

On the Interaction between Power-Aware CAD Algorithms for FPGAs

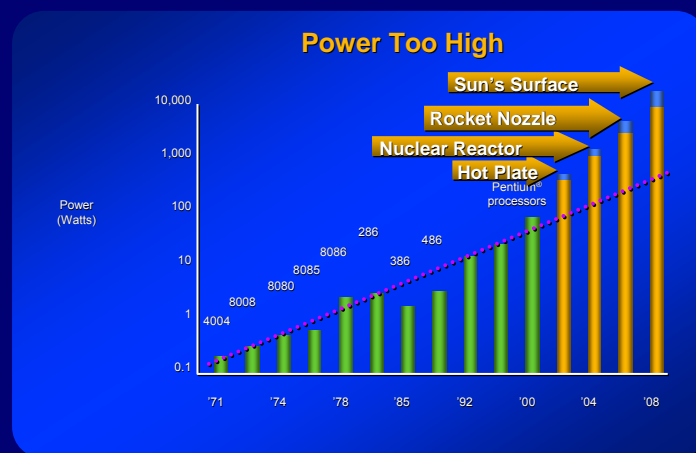
by

Julien Lamoureux

Steven J.E Wilton

University of British Columbia

Why Reduce Power?



Source: Intel

2

Why Reduce Power?

- Longer Battery Life
- Cheaper Packaging
- Improve Reliability



3

Reducing FPGA Power

Two ways of reducing FPGA power consumption

- Optimize the FPGA
 - use shorter wires segments
 - bigger clusters
- Optimize the FPGA CAD Flow
 - hide high-activity wires within LUTs or clusters
 - reduce the routing length of high-activity wires

4

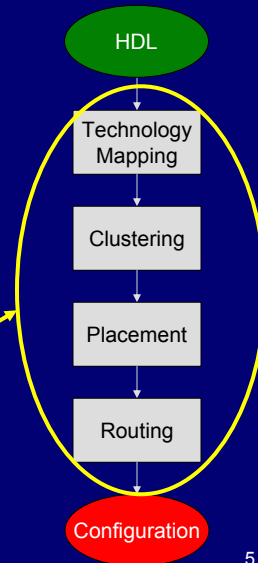
Goals of this Research:

Answer two basic questions:

1. Where do we expect the biggest power savings?
2. Are these savings cumulative?

Use this information to design better FPGAs, CAD Tools

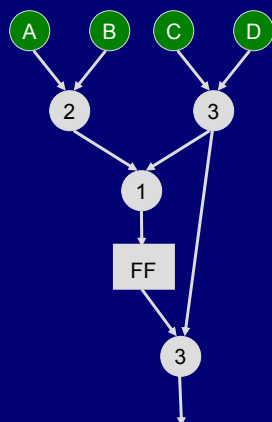
This Talk



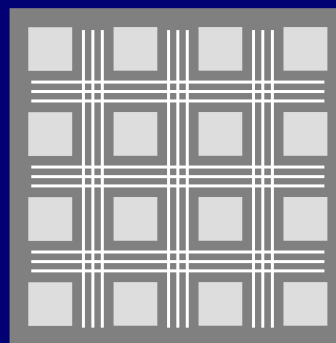
5

What's an FPGA CAD Flow?

