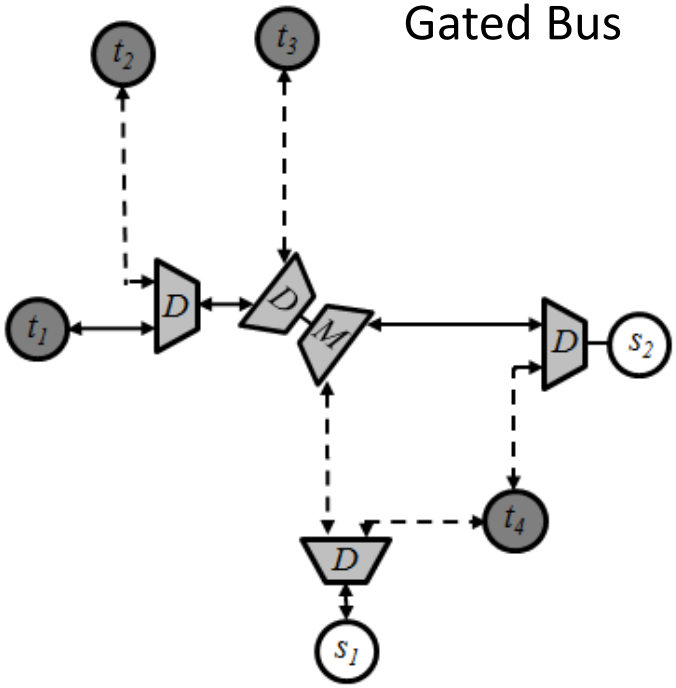
Low-Power Gated Bus Synthesis for 3D IC via Rectilinear Shortest-Path Steiner Graph

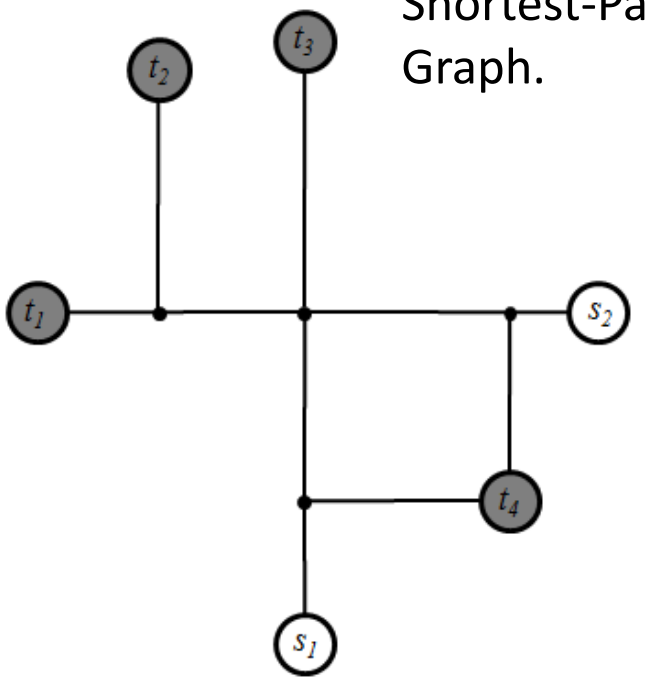
Chung-Kuan Cheng, Peng Du,
Andrew B. Kahng, and Shih-Hung Weng
UC San Diego
Email: ckcheng@ucsd.edu

Outline

- Introduction
- Statement of Problem
- Algorithms
  - Determination of TSV locations
  - Generating Rectilinear Shortest-Path Steiner Graph
- Experimental Results
- Conclusion

Introduction: 2D bus

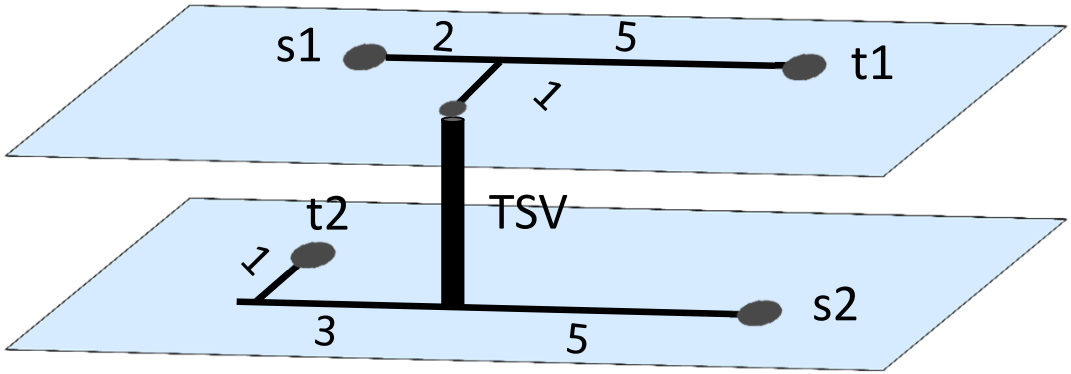




- Problem: a gated bus with multiplexers and demultiplexers to minimize power consumption
- Shorest-Path Steiner Graph: a graph that contains shortest paths between sources and sinks, with minimal total wire length

