Planar Wireless NoC Architectures

Partha Pratim Pande and Deukhyoun Heo
School of Electrical Engineering and Computer Science

Additional Contributors: Amlan Ganguly, Sujay Deb, Paul Wettin, Jacob Murray, Ryan Kim and Turbo Majumder

8th International Symposium on Networks-on-Chip
Motivation

- Massive multicore processors are enablers for ICT innovations
- Need for holistic power optimization and management
- Energy across the layers

“We need research on how to minimize communication, since energy is largely spent in moving data”

“21st Century Computer Architecture" commissioned by the Computing Community Consortium

NSF Workshop on Cross-Layer Power Optimization and Management (Feb ‘12)
Moving a bit across die

Moore’s Law:

Global on-chip movement of data is very power hungry!!

Novel Interconnect Paradigms for Multicore designs

- **Optical Interconnects**
- **Three Dimensional Integration**
- **Wireless/RF Interconnects**
- **High Bandwidth and Low Energy Dissipation**
The mm-wave Small World Wireless Network-on-Chip (mSWNoC)

- Use of on-chip wireless links working in 10-100 GHz range
  - High bandwidth: 10’s of Gbps
  - Latency: True speed-of-light
  - Long distance: ~ 5 mm - 30 mm
  - No physical interconnect layout is necessary

- Reduce latency and energy dissipation in communication

A CMOS Compatible Solution!!
Outline

- Background
- Wireless NoC
  - Network Architecture
  - Communication Resource Management
  - Physical Layer
    - On-Chip Antenna Design
    - Transceiver
- Performance evaluation
- Power & Thermal Management
Classification of Wireless NoC Architectures

- Mesh topology based
  - Wireless links inserted on top of a mesh
    - Wireless NoC based on CMOS UWB technology
    - Wcube architecture with sub-THz wireless links
    - Inter-router wireless scalable express channel in the mm-wave frequency range for NoC (iWISE)

- Small-World network

Power-Law based Small World Network

- Power-law based connectivity
  \[ P(i, j) = \frac{\; l_{ij}^{-\alpha} f_{ij}^{\beta} \;}{\Sigma_{\forall i} \Sigma_{\forall j} l_{ij}^{-\alpha} f_{ij}^{\beta} } \]

- Many short-range local links
  - Wireline links

- A few long-range shortcuts
  - Utilize wireless
Switch Connectivity

- Non-uniform switch ports
- Average number of connections is 4
- An upper bound is imposed
  - One particular switch should not be unrealistically large

- 64-core system
Routing Algorithm

- Deadlock free routing algorithm
- Wireless shortcuts only taken when it saves hops
- Token flow control to avoid hotspot at the WIs
- Token passing protocol for channel assignment among WIs
MROOTS Routing

- Multiple tree roots (MROOTS)
- Traffic-weighted, minimized hop-count, root-node placement

ALASH Routing

- Adaptive Layered Shortest Path (ALASH) Routing
- Only the shortest physical paths considered
- Avoids deadlock by restricting virtual channel (VC) availability to certain messages
  - Creates subsets (layers) of the network using the VCs
  - Layering function determines layer allocation to paths with knowledge from $f_{ij}$.
  - Through a link dependency graph the algorithm ensures that each layer is free of cycles and therefore deadlock free
ALASH Routing (Cont.)

• Inherently avoids creating temperature hotspots through rerouting
• Uses flit density per link as the relevant rerouting parameter
• Different layering mechanisms
  o Uniform - source-destination pair has an equal opportunity for each layer
  o Virtual - evenly distribute source-destination pairs with large $f_{ij}$ values across the different layers
  o Priority - allocates as many layers as possible to source-destination pairs with high $f_{ij}$
MROOTS vs. ALASH

- ALASH distributes traffic across the network better than MROOTS
- ALASH has a reduced flit/hotspot switch
- This results in a better thermal profile for ALASH depending on benchmarks

---

Millimeter-Wave Transceiver

Transceiver system architecture: On-Off-Keying (OOK) modulation for simplicity and low-power.

- Non-coherent OOK receiver: A well-designed envelope detector eliminates the need for a PLL or VCO at the RX side.
- Bandwidth extension design technique.
- Body-enabled design: forward body-bias and bulk-driven techniques.

• Extending to Sub-THz
Millimeter-Wave TX Circuit Design

- OOK modulator chip photo
  - Size: 400um x 550um (core size: 218um x 138um)
  - Transformer baluns provides single-ended to differential conversion for LO and RF ports
Zig-zag Antenna

- 3dB bandwidth of 16 GHz with almost omni-directional response for antennas separated by 20mm
- Three non-overlapping channels
Millimeter-Wave Chip Test Environment

Probed chip under microscope

Agilent 10MHz to 67GHz PNA-X

Mm-wave headers extend frequency range to 110GHz

Cascade Summit 11000 probe station
Performance Evaluation

- Architecture under test
  - 64 core-based system
  - 32-bit flit width
  - Packet Size: 64 flits
  - Die area: 20mm X 20mm
  - Latency, Energy and Thermal profiles
Design Constraints

- Limited Wireless Interfaces (WIs)
- On-chip wireless nodes have associated overhead
  - Transceiver
  - Modulator/Demodulator
  - Antennas

How to efficiently distribute the wireless resources?
Optimum number of Wireless Nodes

- More WIs improve connectivity
- Increases token return period

![Graph showing Bandwidth (Tbps) and Energy Dissipation (nJ) vs. Number of WIs]
Latency Characteristics

![Graph of Average Packet Latency (Cycles) for different benchmarks and NoC topologies]
Energy Dissipation

Normalized Total Network Energy

- FFT
- RADIX
- LU
- CANNEAL
- BODYTRACK
- VIPS
- DEDUP
- SWAPTION
- FLUIDANIMATE
- FREQMINE

Mesh | MROOTS | SWNoC | ALASH | MROOTS | mSWNoC | ALASH
Average Temperature Reduction

Power and Thermal Management

- mSWNoC is an enabling architecture for cross-layer power management
- Thermal management through routing
- No latency penalty compared to a standard mesh

Opens up opportunities for Joint DTM & DVFS in mSWNoC considering both the processing cores and the network

DVFS

- Overall energy dissipation is still dominated by wireline links
- Fine tune voltage and frequency depending on traffic patterns
- History-based DVFS is determined by link utilization
- Utilization determined per link on wireline links

Wireline Link Utilization

- Less utilization of wireline links
- Enables more opportunity for DVFS
VFI-based design

Wireless shortcuts can be used for inter-VFI control and communication

Courtesy: Prof. Radu Marculescu, CMU
Conclusions

- NoC is a reality
  - Limitations: performance & energy
- Alternative interconnect technology
  - High bandwidth, low power
  - Long distance shortcuts
- Adopting nature-inspired topologies
  - Optimization of the network
  - Identify application areas
  - CMOS compatibility
  - A holistic approach considering performance, energy, reliability, and thermal issues is essential
Acknowledgements

◆ Collaborators
  • Behrooz Shirazi, WSU
  • Deukhyoun Heo, WSU
  • Benjamin Belzer, WSU
  • Radu Marculescu, CMU
  • Diana Marculescu, CMU
  • Manos M. Tentzeris, GIT
  • Christof Teuscher, PSU
  • Alireza Nojeh, UBC
◆ Graduate students
◆ Funding from NSF and DOD
Thank you