Effective Power Optimization combining Placement, Sizing, and Multi-Vt techniques

Tao Luo, David Newmark*, and David Z Pan

Department of Electrical and Computer Engineering,
University of Texas at Austin

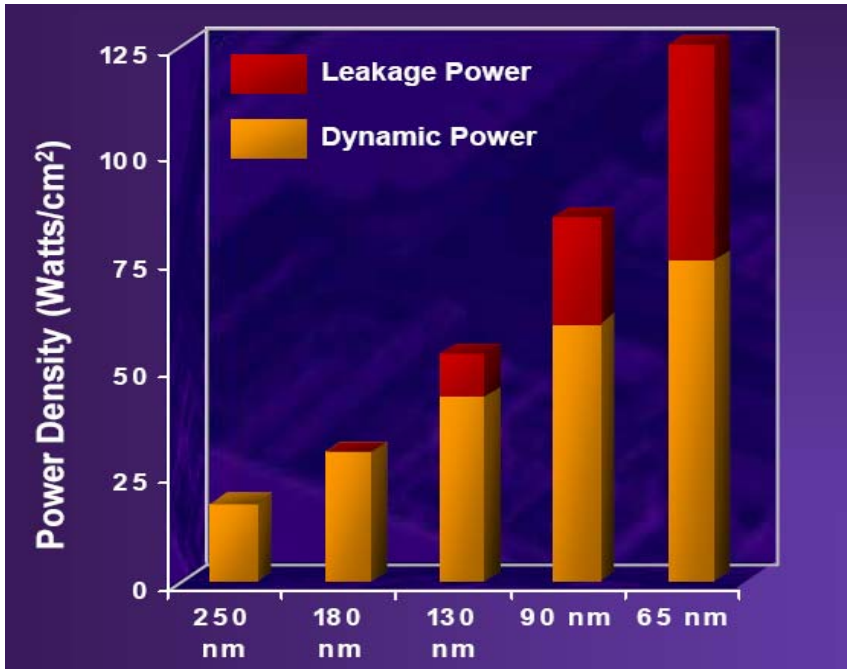
*Advanced Micro Devices, INC.

Overview

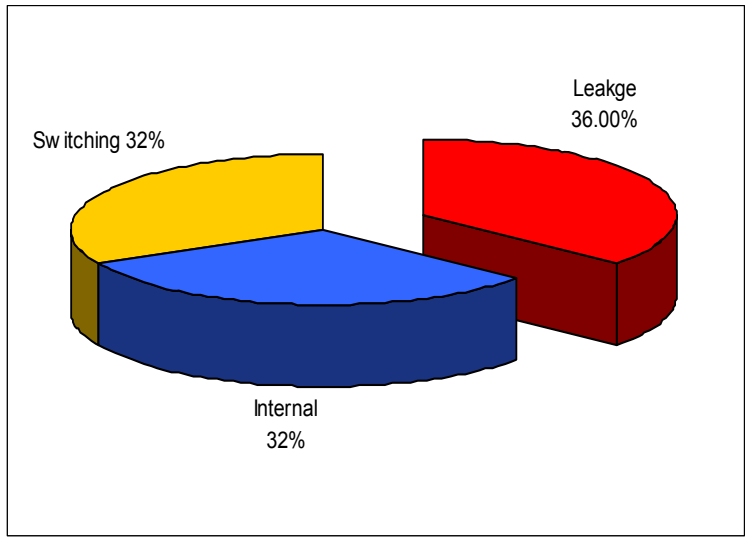
- Introduction
  - Power dissipation increases dramatically
    - As the technology scales down.
  - To maximize the power reduction
    - We combine placement, gate sizing, and multi-Vt techniques
    - How to integrate techniques with disjoint objectives?
- The proposed methodology
  - Integrate placement, gate sizing and Vt-swapping
  - Linear Programming based placement for power
  - Geometric Programming based gate sizing for power
    - Gate sizing on critical paths
    - Gate sizing on non-critical paths
- Experimental results
- Conclusions

Power Consumption Trend

- Dynamic power still dominates
- The percentage of the leakage power increases as technology scales down



