Mixed Integer Linear Programming-based Optimal Topology Synthesis of Cascaded Crossbar Switches

Minje Jun
Sungjoo Yoo
Eui-Young Chung
Contents

- Motivation
- Related Works
- Overview of the Method
- Problem Definition
- Constraints and Formulations
- Experiment
- Conclusion
Motivation

- **On-chip Interconnection Methodology Shifting**
  - Shared bus → Hierarchical bus or Bus matrix
    → Switch-based design

- **Crossbar Switch Sufferings against Its Size**
  - Increase in logic delay
    - Control blocks (decoder, arbiter, etc) grow
    - Results in lower speed
  - Worse then linear increase in silicon area
    - Quadratic increase of # of bus lines
Motivation

- Maximum Freq. of Crossbar Switch against Its Size

Synthesis result of RTL code from AMBA Designer with Samsung 90nm process technology
Motivation

Can achieve faster speed
Can achieve area reduction
Related Works

- Does it need to be fully connected?
  - S. Murali *et al.* and S. Pasricha *et al.* presented synthesis of partially connected bus matrix [1-3]
  - Limitation on modularity and scalability

- Can’t it be shared by multiple masters/slaves?
  - S. Murali *et al.* presented clustering method to minimize the central crossbar switch [4]
  - An extreme case of our work
Related Works

- How about in NoC area?
  - K.Srinivasan et al. presented irregular topology synthesis method using fixed size routers [5]
- J.Yoo et al. first presented cascaded crossbar switch, based on simulated annealing method [6]
  - Does not targeting optimality
Overview of the Synthesis Process

- Communication Requirement of Target System
- Library: Physical Characteristics of Crossbar Switch (frequency, area)
- Proposed Synthesis Flow
- Optimal Cascaded Crossbar Switch Topology
Problem Definition

Communication Requirement Graph CRG

\[ G(V_M, V_S, E) \]

- \( v_m \in V_M \): master node
- \( v_s \in V_S \): slave node
- \( e_{m,s} \in E \): edge between \( v_m \) and \( v_s \)
- \( w(e_{m,s}) \): BW requirement of \( e_{m,s} \)
- \( d(e_{m,s}) \): latency constraint of \( e_{m,s} \)

Crossbar Switch Physical Characteristics

- \( A_{m,s} \): area of \( m \times s \) crossbar switch
- \( F_{m,s} \): maximum freq. of \( m \times s \) crossbar switch
Problem Definition

- Cascaded Crossbar Topology $T(MX,SX,XX)$
  - $MX_{m,x} : 1$ iff master $\nu_m$ is connected to crossbar $x$
  - $SX_{s,x} : 1$ iff slave $\nu_s$ is connected to crossbar $x$
  - $XX_{x',x''} : 1$ iff crossbar $x'$ is connected to crossbar $x''$

- Our problem is
  - to find $T(MX,SX,XX)$ which optimizes the design
Assumptions

- Single clock frequency and data width
  - Bridges for freq. and/or data width conversion already included in IP interfaces

- Single path routing
  - Each $e_{m,s}$ in CRG is accommodated by a single path
  - Productized switch modules [7-8] remain unchanged
  - Still, a master-slave pair can communicate with multiple path by using multiple ports
Topology Feasibility Constraint

○ A master or slave must be connected to one and only one crossbar switch

\[ \forall v_m \in V_M, \sum_{x \in X} MX_{m,x} = 1 \]
\[ \forall v_s \in V_S, \sum_{x \in X} SX_{s,x} = 1 \]

○ A crossbar switch must be

• not used: \( PM_x = PS_x = 0 \)
• or larger than 1x2 or 2x1:

\[ PM_x > 0 \& PS_x > 0 \& PM_x + PS_x > 2 \]
Single Communication Path Constraint

- **Depth-\(n\) path matrix** \(D^n_{m,x_1,x_2,...,x_n,s}\)
  - 1 iff master \(v_m\) is connected to slave \(v_s\) through \(x_1, x_2, ..., x_{n-1}\), and \(x_n\)

\[
D^n_{m,x_1,x_2,...,x_n,s} = MX_{m,x_1} \times \prod_{k=1}^{n-1} XX_{x_k,x_{k+1}} \times SX_{s,x_n}
\]

\[
D^n_{m,x_1,x_2,...,x_n,s} \geq MX_{m,x_1} + \sum_{k=1}^{n-1} XX_{x_k,x_{k+1}} + SX_{s,x_n} - n
\]

\[
D^n_{m,x_1,x_2,...,x_n,s} \leq 1/n \times [MX_{m,x_1} + \sum_{k=1}^{n-1} XX_{x_k,x_{k+1}} + SX_{s,x_n} - 1]
\]
Single Communication Path Constraint

