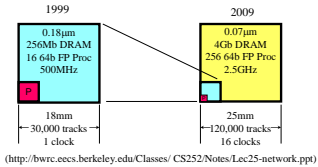


Communication Latency Aware Low Power NoC Synthesis Through Topology Generation and Wire Style Optimization

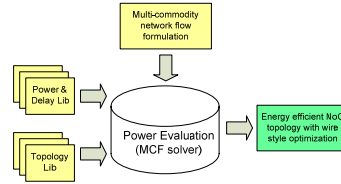
Yuanfang Hu, Yi Zhu, Hongyu Chen, Chung-Kuan Cheng @ UC San Diego VLSI lab

Motivation

- More processing cores are put on a single chip
- On-chip interconnect becomes slower as technology scaling down



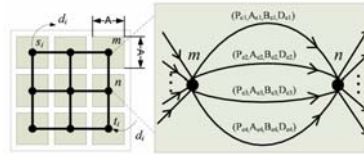
Design Flow



Part I: Multi-commodity Network Flow Formulation

Notations

- Flow graph $G=(V,E)$
- d_j : Commodity j between pair of nodes s_j and t_j
- A : Routing resources on X and Y dimension of chip
- (P_e, A_e, B_e, D_e) : wire style parameters
- f_p : flow on path p



Formulation

$$\begin{aligned} \text{Min : } & \sum_{j=1}^k \sum_{p \in P_j} \sum_{e \in E} f(p) \cdot P_e \\ \text{s.t. } & \sum_{j=1}^k \sum_{p \in P_j} \sum_{e \in E} f(p) \cdot D_e \leq LT \\ & \forall 1 \leq j \leq k : \sum_{p \in P_j} f(p) \geq d_j \\ & \forall q : \sum_{e \in Grid(q)} \sum_{p \in P} f(p) \leq A(q) \\ & \forall p : f(p) \geq 0 \end{aligned}$$

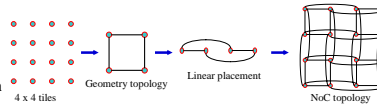
Part II: Topology Library Generation

Regular topologies

- Each row and column have identical connections

Generation

- Step 1: Generate topology on n nodes
- Step 2: Enumerate linear placements on a row/column
- Step 3: Duplicate placements to all rows/columns



Routers

- 0.18um technology node, Using Orion dynamic power simulator
- 1GHz frequency, 4-flit buffer size, 128-bit flit size

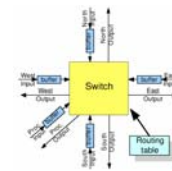
ports	2	3	4	5	6	7	8
P_r (pJ/bit)	0.22	0.33	0.44	0.55	0.66	0.78	0.90
D_r (ns)	0.599	0.662	0.709	0.756	0.788	0.819	0.835

Wires

- 0.18um technology nodes, min global pitch is 1.44um
- Unit wire length (2mm), power and delay of RC wires are proportional to wire length, power and delay of T-line have setup cost: $P(\text{setup}) = 4.4\text{pJ/bit}$, $D(\text{setup}) = 50\text{ps}$

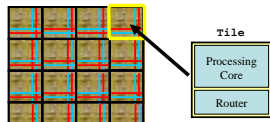
wire type	RC-1x	RC-2x	RC-4x	T-line
P_w (pJ/bit)	2.68	2.15	1.99	0.15
D_w (ns)	0.127	0.112	0.100	0.020

Part III: Power and Delay Library Generation



Solution: Networks-on-Chip (NoCs)

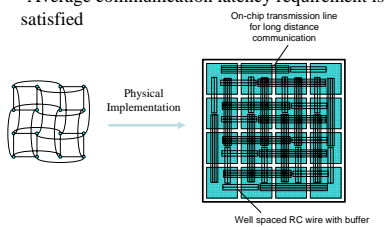
- Tackle multi-cycle signal propagation
- Structure global wires to reduce crosstalk
- Enable the use of aggressive signal circuits
- Efficiently shares on-chip wire resources



(http://cag-www.lcs.mit.edu/raw/)

Problem Statement

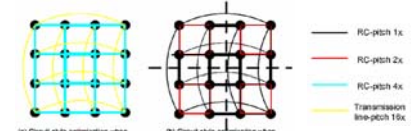
- Given
 - n by n tiles, a library of interconnect wire components
- Input
 - Communication demand matrix [src x dest]
- Output
 - Low power NoC topology and its physical implementation (wire type and capacity)
- Constraints
 - The cross section wiring area cannot exceed the chip dimension
 - Average communication latency requirement is satisfied



Experimental Results

Wire Style Optimization

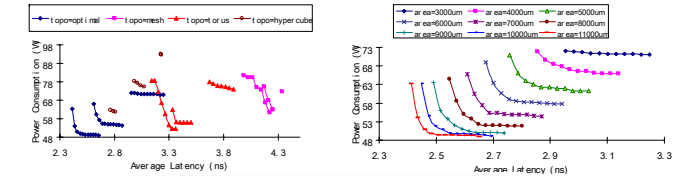
- Given topology: 4x4 torus NoC
- Without wire style optimization
 - RC wires w/ 1x min global pitch
- With wire style optimization
 - RC wires w/ 1x min global pitch
 - RC wires w/ 2x min global pitch
 - RC wires w/ 4x min global pitch
 - Transmission lines with 16um wire width
- Evenly distributed communication demand
- Up to 34.6% power savings



Comm. (Gb/s)	Power (w/o opt.) (W)	Power (w/ opt.) (W)	Impr. (%)
10	43.0	28.1	34.6
20	82.0	58.4	28.8
30	121	103	14.6
40	160	152	5.10

Topology Selection

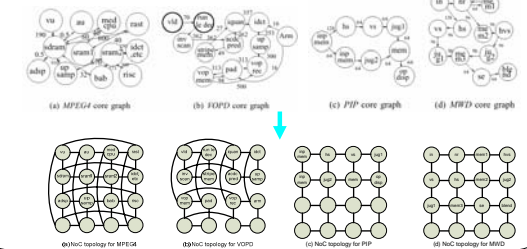
- Various available on-chip resources, evenly distributed communication demand
- Left figure and below table show NoC power and latency comparison among mesh, torus, hypercube and our optimal design
- Right figure Power and latency relations among optimal design under various on-chip resources



area (um)	topo	L (ms)	P (W)	P ² L (W*ms)	Impr. (%)
3000	mesh	4.34	72.7	315.2	26.7
	torus	3.74	76.1	284.7	18.9
	cube	3.23	92.8	299.3	23.0
	optimal	3.25	71.1	230.9	44.5
7000	mesh	4.25	63.0	267.9	44.5
	torus	3.37	56.3	189.6	21.5
	cube	3.04	76.0	231.2	35.6
	optimal	2.69	55.4	148.8	52.1
11000	mesh	4.22	61.2	258.3	52.1
	torus	3.33	52.7	175.3	29.4
	cube	2.76	62.6	173.1	28.5
	optimal	2.48	49.8	123.8	

Video Applications

- Four video applications are mapped to 4x4 NoC, adopting various network topologies



Summary

- We reduce NoC power consumption by simultaneous optimization of network topology and interconnect wire styles, while satisfying communication latency constrains
- Wire style optimization reduces NoC power consumption by up to 34.6%, for 4x4 torus
- Comparing with mesh, torus and hypercube, our optimized design for 8x8 NoC can improve power latency product by up to 52.1%, 29.4%, and 35.6%, respectively