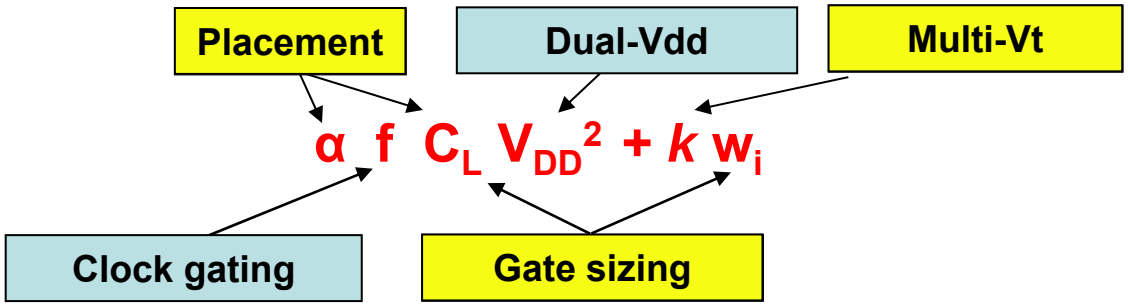
Source: Synopsys, Inc.



AMD 65 nm

Power Optimization Techniques

- Power optimization techniques
  - $P_{switch} + P_{leak} = \alpha\ f\ C_L\ V_{DD}^2 + k\ w_i$
  - α : switching factor
  - f : frequency
  - C_L : capacitive load
  - k : parameter for leakage
  - W_i : size of gate i



Maximize the Power Reduction

- Physical synthesis techniques that affect the power
  - Placement
    - Reduce the length of nets with high switching rate
  - Gate sizing
    - Choose the minimum gate sizes satisfying the timing constraints
  - Vt swapping:
    - LowVt (LVT) is used to close timing on critical paths
    - HighVt (HVT) is used to reduce leakage on non-critical paths
    - RegularVt (RVT) for other cells.

Maximize the Power Reduction

- How to integrate above techniques to maximize the power reduction?
  - Timing and power are often conflict objectives
  - Those techniques can be used for either timing and power
  - Above techniques affect each other
    - Using smaller gate sizes improves the power but increases the delay.
    - HVT reduces the power but hurts the timing.
    - We can use large gate sizes and less LVTs, or smaller gate sizes and more LVTs, etc.

Improve the Slack Distribution

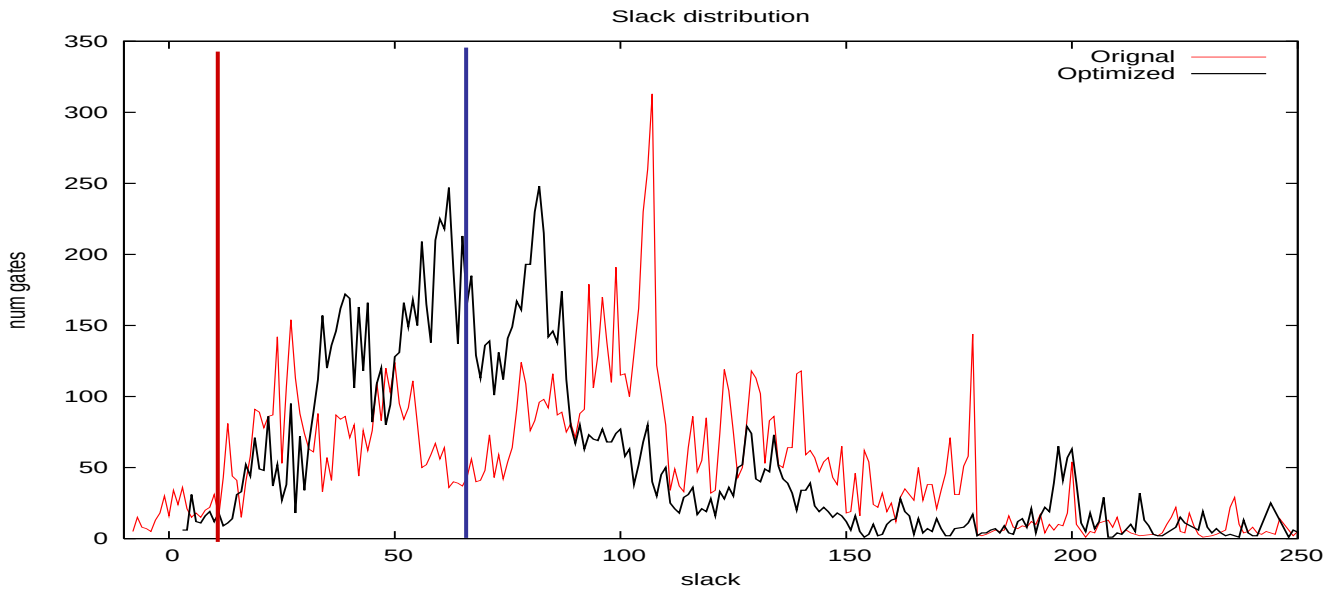
- Vt swapping is the most effective technique for leakage reduction
  - Using high threshold voltage has exponential saving on leakage
    - Leakage current is exponential to Vt, and linear to gate size