- For a master-slave pair with $w(e_{m,s}) > 0$, there must exist a communication path

$$\sum_{x \in X} D_{m,x,s}^1 + \sum_{x_1,x_2 \in X} D_{m,x_1,x_2,s}^2 + \cdots + \sum_{x_1,x_2,\ldots,x_N \in X} D_{m,x_1,x_2,\ldots,x_N,s}^N = 1$$

- When latency constraint exists, $d(e_{m,s}) = k$

$$\sum_{x \in X} D_{m,x,s}^1 + \sum_{x_1,x_2 \in X} D_{m,x_1,x_2,s}^2 + \cdots + \sum_{x_1,x_2,\ldots,x_k \in X} D_{m,x_1,x_2,\ldots,x_k,s}^k = 1$$
Bandwidth Constraint

- Link between $x_1$ and $x_2$ is loaded
  - by depth-2 connection ($m_1$-$s_1$)
    \[ \varepsilon_{x_1,x_2,m_1,s_1}^2 = w(e_{m_1,s_1}) \times D_{m_1,x_1,x_2,s_1}^2 \]
  - by depth-3 connection ($m_2$-$s_2$)
    \[ \varepsilon_{x_1,x_2,m_2,s_2}^3 = w(e_{m_2,s_2}) \times \left( \sum_{x' \in X} D_{m_2,x_1,x_2,x',s_2}^3 \right) \]
Bandwidth Constraint

- Total weight between $x_1$ and $x_2$ in above example

$$E_{x_1,x_2} = \varepsilon^2_{x_1,x_2,m_1,s_1} + \varepsilon^3_{x_1,x_2,m_2,s_2}$$

- General expression

$$\forall x_1 < x_2 \in X, \ E_{x_1,x_2} = \sum_{k=2}^{N} \sum_{v_m \in V_M} \sum_{v_s \in V_S} \varepsilon^n_{x_1,x_2,m,s}$$

- Bandwidth Constraint

$$\forall x_1 < x_2 \in X, \ E_{x_1,x_2} \leq \text{CostFreq} \times \text{channelwidth}$$

- Crossbar network frequency
- Channel width of the switches
Table Referencing with MILP Variables

- Freq. / Area of crossbar $x$
  
  $F_{PM_x,PS_x} / A_{PM_x,PS_x} \rightarrow \sum_{m=0}^{\mid V_M \mid} \sum_{s=0}^{\mid V_S \mid} F_{m,s} \times K_{x,m,s} / \sum_{m=0}^{\mid V_M \mid} \sum_{s=0}^{\mid V_S \mid} A_{m,s} \times K_{x,m,s}$

- Cannot be directly obtained!!

- Obtain indexing matrix $K_{x,m,s}$
  
  \[ K_{x,m,s} \leq 1 - \beta \{ m - PM_x + \alpha(s - PS_x) \} \]

  $K_{x,m,s} \leq 1 + \beta \{ m - PM_x + \alpha(s - PS_x) \}$

  $\sum_{m=0}^{\mid V_M \mid} \sum_{s=0}^{\mid V_S \mid} K_{x,m,s} = 1$

  $K_{x,m,s}$ is 0 if inside the brace is positive

  $K_{x,m,s}$ is 0 if inside the brace is negative

  $K_{x,m,s}$ can and must be 1 if inside the brace is $0$

  \[ K_{x,m,s} \text{ is less than 1 by } \beta \]

  \[ K_{x,m,s} \text{ is always less than 1 by } \alpha, \text{ thus fractional part} \]

  \[ K_{x,m,s} \text{ is 0 if inside the brace is positive} \]
Experiment Setting

Applications
- App I : Industrial strength SoC (12x4)
- App II : Mpeg4 decoder example (9x3) [9]

Objectives
- Frequency maximization w/o area upper bound
- Frequency maximization w/ area upper bound (30%)
- Area minimization w/o freq. lower bound
- Area minimization w/ freq. lower bound (380Mhz)
Result for App I

- Freq. improvement is up to 37.3%, but with 43.6% area overhead (Objective I)
- Area reduction is up to 12.7%, and with 21.9% freq. improvement (objective III)
Result for App I

Objective I

Objective II

Objective III

Objective IV
Result for App II

- Freq. improvement is up to 22.9%, but with area overhead of 68.4% (Objective I)
- For area minimization, the single crossbar is the best solution
- Freq. lower bound 380Mhz achieved with only 4% area overhead
Result for App II
Conclusion

- We proposed
  - cascaded crossbar switch network
  - using arbitrary sized crossbar switches
  - MILP-based exact topology synthesis method

- Experimental result shows
  - up to 37.3% (12.7%) freq. (area) improvement
  - synthesis time: 15.7 hours (App I) / 0.36 hours (App II) on average

- Future Work
  - Time-efficient heuristic algorithm
  - Adding objective for power consumption
Thank You
References


[7] ARM, \url{www.arm.com}

[8] Sonics Inc., \url{www.sonicstinc.com}