Introduction: 3D Bus

- Through Silicon Vias (TSV) for inter-silicon connection
  - Silicon area
  - Feature size
  - Yield
- Implication:
  - The z segment is more expensive than x & y segments
  - Routing distance between different layers may not be the shortest



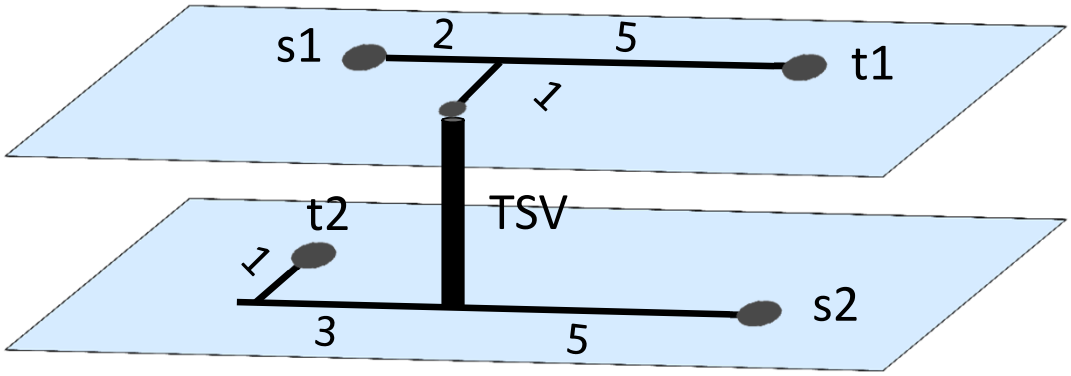
Statement of Problem

- Given: A set of masters (*src*) and a set of slaves (*dst*) on *L* silicon layers, and traffic demands between all (*src*, *dst*) pairs
- Assumption: time sharing bus, one channel on each direction. Routing is optimized and fixed.
- Objective: (1) Power consumed by the traffic and (2) total wire length
- Output: 3D Steiner graph
- Constraint: bounded #TSVs one each silicon layer