AMD 65-nm cell			
	LVT	**RVT**	**HVT**
Delay	1.0	1.1	1.3
Leakage	17.3	2.4	1.0

- We may use gate sizing to help Vt-swapping

Improve the Slack Distribution

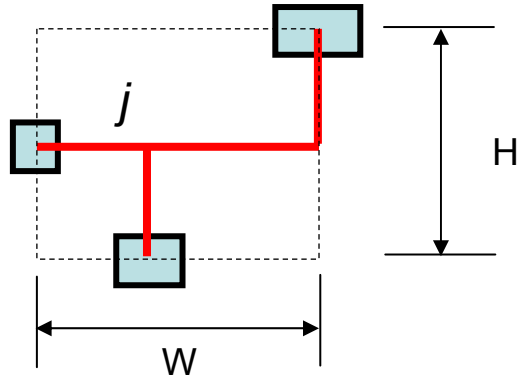
- The effectiveness of Vt-swapping relies on the slack profile.
  - Integrate those techniques through the slack profile management
  - Placement and gate sizing are formulated to improve the slack profile
    - For more effective Vt swapping finally
    - use more HVTs and less LVTs overall



Linear Program formulation

- Variables: cell coordinates

$$l_j \le x_i + pin_x(i,j)$$
$$r_j \ge x_i + pin_x(i,j)$$
$$t_j \le y_i + pin_y(i,j)$$
$$b_j \ge y_i + pin_y(i,j), \qquad i = 1,2,..,k$$
$$L_j = r_j - l_j + t_j - b_j$$
$$Cap_j = c \cdot L_j + Cpin_j$$



- Delay models

$$Dp_i = dp_I + a1_i \cdot Slew_i + a2_i \cdot Cap_i$$
$$Sp_i = sp_I + u1_i \cdot Slew_i + u2_i \cdot Cap_i$$

LP based Power Aware Placement

- Select a set of critical paths
  - Linear program to minimize the sum of weighted nets.
  - Two major components on net weights
    - The timing weight is based on the delay sensitivity of the net.
    - The power weight is based on nets switching factor