User Circuit

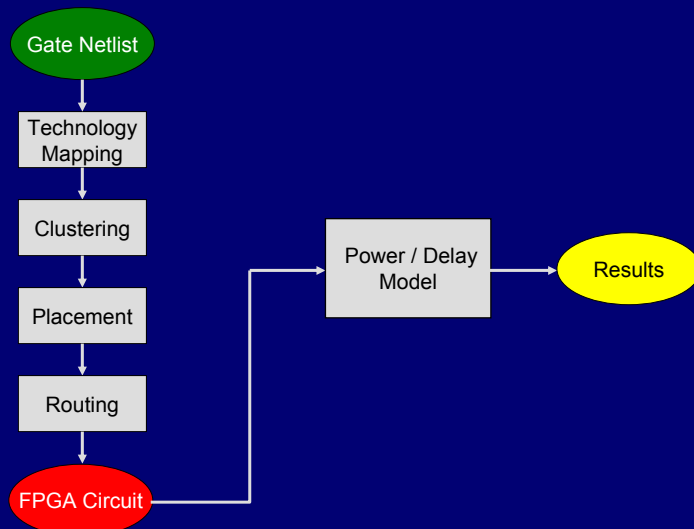


FPGA



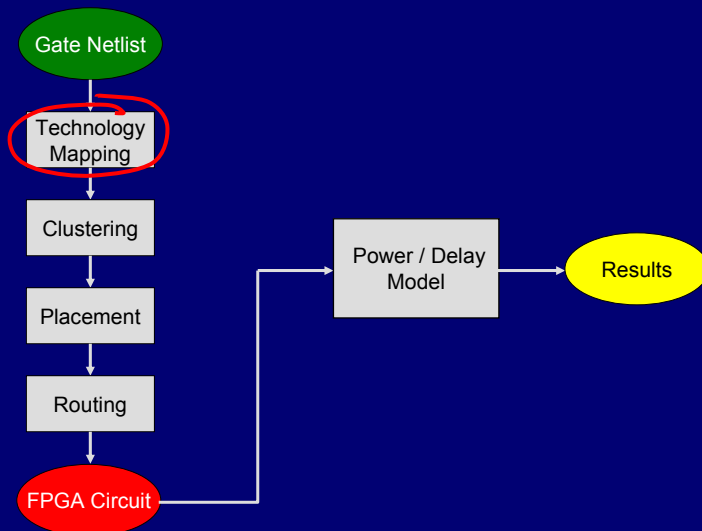
6

Experimental Framework



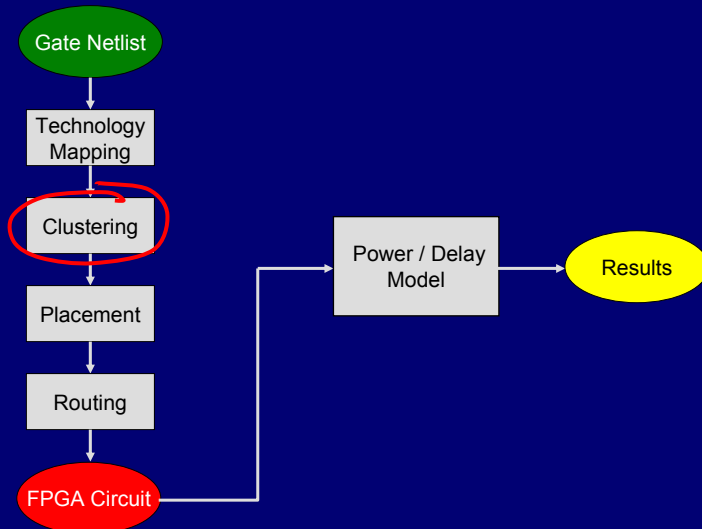
7

Experimental Framework



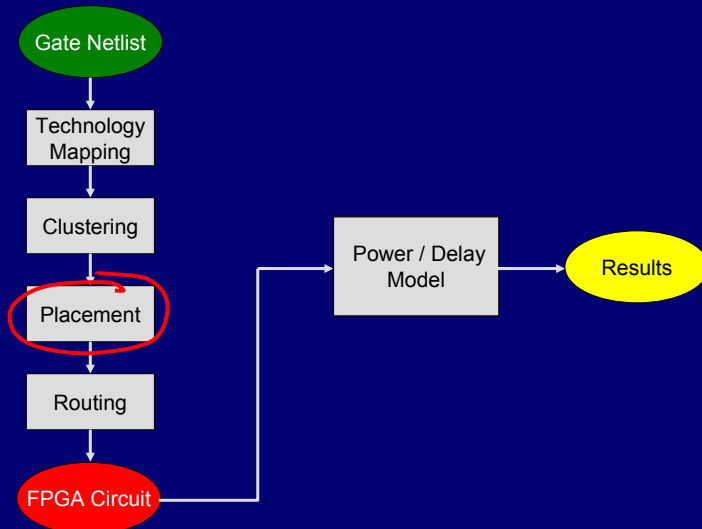
8

Experimental Framework



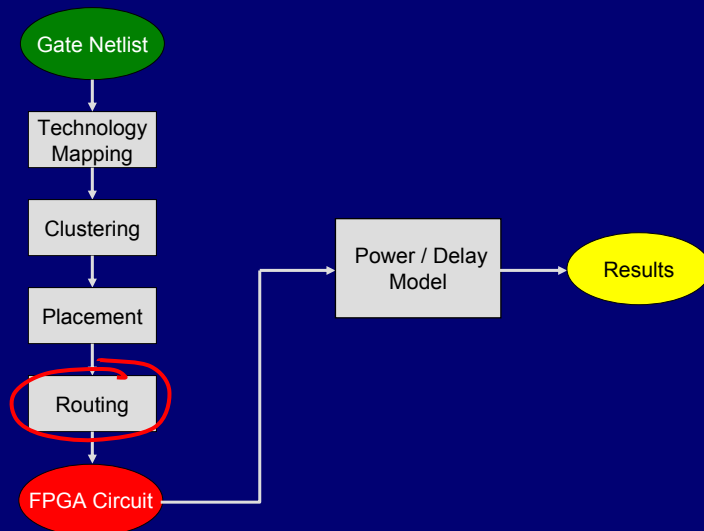
9

Experimental Framework



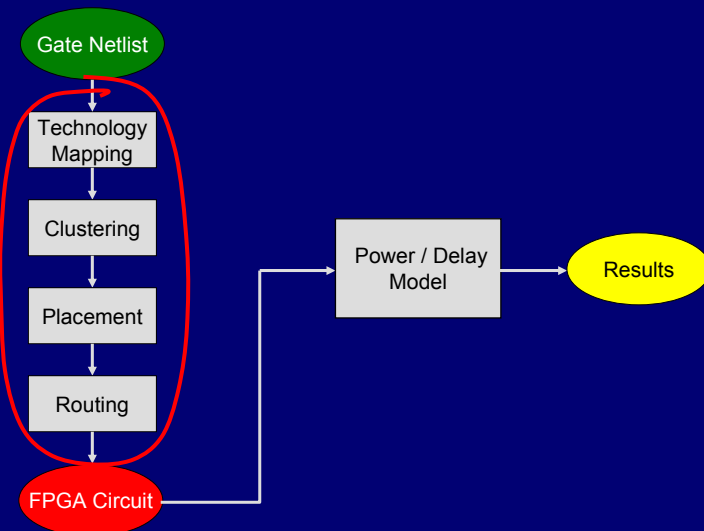
10

Experimental Framework



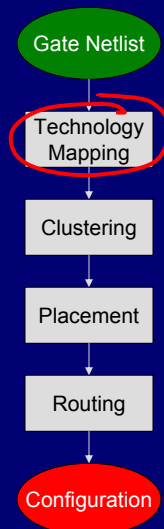
11

Experimental Framework



12

Technology Mapping



13

Technology Mapping

Operation: *Transforms a netlist of gates into a netlist of lookup tables (Cutmap)*

Goal: *Minimize energy (power*delay)*

Delay → optimize circuit depth

Power → 2 ways:

- hide high-activity nets in LUTs
- minimize unnecessary node duplication

14

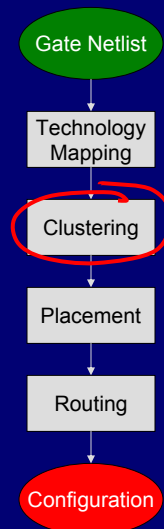
Technology Mapping Results

	LUTs	Connections	Activity	Energy (nJ)
CutMap	2576	10746	0.330	2.18
EMap	2441	9705	0.323	2.01
% Diff	-5.2	-9.7	-2.1	-7.6

For EMap, most of the savings come from minimizing unnecessary node duplication.