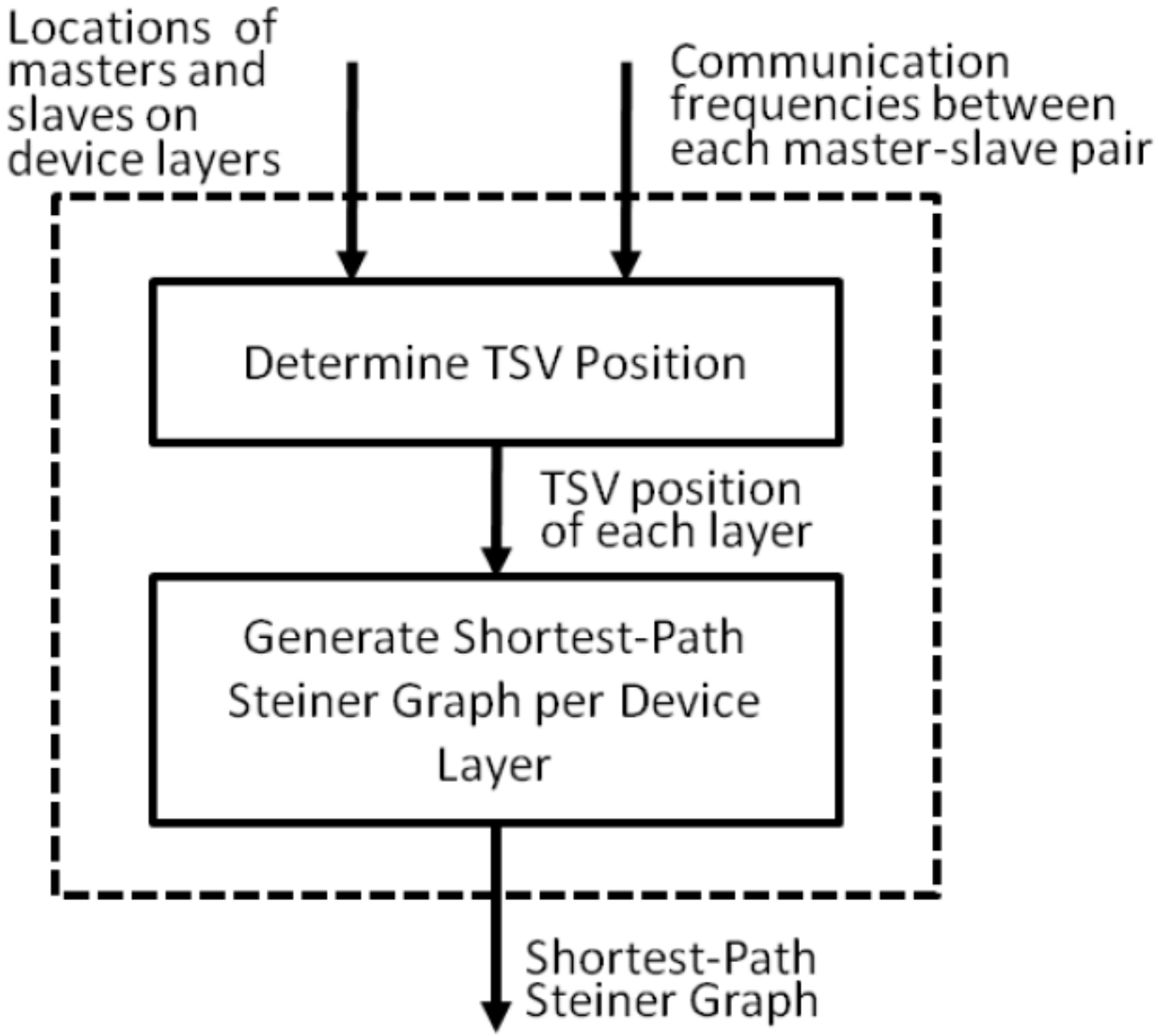
Motivational Example

- src: s1, s2, dst: t1, t2
- Traffic Demands:
  - (s1, t1) = 5, (s1, t2) = 1
  - (s2, t1) = 3, (s2, t2) = 4
- #TSV/layer= 1
- Wire length
  - (2+5+1)+(1+3+5)
- Power consumption
  - 5x7+1x7+3x11+4x9

One channel for each direction
Power = demand x length



Overall Flow



Problem Formulation

- **TSV Placement**: Place TSVs between adjacent layers so that the total traffic power (length of weighted shortest paths between *src-dst* pairs) is minimized.
- **Steiner Graph on Each Layer**: Given a silicon layer *k* with TSV locations on both sides, construct a shortest-path Steiner graph to connect all traffics between *srcs*, *dsts*, and TSVs on layer *k*.

TSV Placement (#TSV/layer=1)

- For #TSV=1, we can decompose 2D placement into 1D.
- A dynamic programming algorithm is proposed to find optimal TSV locations.
  - Let $Opt(k,r)$ be the minimal total traffic power among terminals (*src, dst*) in the first k layers and the TSV between layers k and $k+1$ at location r.
- Algorithm complexity is $O((n+m)^2 L)$, where n=#*srcs*, m=#*dsts*, L=#*layers*.

TSV Placement (#TSV/layer>1)

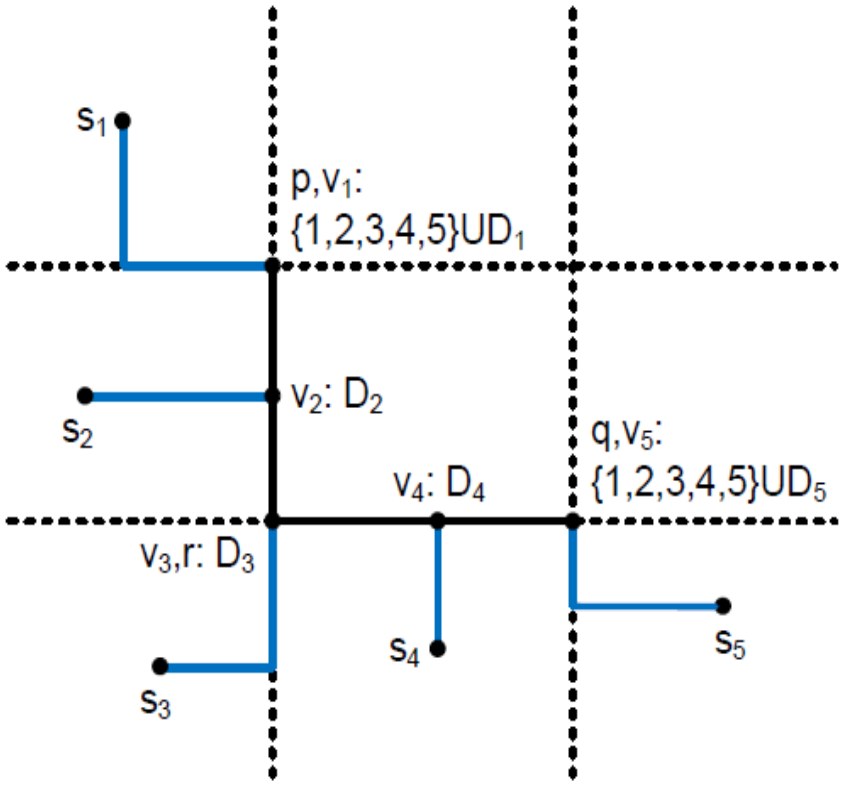
1. Snap the Hanan points into a coarse grid, e.g. 5x5
2. Find the best TSV placement on the snapped Hanan points using exhaustive search
3. For every TSV, refine the placement.
4. Repeat step 3 until there is no improvement.