$$\min \quad \sum wt_j L_j$$

$$\forall j \in Selected \qquad critical \qquad nets$$

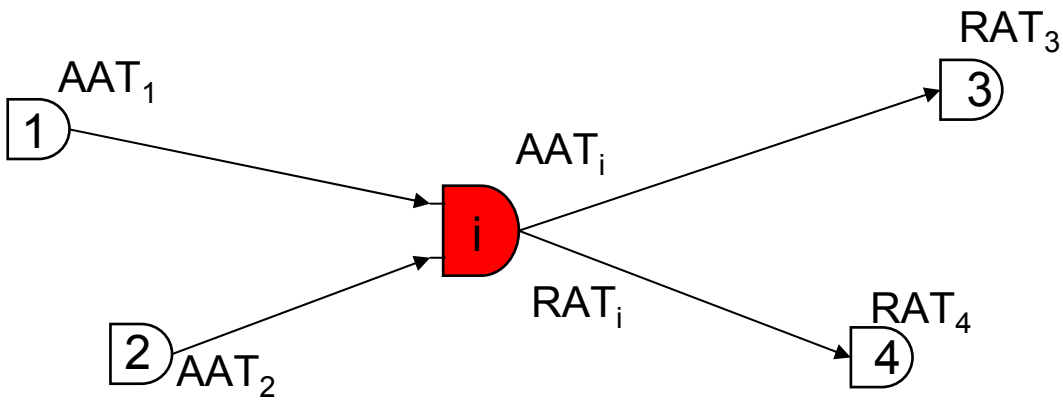
$$wt_p = 0.5\alpha_i \cdot F \cdot V^2$$

$$wt_t = a2_i + a1_{i+1} \cdot u2_i$$

$$wt_j = \beta wt_p + (1-\beta) wt_t$$

Geometric Programming (GP) Preliminary

- Static timing analysis
  - $AAT_i = \max\{AAT_1, AAT_2\}+D_i$
  - $RAT_i = \min\{ RAT_3-D_3,\ RAT_4-D_4\}$
  - $D_i = d_i + h_i/W_i$
  - W_i: the size of gate i



GP based Gate Sizing for Near Critical Cells

- Effort based delay models

$$Dg_i = dg_I + (h_i / W_i) \cdot Cap_i$$

$$Sg_i = sg_I + (v_i / W_i) \cdot Cap_i$$

$$Cap_i = \sum_{k=1}^{a} (e_k + f_k W_k) + \sum_{l=1}^{b} (Cp_l) + Cap_{wire}$$

$$Cp_i = e_i + f_i W_i$$

$$P_i = 0.5\alpha \cdot F \cdot V^2 \cdot Cap_i$$

$$L_i = leak_i \cdot W_i$$

- For critical and near critical cells
  - Maximize the sum of slack on critical primary outputs
  - Increase the size of critical cells and push the slack of near critical cells to be more positive, but not larger than a threshold.

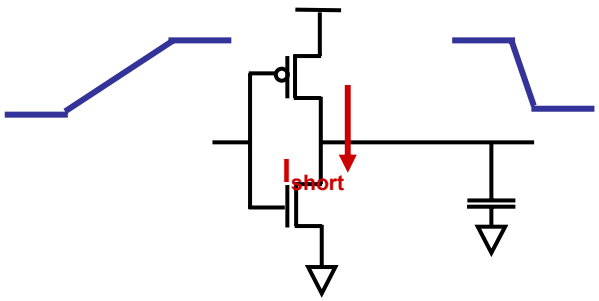
GP Gate Sizing for Non-Critical Cells

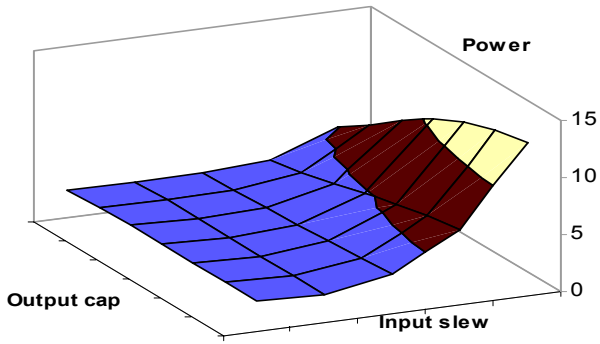
- Sizing down cells with large slacks directly

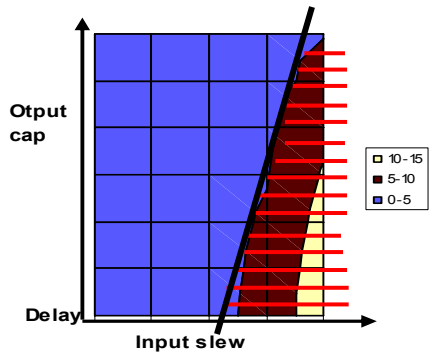
$$\min \quad \sum(\Delta Pg_i / \Delta W_i) \qquad i \in NC$$

$$s.t. \quad AT_i \geq \max\{(T_{cycle} - T_{threshold}), AT_{orig_i}\}, i \in PO$$

