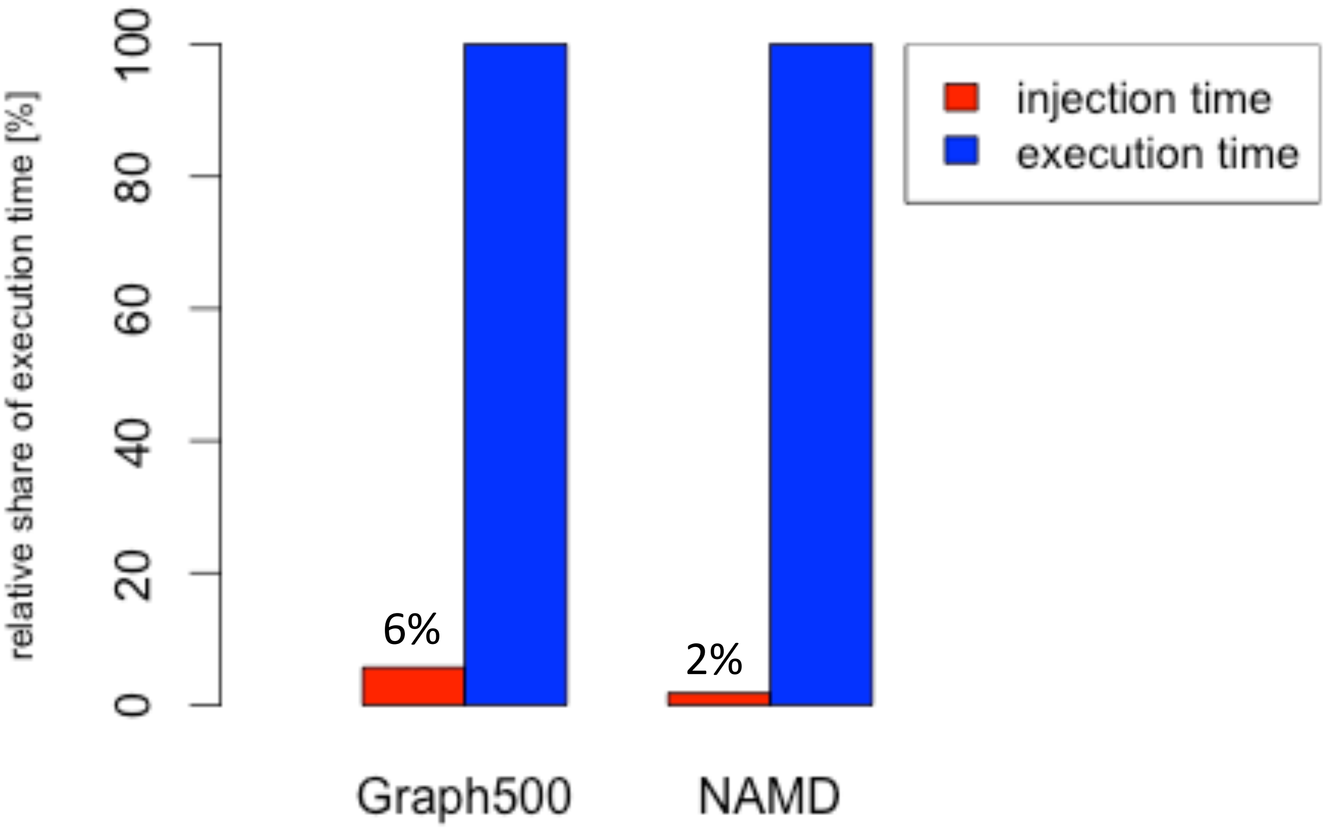


Results – Utilization

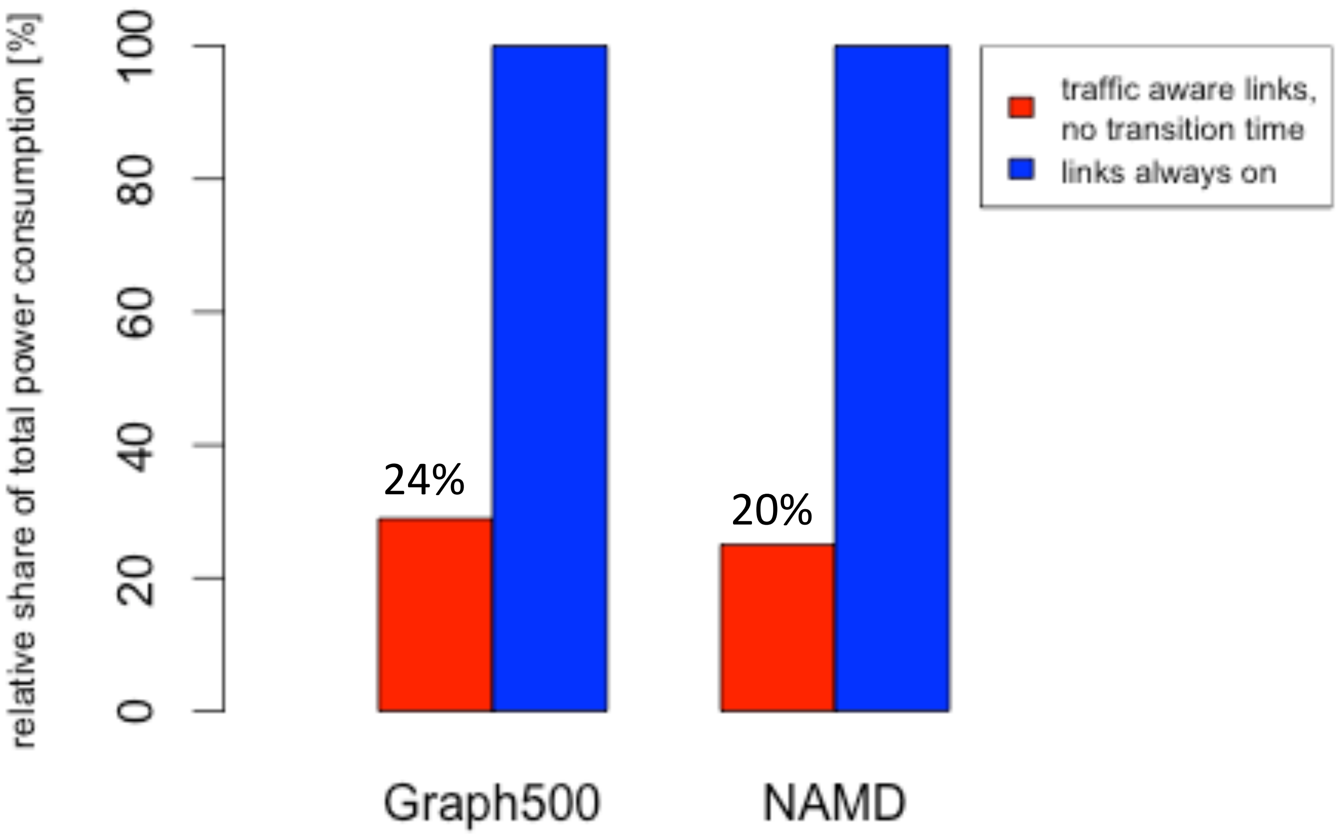
- Link utilization is highly volatile due to:
  - Application
  - Dimension
  - Routing algorithm



Results – Idle time



Results – Power saving potentials



Discussion

- Network power matters
  - Today's interconnection networks contribute up to 20% to the total power consumption of large computing systems
- Interconnects are highly energy dis-proportional
  - But they provide the opportunity to implement power saving strategies
- Huge potential power savings
  - An optimal and fully energy-proportional NIC would save about 75% of the network power

Outlook

- Understanding energy consumption
  - Power saving possibilities depend highly on workloads
  - Communication for HPC applications is complex to model => trace based network simulation
- Analysis of different hardware parameter
  - Transition time, granularity of link width, etc.

  => Useful input on design decisions for future hardware
- Network energy model
  - Predicting energy consumed by the network based on communication characteristics

Credits

Discussions: Maximilian Thürmer, Markus Müller, Benjamin Klenk, Alexander Matz, Daniel Schlegel (Heidelberg University), Sébastien Rumley (Columbia University), Francisco Andujar, Jesus Escudero, Juan Villar (Universidad de Castilla-La Mancha)

Current main interactions

UCLM
UNIVERSIDAD DE CASTILLA~LA MANCHA

COLUMBIA UNIVERSITY

Thank you!

Questions?