Steiner Graph on Each Layer (tree merge)

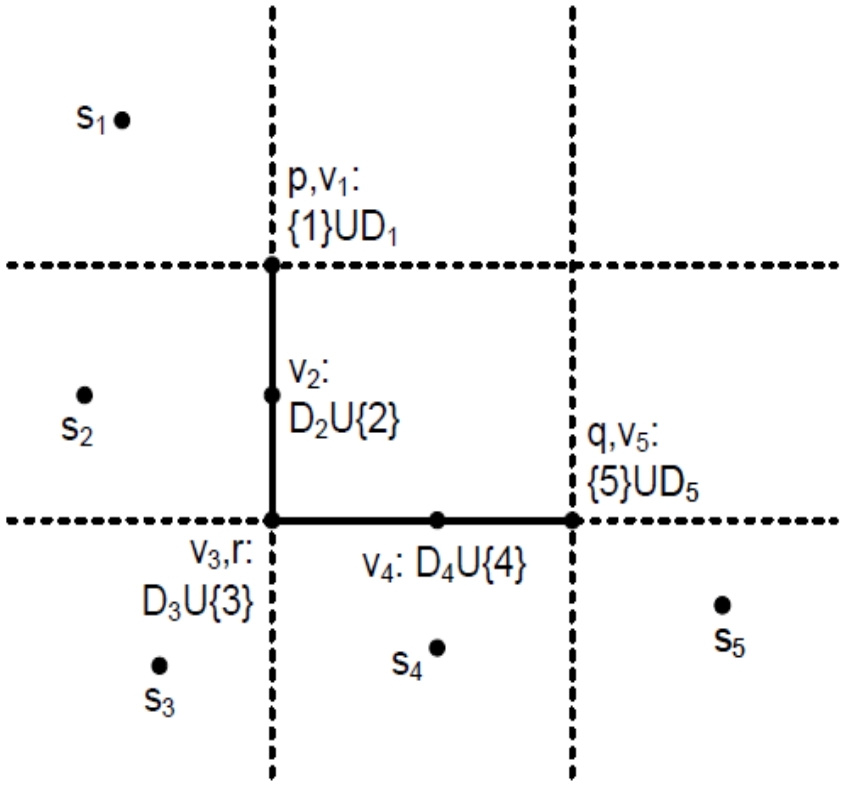
1. Start with m *dst*s as m trees. Each root of the tree contains an *src* list to be connected.
2. Merge a pair of roots p and q with the largest benefit. Update the *src* list on the new root.
3. Repeat step 2 until there is no more pairs to be merged.
4. For the roots of nonempty *src* list, route to the *src*s on the list.
5. Remove redundant edges.

Computational Complexity $O(nm^2)$

Steiner Graph on Each Layer (tree merge)



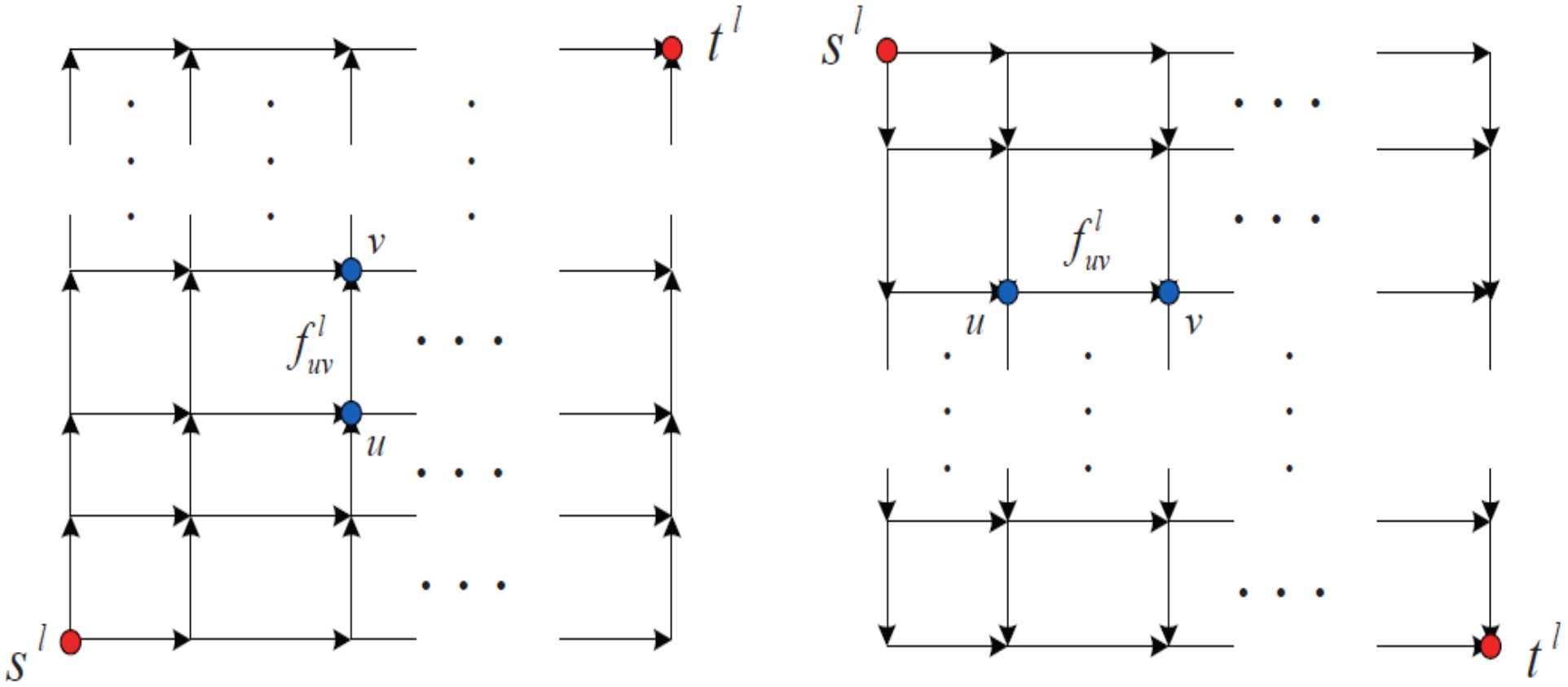
Original demand set.



