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

Minje Jun

Sungjoo Yoo

Eui-Young Chung

Contents

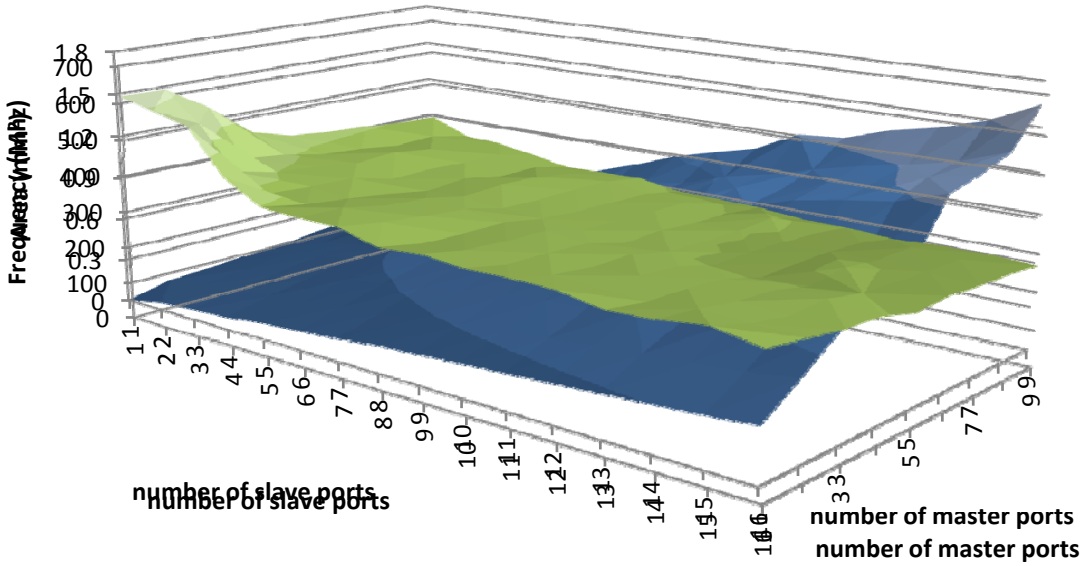
- Motivation
- Related Works
- Overview of the Method
- Problem Definition
- Constraints and Formulations
- Experiment
- Conclusion

