Min-Cut Partitioning with Functional Replication for Technology Mapped Circuits using Minimum Area Overhead

Wai-Kei Mak
Dept. of Computer Science and Engineering
University of South Florida
Tampa, FL 33620
wkmak@csee.usf.edu

Motivation for Our Work

Cut Size Reduction

The number of cut nets in a partitioned circuit can often be reduced by replicating some logic cells in two or more components.

Area Consideration

Due to area constraint on a component, excessive amount of replication should be avoided.

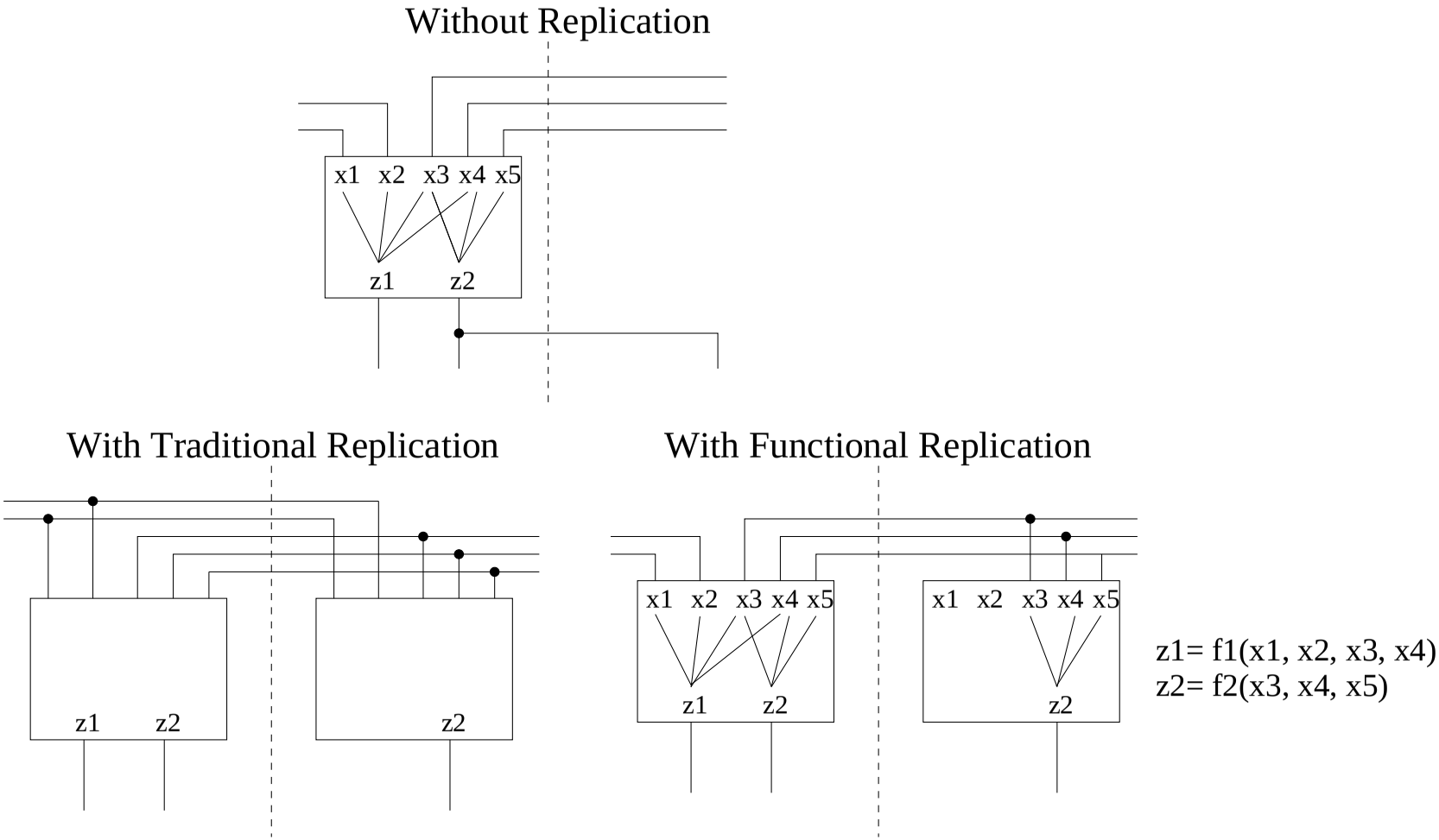
A Desirable Solution

Optimize the cut size with the least possible area overhead resulted from replication.

Functional Replication vs. Traditional Replication

- Logic cells can have **multiple outputs** (e.g. those in FPGA).
- Each cell output can depend on a different subset of cell inputs.
- Traditional Replication
  - Preserve all input signals for both cell copies when replicating a cell.
- Functional Replication
  - Preserve for each cell copy only the input signals for its required outputs when replicating a cell.
- **Advantages of functional replication**
  - More flexible
  - Bigger reduction in cut size

Functional Replication vs. Traditional Replication Example



Functional replication considers the **dependency** of different cell outputs on the cell inputs.