Updated demand set.

- Our objective is to connect each one of s_1,s_2,s_3,s_4,s_5 to p and q.
- By merging p and q, the benefit is the total length of blue segments.

Steiner Graph on Each Layer (LP Rounding)



S^l is below t^l

S^l is above t^l

- The figure depicts the directed network N_l on the Hanan grid.
- The rectilinear shortest path from s^l to t^l corresponds to a flow with amount one in N_l.

Steiner Graph on Each Layer (LP Rounding)

$$\min \quad \sum_{(u,v)\in E_H} d_{uv}x_{uv}$$

$$s.t. \quad \sum_{(u,v)\in E_l} f_{uv}^l - \sum_{(v,u)\in E_l} f_{vu}^l \geq 0,$$

$$1 \leq l \leq Q \text{ and } v \neq s^l, t^l;$$

$$\sum_{(s^l,u)\in E_l} f_{s^l u}^l \geq 1, \quad 1 \leq l \leq Q;$$

$$\sum_{(u,t^l)\in E_l} f_{ut^l}^l \geq 1, \quad 1 \leq l \leq Q;$$

$$x_{uv} - f_{uv}^l \geq 0,$$

$$(u,v) \in E_H \text{ and } l \in \{l_0 : (u,v) \in E_{l_0}\};$$

$$f_{uv}^l \in \{0,1\}, \quad 1 \leq l \leq Q \text{ and } (u,v) \in E_l;$$

$$x_{uv} \in \{0,1\}, \quad (u,v) \in E_H.$$

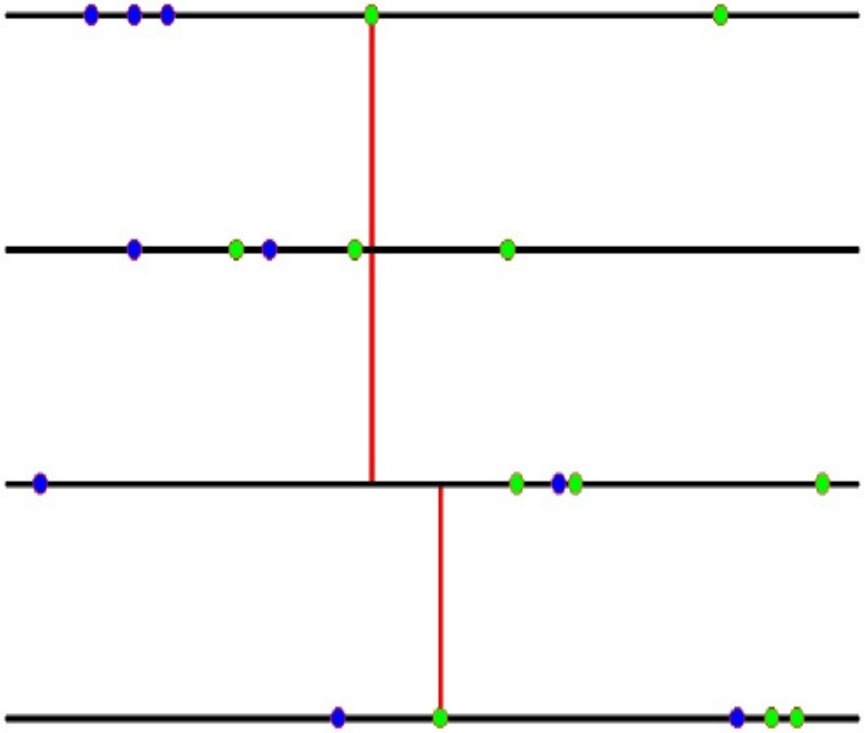
- E_h: undirected edge set of Hanan grid.
- E_l: directed edge set on top of E_h for each demand l
- $f^l_{u,v}$: flow from u to v on edge (u,v) in E_l.
- Q: # demands (src, dst)
- x: a binary variable to denote the selection of edge (u,v) in the graph.
- d: wire length of edge (u,v).