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

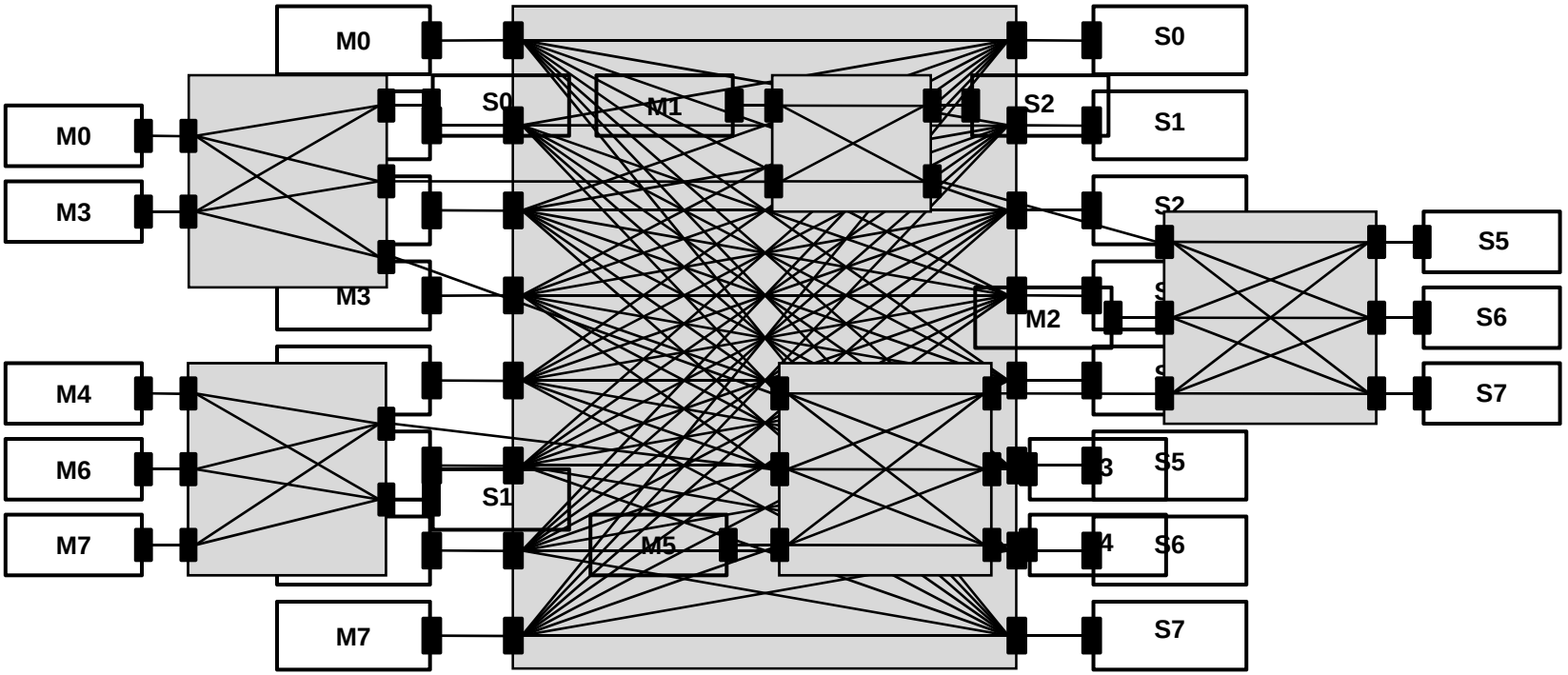
Motivation

- Maximum Frequency Versus Area Against Its Size



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

Motivation



Proposed Gated Shield Crossbar Switch Network

Conventional Single Crossbar Switch Network

Can achieve faster speed

Can achieve area reduction

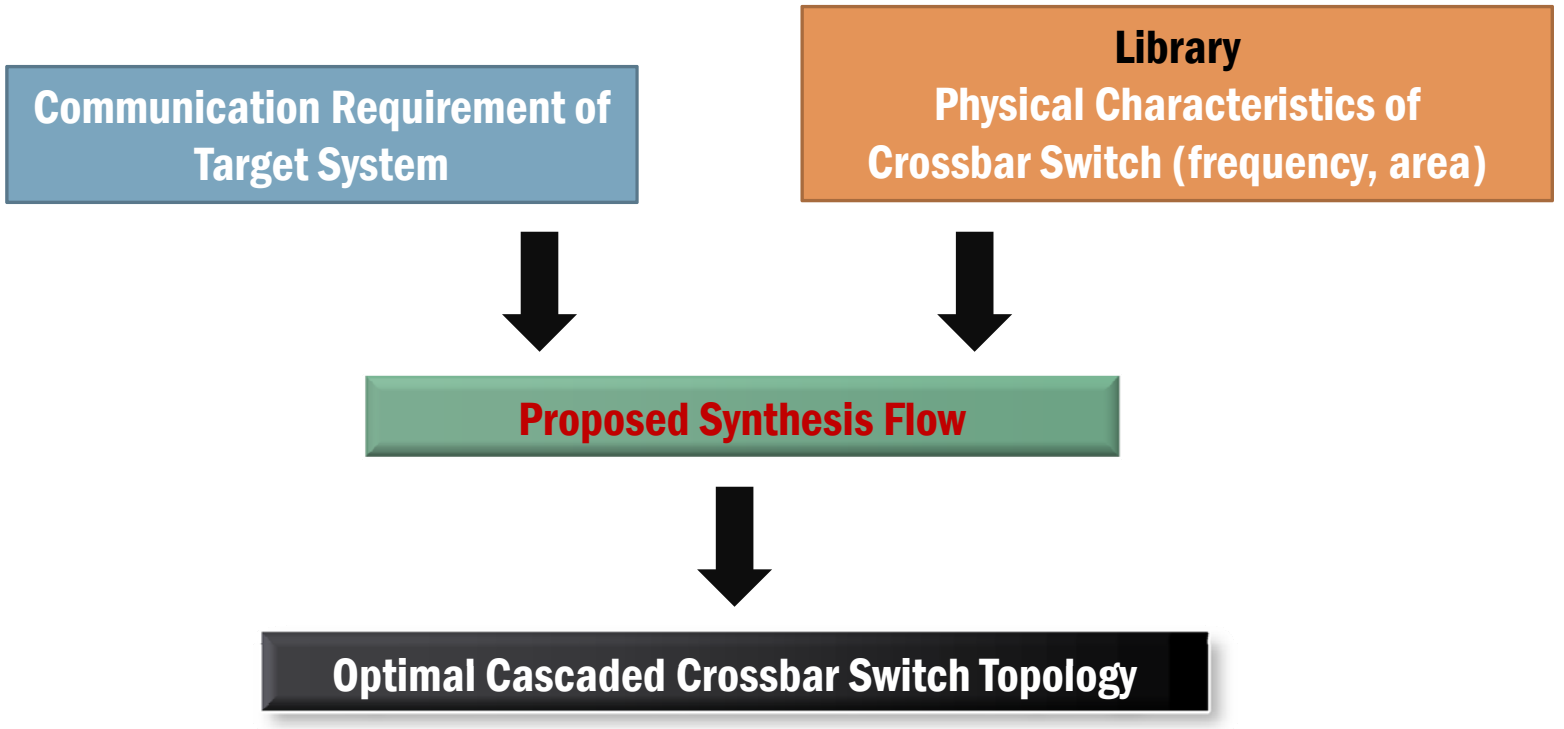
Related Works

- 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 extereme 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 targetting optimality

Overview of the Synthesis Process



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 mxs crossbar switch
  - $F_{m,s}$: maximum freq. of mxs crossbar switch

Problem Definition

- Cascaded Crossbar Topology *T(MX,SX,XX)*
  - $MX_{m,x}$: 1 *iff* master v_m is connected to crossbar x
  - $SX_{s,x}$: 1 *iff* slave v_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** $D^n_{m,x_1,x_2,\ldots x_n,s}$