- Additional constraints
  - Short-circuit power constraint,
    - Short circuit power could be large if not properly controlled

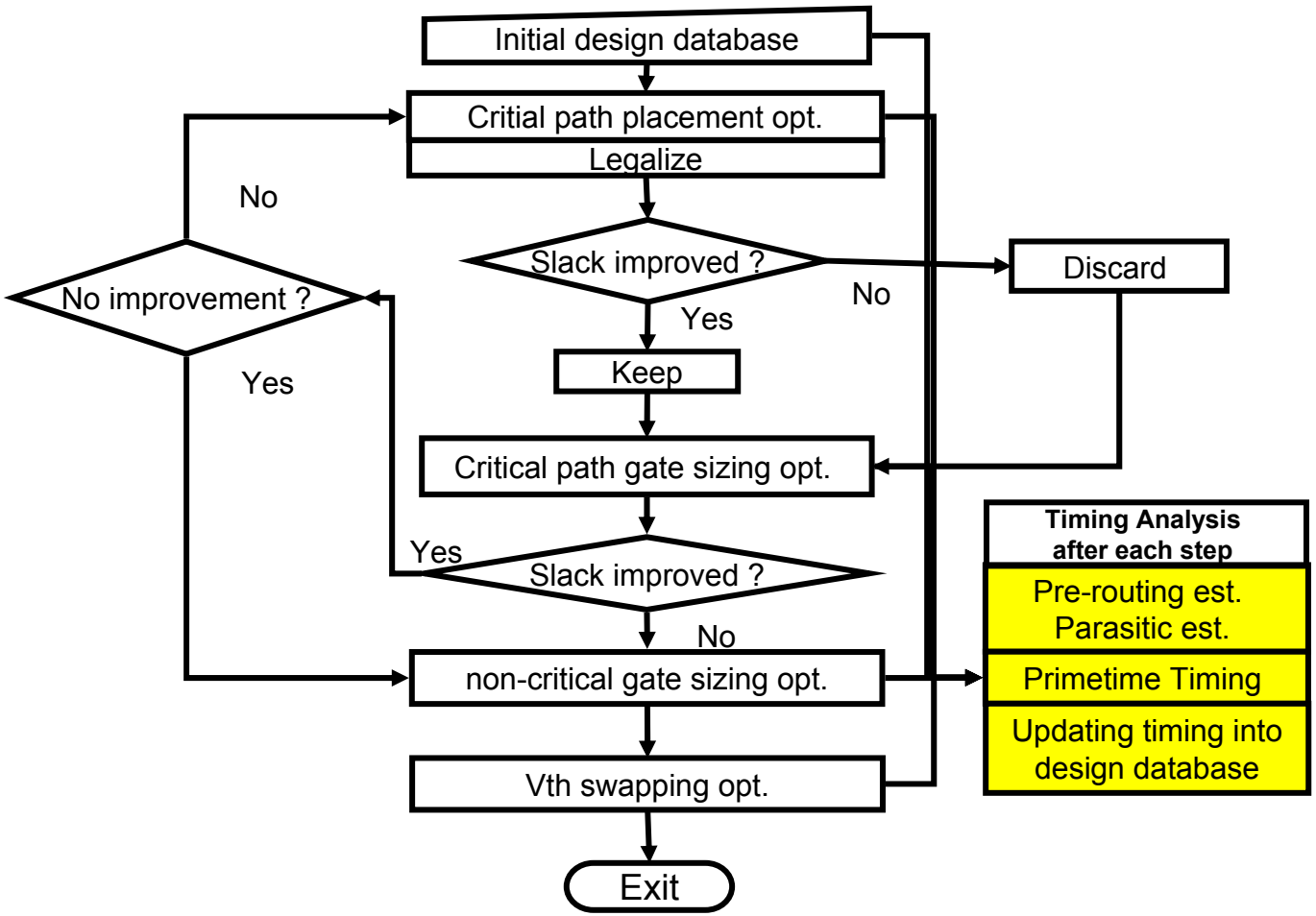






  - Maximum slew constraint
  - Effective fan-out constraint (control the noise)