15

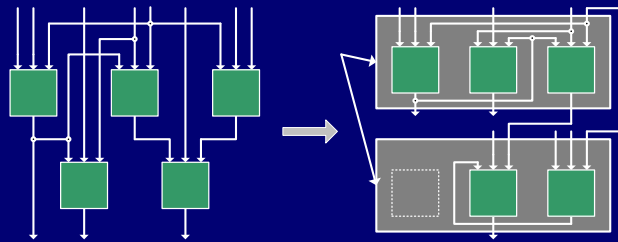
Clustering



16

Clustering

Operation: *Transforms a netlist of LUTs into a netlist of clusters (T-VPack)*



17

Clustering Cost Functions

Seed Attraction Function

T-VPack: *most timing-critical LUT*

P-T-VPack: *LUT with highest activity pins*

Second Attraction Function

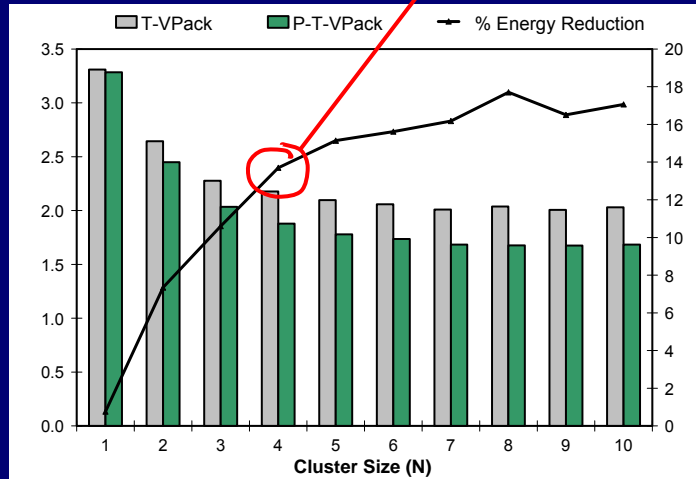
T-VPack: $\lambda * Criticality(B) + (1 - \lambda) * \frac{|SharedNet(B,C)|}{K}$

P-T-VPack: $Criticality(B) + \alpha \frac{\sum Weight(i) | i \in SharedNets(B,C)}{K} + \beta \frac{\sum Act(i) | i \in SharedNets(B,C)}{K * Act_{avg}}$

18

Clustering Results

12.6% energy reduction

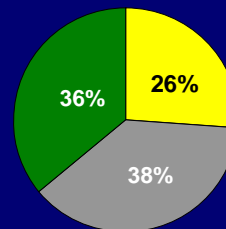


19

Clustering Results

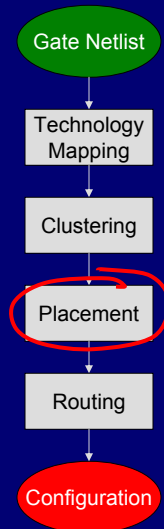
- Clustering savings are greater than tech. map savings since clusters are bigger than LUTs
- Savings are obtained by hiding high activity wires within clusters

- Wires Inside LUTs
- Wires Inside Clusters
- Global Wires



20

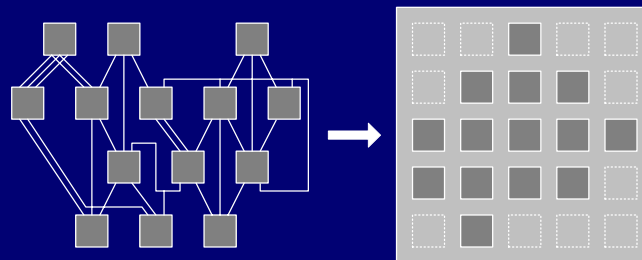
Placement



21

Placement

Operation: *Places clusters to physical locations on the FPGA (T-VPlace)*



22

Placer Cost Functions

Original Algorithm:

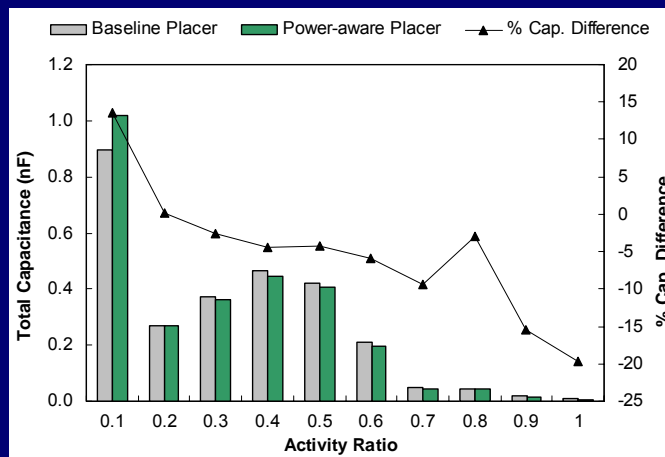
$$\Delta C_{\text{old}} = \lambda \frac{\Delta \text{timing cost}}{\text{previous timing cost}} + (1 - \lambda) \frac{\Delta \text{wiring cost}}{\text{previous wiring cost}}$$

Power-Aware Algorithm:

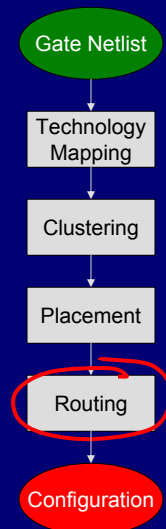
$$\Delta C_{\text{old}} = \lambda \frac{\Delta \text{timing cost}}{\text{previous timing cost}} + (1 - \lambda) \frac{\Delta \text{wiring cost}}{\text{previous wiring cost}} + \beta \frac{\Delta \text{power cost}}{\text{previous power cost}}$$

Placement Results

3.0% energy reduction



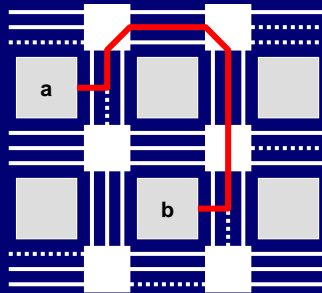
Routing



25

Routing

Operation: *Connect logic blocks using routing fabric (VPR Router)*



26

Router Cost Functions

Original Cost Function:

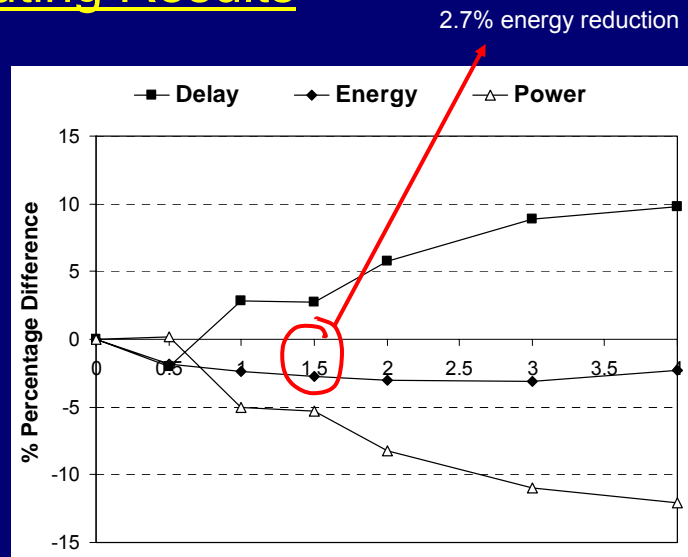
$$\text{Cost}(n) = \text{Crit}(i, j) \cdot \text{delay}(n) + (1 - \text{Crit}(i, j)) \cdot \text{cong}(n)$$