  - **1 *iff* master v_m is connected to slave v_s through $x_1,x_2,\ldots,x_{n-1}$, and x_n**

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

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

$$D^n_{m,x_1,x_2,\ldots,x_n,s} \leq 1/n \times \left[MX_{m,x_1} + \sum_{k=1}^{n-1} XX_{x_k,x_{k+1}} + SX_{s,x_n} - 1\right]$$

Single Communication Path Constraint

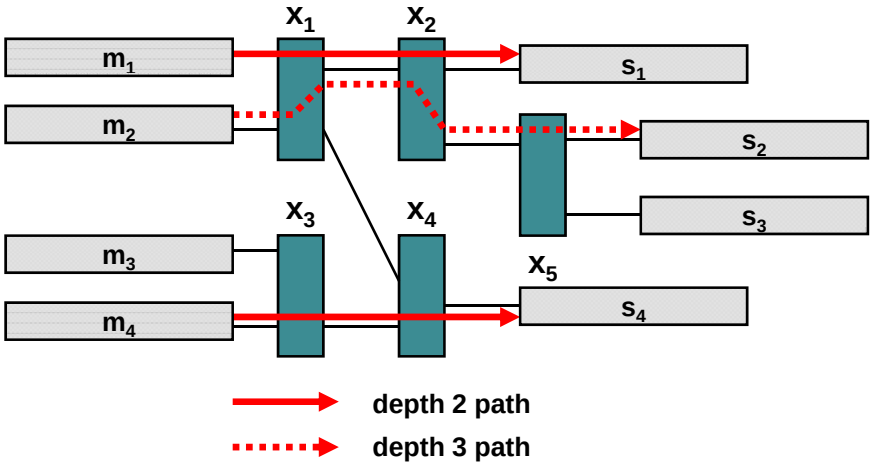
- For a master-slave pair with $w(e_{m,s})>0$, there must exist a communication path

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

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

$$\sum_{x\in X} D^1_{m,x,s} + \sum_{x_1,x_2\in X} D^2_{m,x_1,x_2,s} + \cdots + \sum_{x_1,x_2,\ldots,x_k\in X} D^k_{m,x_1,x_2,\ldots,x_k,s} = 1$$

Bandwidth Constraint



- Link between x_1 and x_2 is loaded
  - by depth-2 connection (m_1-s_1)

$$\varepsilon^2_{x_1,x_2,m_1,s_1} = w(e_{m_1,s_1}) \times D^2_{m_1,x_1,x_2,s_1}$$