The Overall Flow



Experimental Results

Table 1: Total power comparison

	65 nm		Total power (mw)				Improvement %		
	Gates	Nets	Base	VT	PV	PGV	*VT|Base*	*PV|Base*	*PGV|Base*
ckt1	1765	2360	29.79	24.60	22.06	20.52	17.4	25.9	31.1
ckt2	2334	2881	30.26	22.41	21.93	19.69	25.9	27.5	34.9
ckt3	6640	8644	142.51	103.45	101.31	96.11	27.4	28.9	32.6
ckt4	9254	7928	110.86	93.57	93.57	86.38	15.6	15.6	22.1
ckt5	9541	9539	233.56	151.81	147.23	123.40	35.0	37.0	47.2
ckt6	12716	14042	241.27	155.89	154.14	140.43	35.4	36.1	41.8
ckt7	15486	18360	287.22	233.63	226.96	217.14	18.7	21.0	24.4
ckt8	27103	26991	499.15	377.34	372.51	354.58	24.4	25.4	29.0
							25.0	27.2	32.9

Table 2: Leakage power comparison

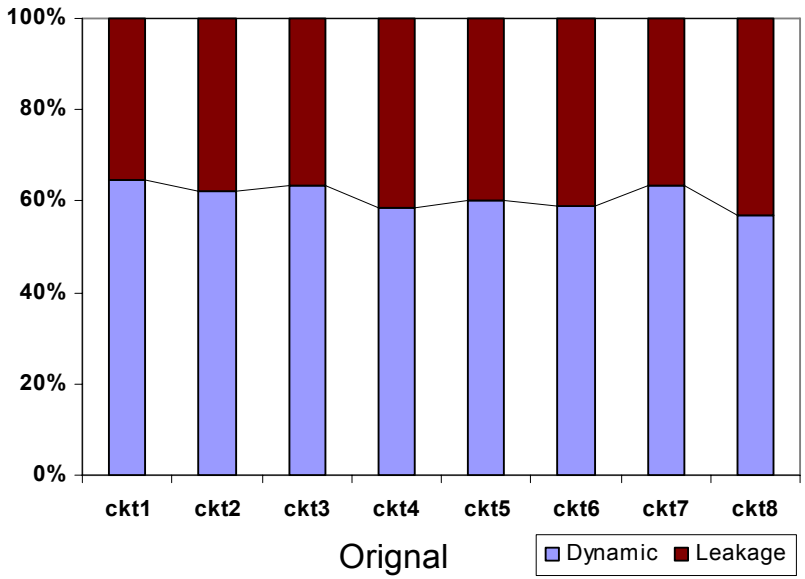
	Base	VT	PGV	*VT|Base %*	*PGV|Base %*
ckt1	10.50	6.09	3.28	42.0	68.8
ckt2	11.49	4.79	3.67	58.3	68.1
ckt3	52.11	20.10	17.38	61.4	66.6
ckt4	45.76	30.42	28.06	33.5	38.7
ckt5	93.04	26.62	18.28	71.4	80.4
ckt6	99.29	24.78	19.64	75.0	80.2
ckt7	104.77	60.25	49.86	42.5	52.4
ckt8	215.24	108.46	96.28	49.6	55.3
				54.2	63.8