Comparison with Previous Work on Functional Replication

Previous Work

- The only previous work: Kuznar, Brglez, Zajc in DAC'94
- Use a **Fiduccia-Mattheyses type** heuristic
- Shortcomings
  - May functionally replicate some cells unnessarily
  - Final cut size is not guaranteed to be optimal

Our Work

- Based on **max-flow min-cut** computation
- Advantages
  - A cell is functionally replicated only if it is necessary for attaining the minimum cut size
  - Final cut size is always optimal

Related Work

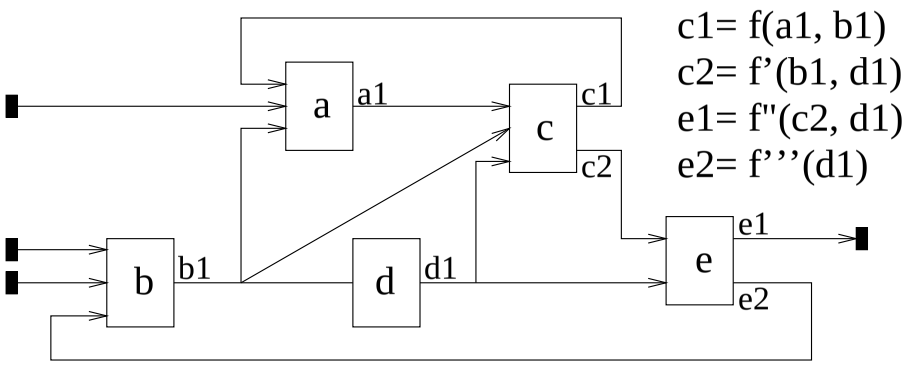
- **Liu, Kuo, Cheng, Hu in TCAD'95** ("A replication cut for two-way partitioning")
  - Use **traditional** replication
  - Cut size $|C_{trad\_rep}|$ found is optimal under traditional replication
  - No control on amount of replication
- **Mak, Wong in ICCAD'96** ("Minimum replication min-cut partitioning")
  - Use **traditional** replication
  - Cut size $|C_{trad\_rep}|$ found is optimal under traditional replication
  - Optimize amount of replication
- **Current Work**
  - Use **functional** replication
  - Cut size $|C_{func\_rep}|$ found is optimal under functional replication, moreover, $|C_{func\_rep}| \leq |C_{trad\_rep}|$
  - Optimize replication area overhead

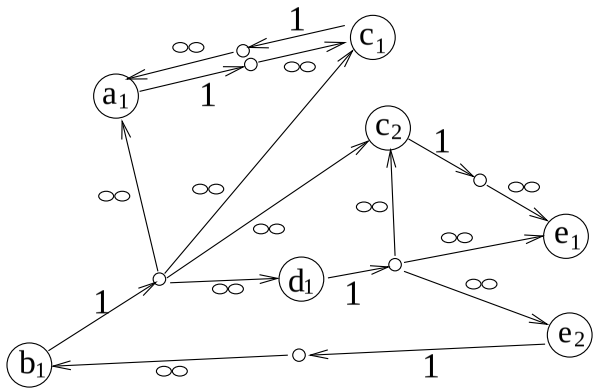
Method

A **2-phase network-flow** approach.

Phase 1

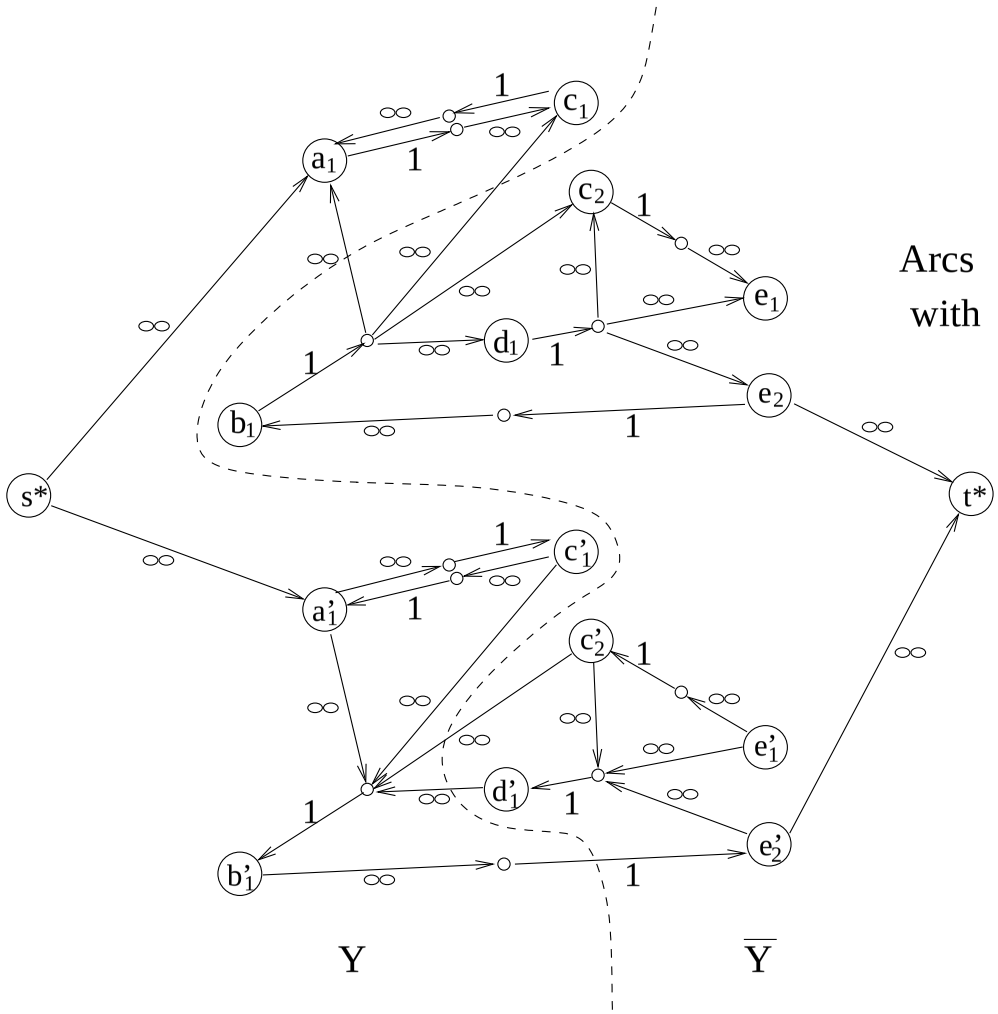
- The circuit is first represented using a **function graph** G.