  - by depth-3 connection (m_2-s_2)

$$\varepsilon^3_{x_1,x_2,m_2,s_2} = w(e_{m_2,s_2}) \times \left(\sum_{x' \in X} D^3_{m_2,x_1,x_2,x',s_2} + \sum_{x' \in X} D^3_{m_2,x',x_1,x_2,s_2}\right)$$

Bandwidth Constraint

- Total weight bewteen 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

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

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} ⇒ $\sum_{m=0}^{|V_M|}\sum_{s=0}^{|V_S|} F_{m,s} \times K_{x,m,s}$ / $\sum_{m=0}^{|V_M|}\sum_{s=0}^{|V_S|} A_{m,s} \times K_{x,m,s}$
  - Cannot be directly obtained!!
- Obtain indexing matrix $K_{x,m,s}$
  - 1 *iff* crossbar x is mxs, otherwise 0

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

less than 1 by β

integer part — always less than 1 by α, thus fractional part

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

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

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

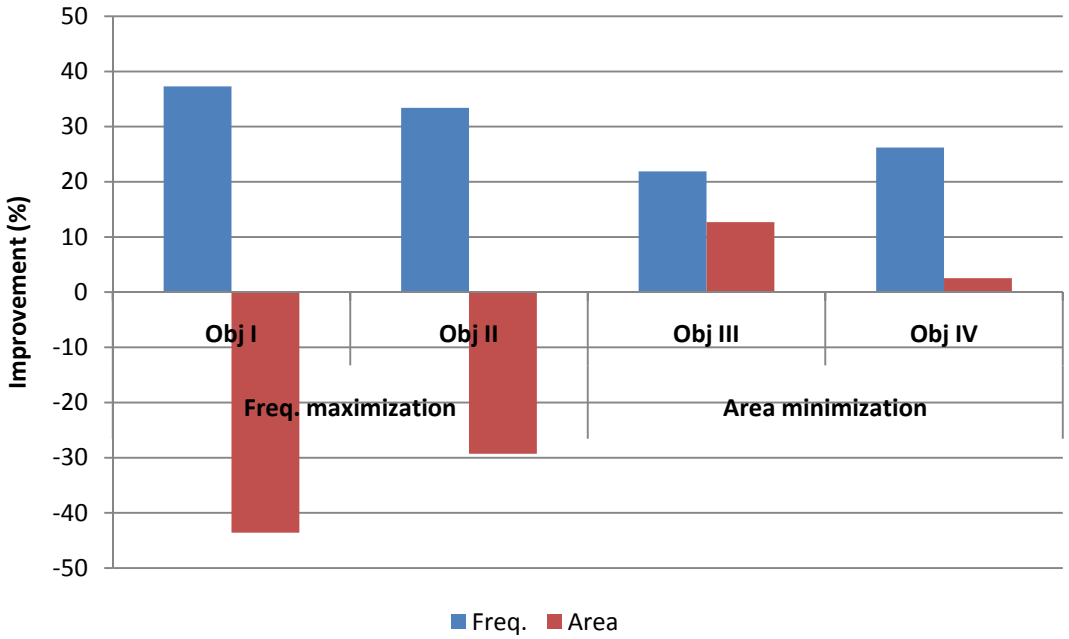
$K_{x,m,s}$ can and must be 1 if inside the brace is 0