Steiner Graph on Each Layer (LP Rounding)

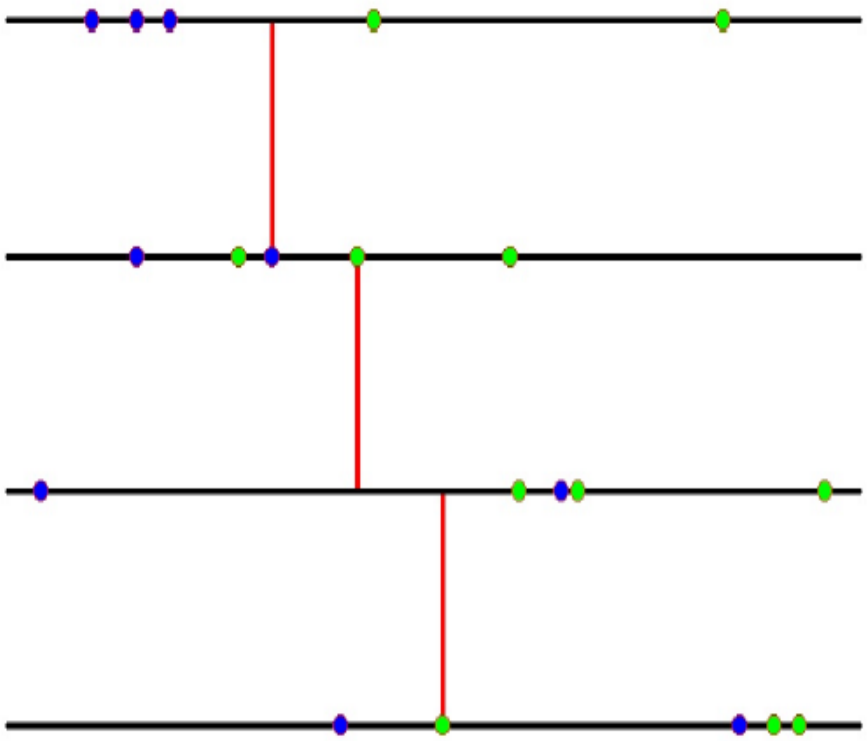
- Solve the LP relaxation of the ILP formulation.
- Sort the edges with respect to the decreasing order of the x variables.
- Delete edges as long as the remaining graph contains necessary shortest paths.

#variables: $O((n+m)^2Q)$

Experimental Results (#TSV/layer=1)

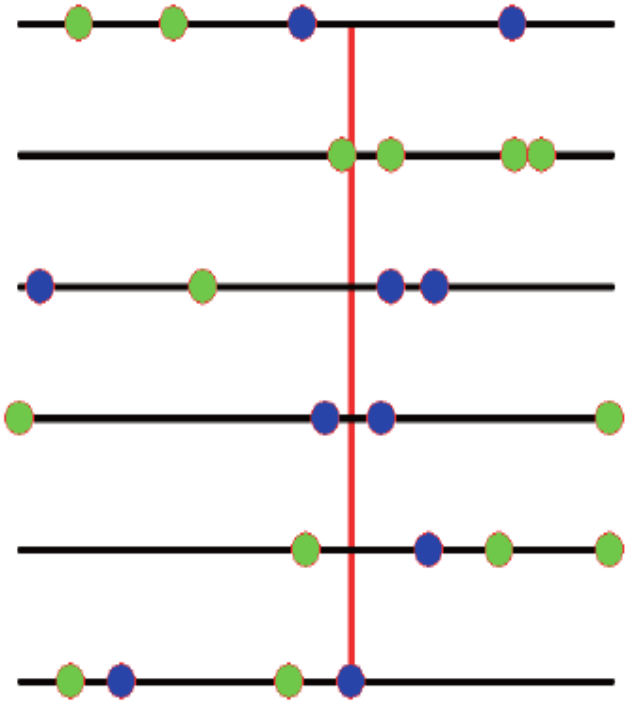


The same communication frequencies for all master-slave pairs.

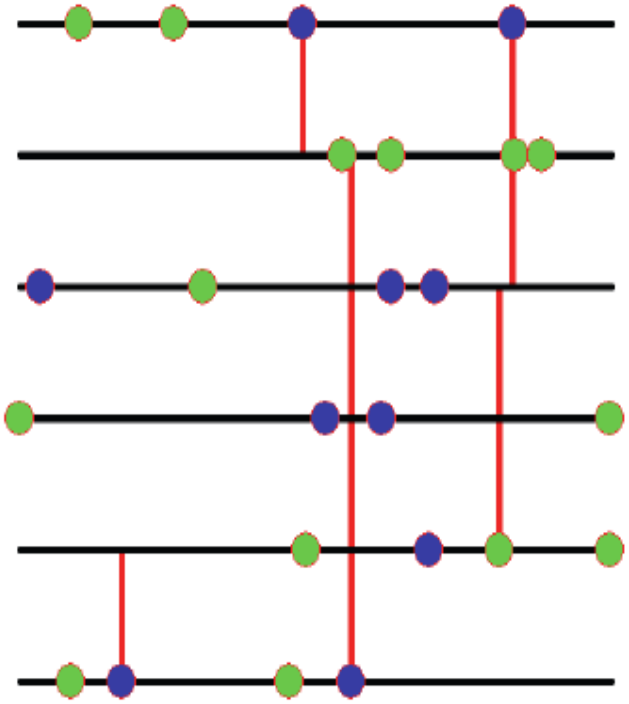


(src, dst) pairs in first two layers communicate 5 times freq.