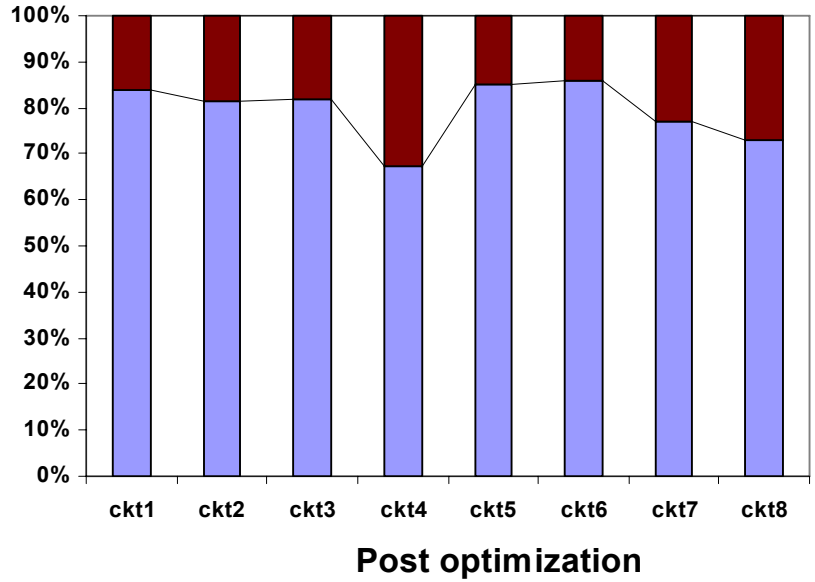
Table 3: Dynamic power comparison

	Base	VT	PGV	*VT|Base %*	*PGV|Base%*
ckt1	19.29	18.51	17.24	4.0	10.6
ckt2	18.77	17.62	16.02	6.1	14.7
ckt3	90.40	83.35	78.73	7.8	12.9
ckt4	65.10	63.15	58.32	3.0	10.4
ckt5	140.52	125.19	105.12	10.9	25.2
ckt6	141.98	131.11	120.79	7.7	14.9
ckt7	182.45	173.38	167.28	5.0	8.3
ckt8	283.91	268.88	258.30	5.3	9.0
				6.2	13.3

Leakage Power Breaks Down

- Leakage power consumption before and after optimization
  - Leakage is reduced effectively





Vt Percentages and Runtime

- Use a few low Vt cells only

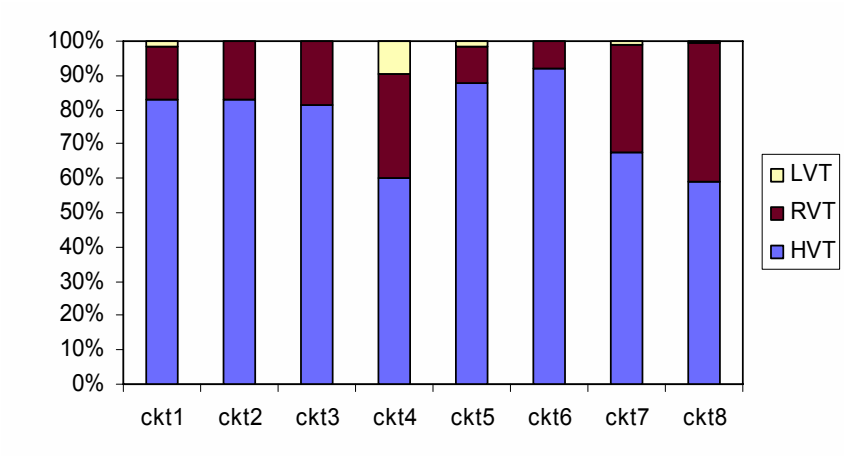


Table 4: Runtime breakup(s)

Testcases	*Place*	*Size*	*Swap*	*Timing*
ckt1	4	14	55	168
ckt2	3	12	69	90
ckt3	30	75	124	453
ckt4	25	42	137	217
ckt5	26	68	149	324
ckt6	32	228	171	262
ckt7	16	193	186	342
ckt8	43	384	245	538

Conclusions

- We propose the methodology to integrate several physical synthesis techniques
  - To maximize the leakage and total power reduction through the slack profile management.
- Our LP based placement and GP based gate sizing algorithms maximize the effectiveness of the multi-Vt swapping technique for power reduction.
- Various practical design constraints are included in our formulation
  - Short circuit power is not ignorable
- Our proposed method is very effective
  - Achieved 63.8% leakage power reduction and 32.8% overall power reduction on 65nm microprocessor circuits.