Experiemt Setting

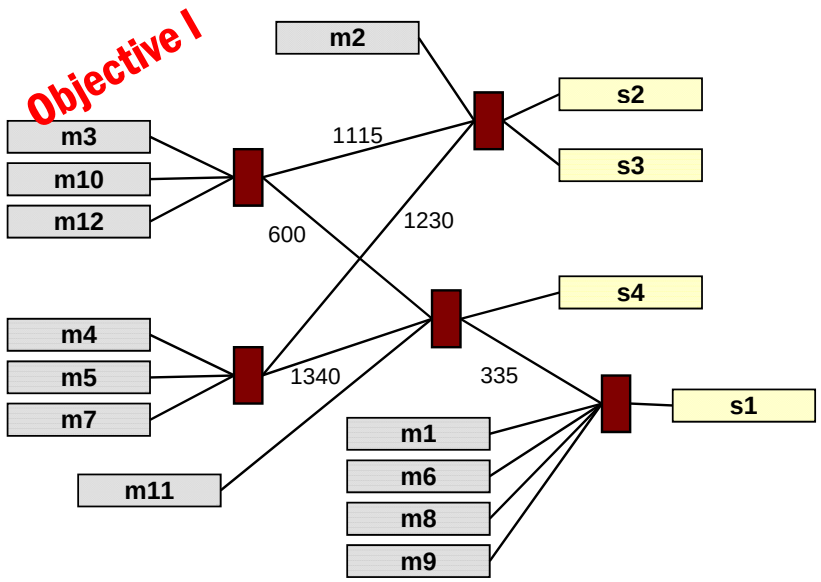
- Applications
  - App I : Industrial strength SoC (12x4)
  - App II : Mpeg4 decoder example (9x3) [9]
- Objectives
  - Frequency maximization w/o area upper bound
  - Frequency maximuzation w/ area upper bound (30%)
  - Area minimization w/o freq. lower bound
  - Area minimization w/ freq. lower bound (380Mhz)

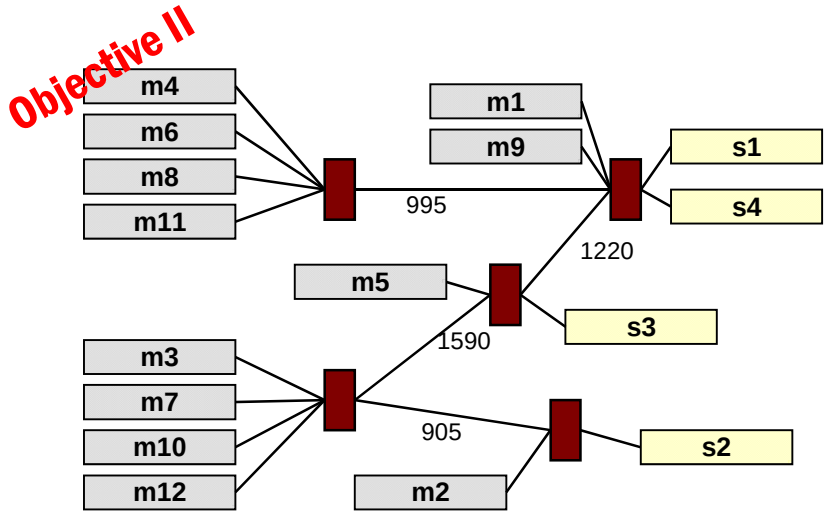
Result for App I

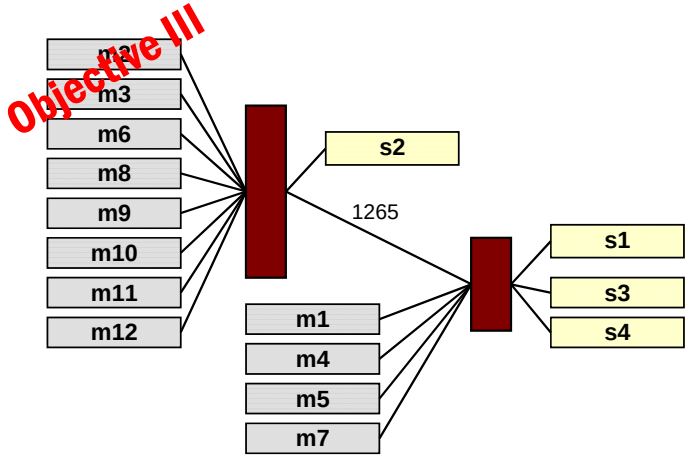
- Freq. improvement is up to 37.3%, but with 43.6% area overhead (Objective I)
- Area reduction is up to 12.7%, and with 21.9% freq. improvement (objective III)