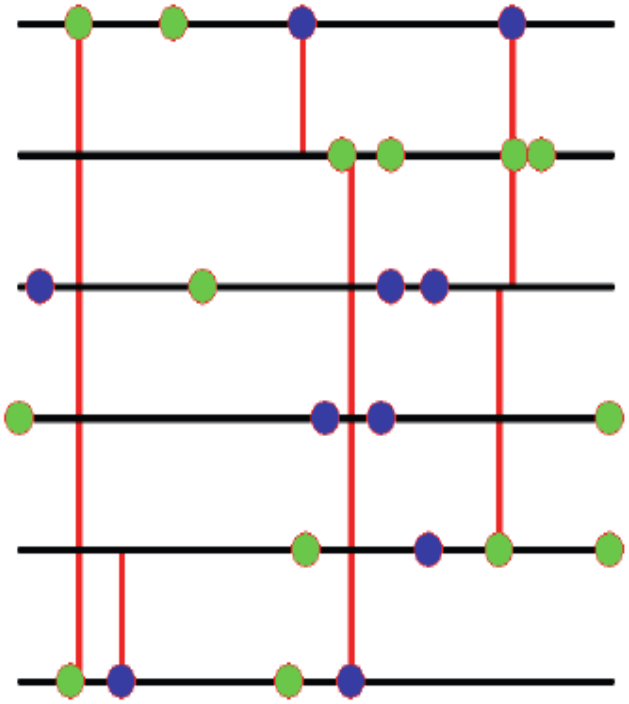
Experimental Results



#TSV/layer=1
Power=439



#TSVs/layer=2
Power=395



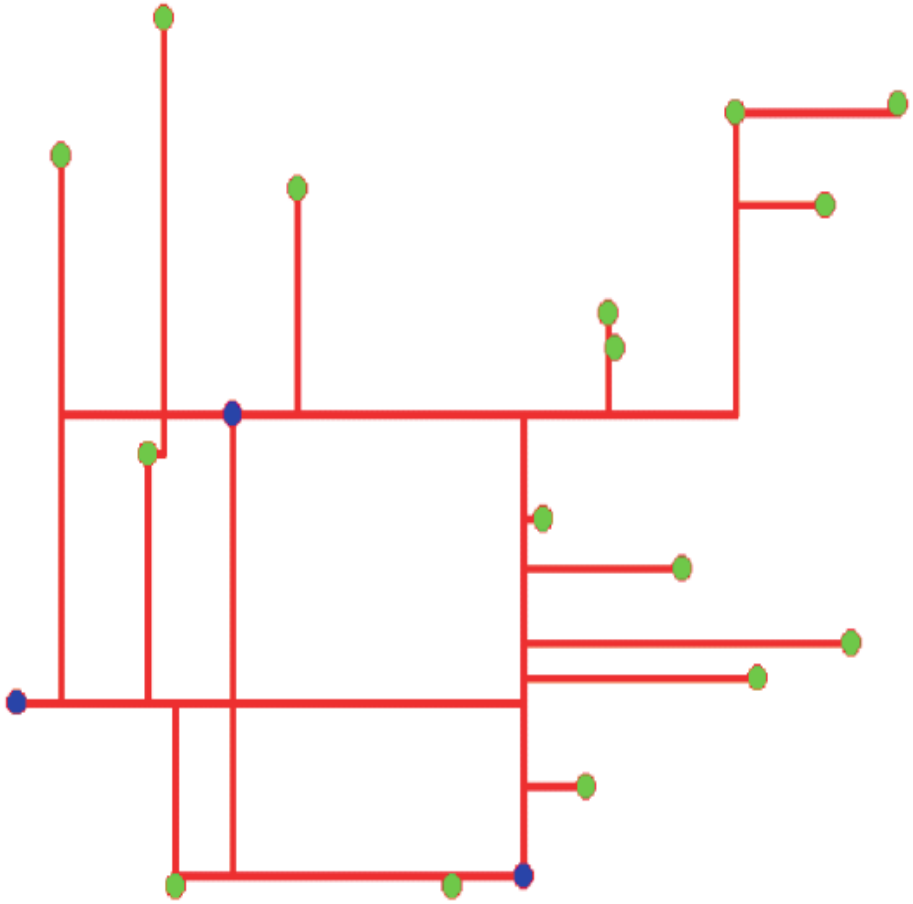
#TSVs/layer=3
Power=348

Experimental Results: Power

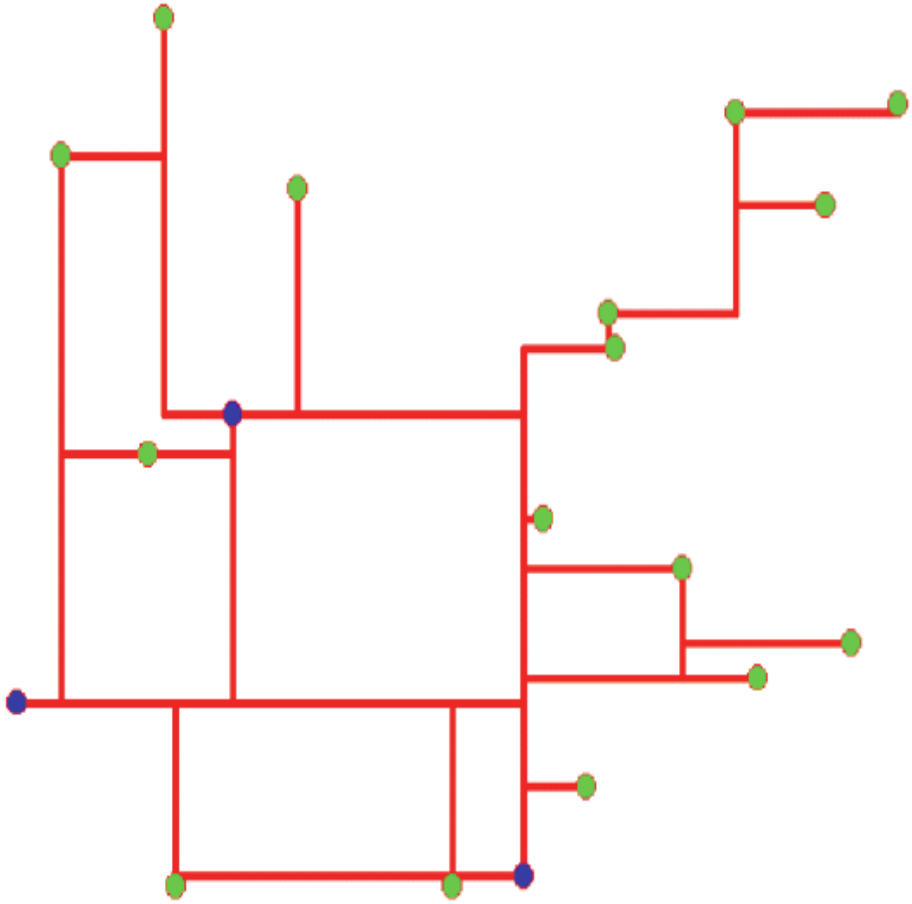
(L, N)	$B = 1$	$B = 2$	$B = 3$	$B = \infty$
(3,10)	456.46 (31.79%)	375.93 (8.54%)	360.86 (4.19%)	346.36
(3,20)	384.34 (31.45%)	336.13 (14.96%)	305.50 (4.49%)	292.39
(3,50)	453.09 (34.34%)	393.07 (16.55%)	355.22 (5.32%)	337.27
(4,20)	476.15 (38.45%)	413.28 (20.16%)	367.62 (6.89%)	343.93
(5,20)	492.00 (41.92%)	431.04 (24.33%)	367.39 (5.97%)	346.68
(5,50)	452.50 (39.77%)	401.23 (23.93%)	346.31 (6.96%)	323.76

- *(L,N)*: (# layers, # masters and slaves in each layer)
- B: #TSVs/layer

Experimental Results (Steiner Graph)



Length=6006, 5.38% extra
Tree merge



Length=5683, 0% extra
LP relaxation and rounding

Experimental Results (Steiner Graph)

(n, m)	Previous	Greedy	LP(Obj)	LP(Round)	Improvement
$(3, 16)$	5388	5068	4882	4882	9.39%
$(5, 15)$	7024	6586	6342	6342	9.71%