Each node represents a function and the edges show the dependency of the functions.

- A **flow network** N is then constructed based on the function graph.



Arcs $(a_1, a'_1), (b_1, b'_1), (c_1, c'_1), (c_2, c'_2), (d_1, d'_1), (e_1, e'_1), (e_2, e'_2)$
with ∞ capacity are not shown

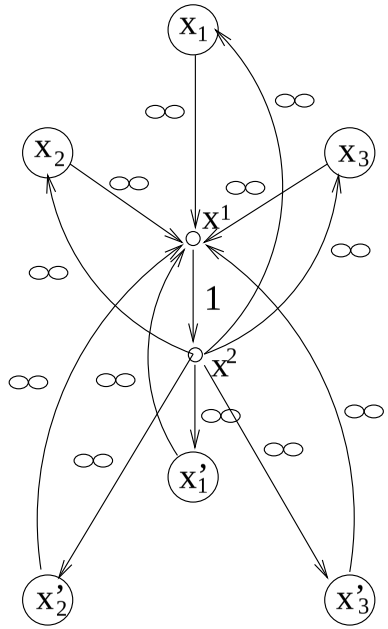
Theorem: Any minimum cut of the flow network N induces a min-cut replication partition of the circuit.

However, to **minimize the area overhead** to attain the minimum cut size, a Phase 2 is required.

Fact: A cell x is replicated in the induced partition iff the cut divides $\{x_i, x_i' : i = 1, 2, \ldots, p\}$ in N.

Phase 2

- The network N is modified into a network N_{Area} so that the min-cut partition requiring the smallest area overhead will stand out from other min-cut replication partitions.
  - Network N_{Area} contains a **penalty arc set** for each cell x so that a penalty equals to the area of cell x is incurred if a cut divides $\{x_i, x_i' : i = 1, 2, \ldots, p\}$ (i.e., if cell x is replicated).



Theorem: A minimum cut of network N_{Area} induces a min-cut replication partition of the circuit that uses the smallest area overhead.

Area-Constrained Partitioning

- Our algorithm can be used to **improve** the solution produced by any area-constrained functional replication partitioning heuristic.
  - Suppose a heuristic partitioned the functions of a circuit into three sets: S^h, T^h, R^h with R^h being the set of replicated functions.
  - By collapsing all nodes in S^h to the source s^* and all nodes in T^h to the sink t^* in network N, we can compute a new partition satisfying $S^h \subset S$ and $T^h \subset T$ using our algorithm.
  - The optimality of our algorithm guarantees that
    i. The new **cut size** will be **smaller**
    ii. The **area overhead** incurred will be **minimized**

Experimental Results

- A simulated-annealing based heuristic was first used to compute a good partition with replication.
- Our algorithm was applied to further optimize the partition.
- Large reduction in area overhead was obtained.

Experimental Results (Cont'd)

Circuit	S.A.		Optimized		Overhead reduction
	cut	overhead	cut	overhead	
c3540	14	9.5%	12	8.5%	10.5%
c5315	6	9.8%	6	6.9%	29.6%
c6288	13	9.5%	12	4.1%	56.8%
c7552	3	9.8%	3	8.4%	14.3%
s5378	19	9.7%	19	7.9%	18.6%
s9234	35	9.9%	35	7.7%	22.2%
s15850	40	10.0%	40	7.7%	23.0%
s38417	98	10.0%	69	6.3%	37.0%
s38584	101	10.0%	80	4.5%	55.0%