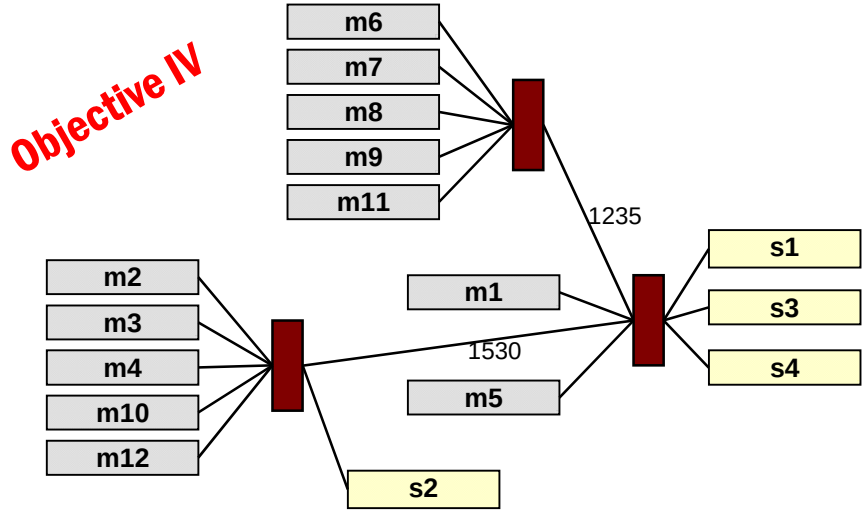


Result for App I



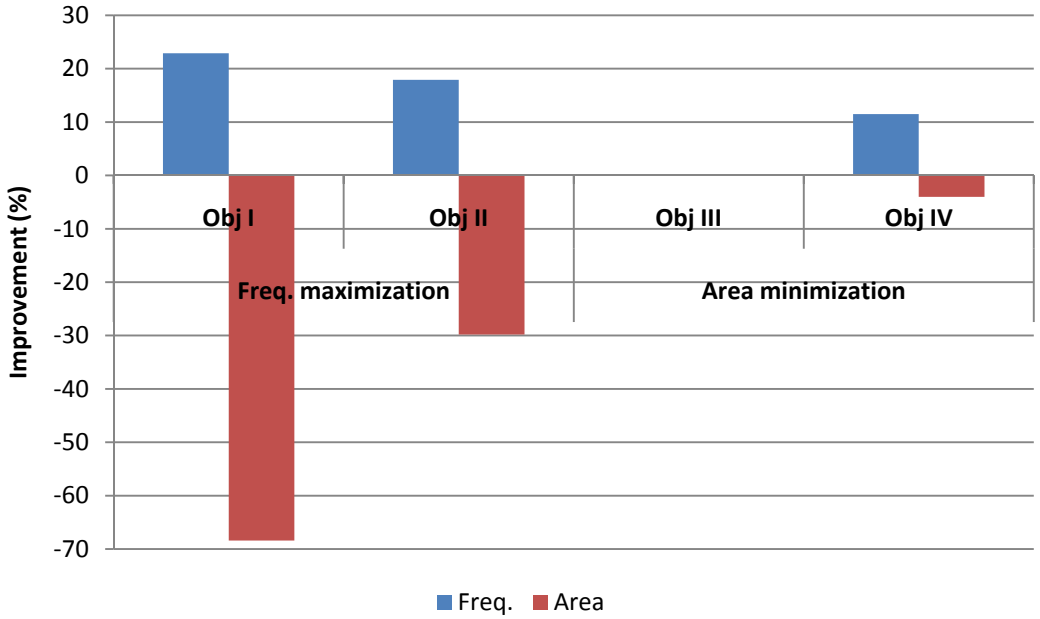




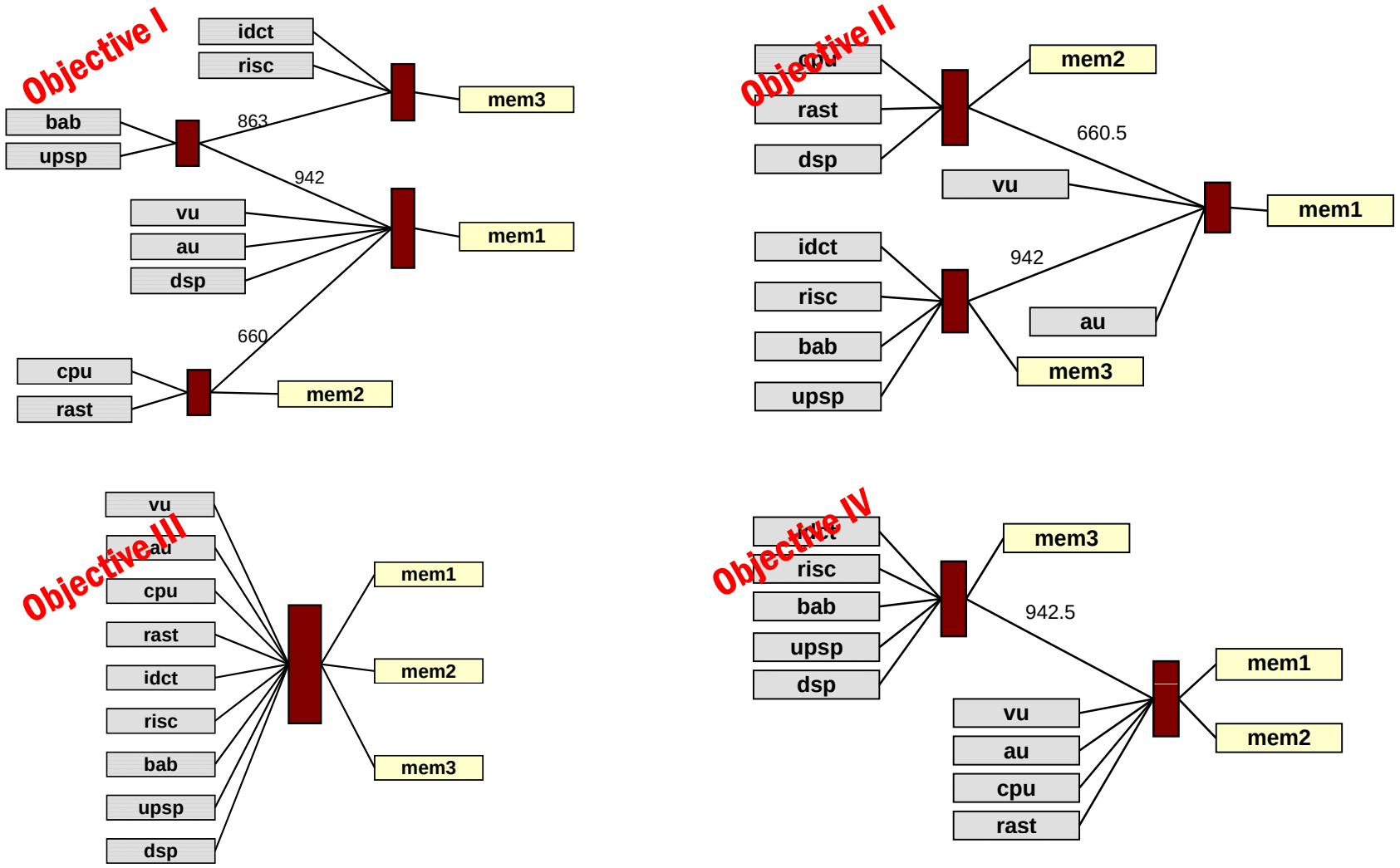


Result for App II

- 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 bbjective for power consumption

Thank You

References

[1] S. Murali and G. De Micheli, "An Application-Specific Design Methodology for STbus Crossbar Generation", *DATE 2005*, pp. 1176-1181

[2] S. Pasricha, N. Dutt, M. Ben-Romdhane, "Constraint-Driven Bus Matrix Synthesis for MPSoC", *ASPDAC 2006*, pp. 30-35

[3] S. Pasricha, N. Dutt, “COSMECA: application specific co-synthesis of memory and communicat ion architectures for MPSoC”, DATE 2006, pp. 700-705

[4] S. Murali, L. Benini, and G. De Micheli, "An Application- Specific Design Methodology for On-C hip Crossbar Generation", Trans. VLSI (to appear), available at http://infoscience.epfl.ch/

[5] K. Srinivasan, K. S. Chatha, and G. Konjevod, "Linear Programming Based Technique for Synth esis of Network-on- Chip Architectures", *IEEE TVLSI,* April 2006*,* vol. 14, pp. 407-420

[6] J. Yoo, S. Yoo, and K. Choi, "Communication Architecture Synthesis of Cascaded Bus Matrix", *ASPDAC 2007*, pp. 171-177

[7] ARM, www.arm.com

[8] Sonics Inc., www.sonicsinc.com

[9] K. Srinivasan, K. S. Chatha, “A low complexity heuristic for design of custom network-on-chip a rchitectures”, DATE 2006, pp.130-135