$(12, 6)$	6698	6306	5915	5915	11.69%
$(6, 12)$	7127	6160	5575	5575	21.78%
$(12, 12)$	11559	10385	10319	10319	10.73%
$(20, 20)$	16236	15921	13847	13847	14.71%
$(30, 30)$	302619	287042	251841	251968	16.78%

Lengths of LP(Obj) and LP(Round) are almost the same with 1.0005 ratio on the last case

- Previous: [Wang DAC09]
- Greedy: Tree merge
- Improvement: Previous vs VP(Round)

CPU Time of LP Relaxation and Rounding

(n, m)	# Variables	# Constraints	Time	Memory
$(3, 16)$	4326	6209	<1s	<50MB
$(5, 15)$	8019	11458	<1s	<50MB
$(20, 20)$	155396	239037	3m43s	234MB
$(30, 30)$	778358	1175974	4h22m	1.2GB

CPU: Intel Core i3, 2.4GHz; Memory: 4GB

Conclusion

- A framework and algorithms to synthesize the gated bus in 3D ICs.
- Optimal TSV placement when #TSV/layer=1
  Exhaustive search on coarse grid + iterative improvement when #TSV/layer>1
- New Steiner graph algorithms with total wire length reduction of up to 22%.
- Future Works
  - Multiple Path Graph
  - Control Systems

Thank you

for

your attention!