Power-Aware Cost Function:

$$\text{Cost}(n) = \text{Crit}(i, j) \cdot \text{delay}(n) + (1 - \text{Crit}(i, j)) \cdot [(1 - \text{ActCrit}(i)) \cdot \text{cong}(n) + \alpha \cdot \text{ActCrit}(i) \cdot \text{cap}(n)]$$

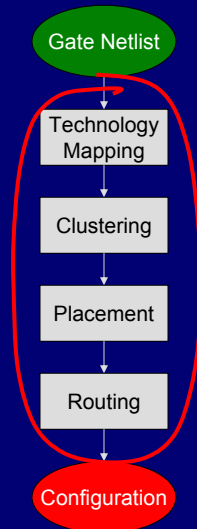
27

Routing Results



28

Combined Results



29

Combined Results

Tech. Map	power	baseline	baseline	baseline
Clustering	baseline	power	baseline	baseline
Placement	baseline	baseline	power	baseline
Routing	baseline	baseline	baseline	power
Energy (nJ)	2.01	1.90	2.11	2.12
% Diff	-7.6	-12.6	-3.0	-2.7

Ideal savings = -25.8%
Actual savings = -22.6%

Tech. Map	baseline	baseline	baseline	baseline	baseline	power	power	power	power	power	power	power
Clustering	baseline	baseline	power	power	power	baseline	baseline	baseline	power	power	power	power
Placement	baseline	power	baseline	power	power	baseline	power	power	baseline	baseline	power	power
Routing	baseline	power	power	baseline	power	power	baseline	power	baseline	power	baseline	power
Energy (nJ)	2.18	2.05	1.85	1.83	1.79	1.96	1.95	1.90	1.79	1.75	1.73	1.68
% Diff	0.00	-5.72	-14.8	-15.9	-17.9	-9.95	-10.2	-12.6	-17.6	-19.5	-20.6	-22.6

Energy savings are mostly cumulative!

30

Combined Results (overlap)

$\text{Overlap}(a,b) = \text{Individual}(a) + \text{Individual}(b) - \text{Combined}(a,b)$

Overlap (%)	Tech. Mapper	Clusterer	Placer
Router	0.27	0.04	-0.10
Placer	0.39	-0.33	
Clusterer	2.61		

31

Validation

Do our results mean anything?

We tested our clustering algorithm using:

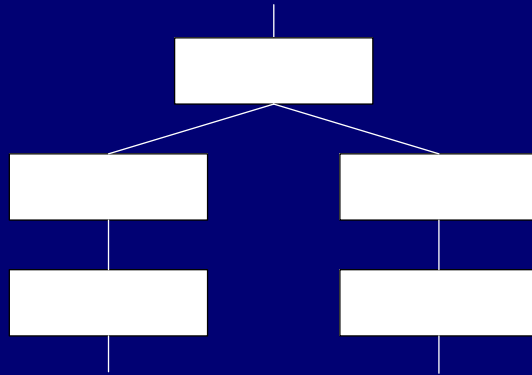
- An Altera Stratix (0.13 um) chip
- Industrial Place and Route tools (Quartus)

Altera's Quartus University Interface Program (QUIP):

- Allowed us to integrate our cluster algorithm with Quartus

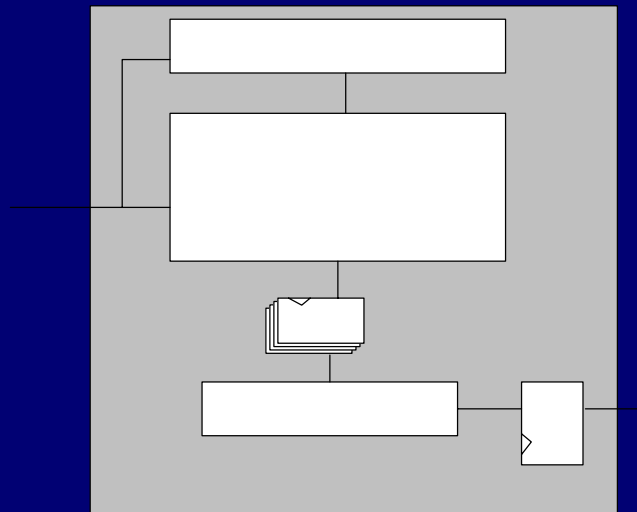
32

Validation (setup)

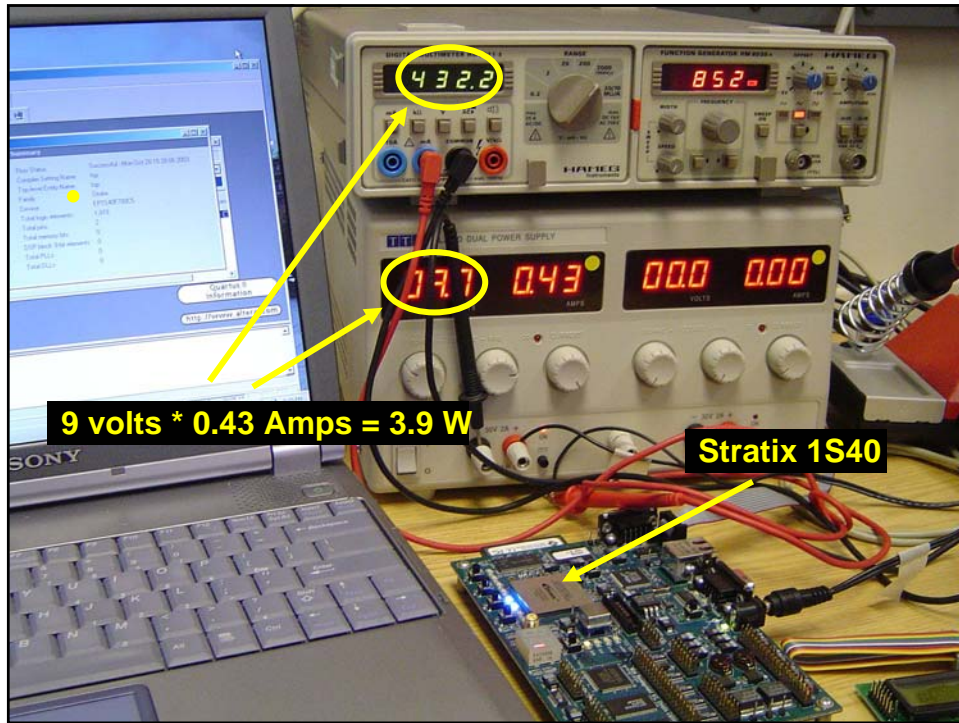


33

Validation (setup)



34



Validation (clustering results)

Circuit	Total Current		Static Current	Dynamic Current		% Improvement
	Non-Power-Aware	Power-Aware		Non-Power-Aware	Power-Aware	
alu4	259	260	222	37	38	-2.7 %
apex4	237	236	222	15	14	6.7 %
diffeq	242	240	222	20	18	10.0 %
elliptic	272	267	222	50	45	10.0 %
ex5p	240	237	222	18	15	16.7 %
ex1010	245	241	222	23	19	17.4 %
frisc	258	252	222	36	30	16.7 %
misc3	251	247	222	29	25	13.8 %
pdcc	251	243	222	29	21	27.6 %
s298	249	244	222	27	22	18.5 %
s38417	320	316	222	98	94	4.1 %
s38584	366	337	222	144	115	20.1 %
seq	256	252	222	34	30	11.8 %
spla	249	243	222	27	21	22.2 %
tseng	240	241	222	18	19	-5.6 %
					Average	12.5 %

Contributions

- Technology mapping and clustering are most suited to power minimization
(7.6% and 12.6% energy reduction respectively)
- Energy reductions are mostly cumulative
[only 3.2% overlap (22.6% instead of 25.8%)]
- Reduced overall energy by 22.6%

37