Peak Power and Dynamic IR-drop Assessment via Waveform Augmenting

Yihan Wen[1], Juan Li[1], **Bei Yu**[2], Xiaoyi Wang

[1] Beijing University of Technology
[2] The Chinese University of Hong Kong

Oct. 29, 2024

Outline

1. Introduction
2. Proposed Method
3. Experimental Results
4. Conclusion

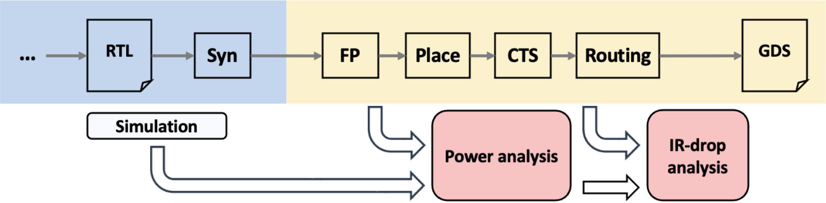
Introduction

Power & IR-drop analysis are critical

- Power consumption on instances and voltage reduction on the power grid
- Essential metrics of chips performance

Power & IR-drop analysis are still facing challenges

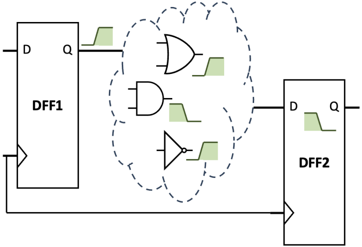
- Implemented in later stages of the chip design flow
- Coverage of the worst-case scenario



Previous Works – Dynamic IR-drop Analysis

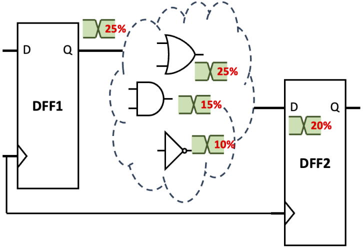
Vector-based analysis

Input vectors generated from gate-level simulation[1]



Vectorless analysis

Input vectors generated from instances toggle probability propagation[2,3]



---

[1] Rouatbi et al. "Power estimation tool for sub-micron CMOS VLSI circuits" 1992
[2] M. G. Xakellis et al. "Statistical estimation of the switching activity in digital circuits" 1994
[3] R. Marculescu et al. "Probabilistic modeling of dependencies during switching activity analysis" 1998

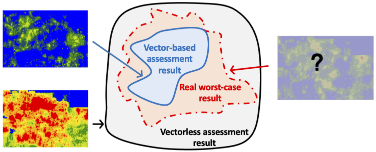
Defects of previous vector-based & vectorless analysis

Vector-based analysis

- Hard to cover the worst case
- Need waveform to activate true worst-case scenario
- Very challenging[4]

Vectorless analysis

- Overly pessimistic results
- Real waveforms only activate logic in a small region
- vectorless propagates toggles throughout the entire netlist[5]



[4]C.-T. Hsieh et al. "Vectorless estimation of maximum instantaneous current for sequential circuits" 2006

[5]S. Soman et al. "Ensuring On-Die Power Supply Robustness in High-Performance Designs" 2011

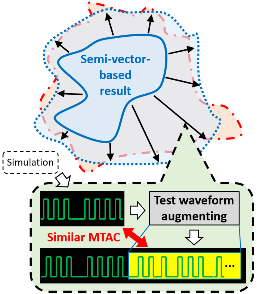
Proposed Methods

Proposed Method

Motivation:

Semi-vector-based analysis via waveform augmenting

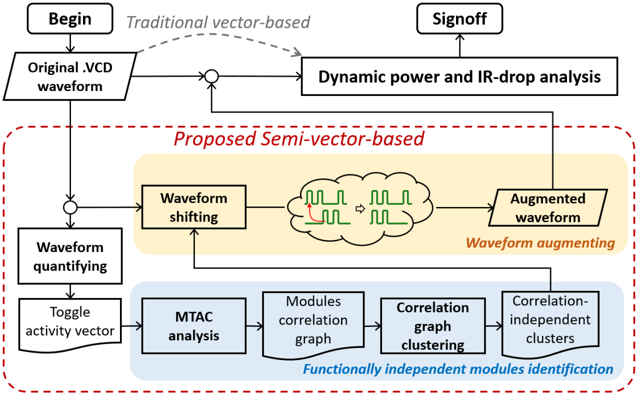
- **Step 1**: Analysis toggle statistics in existing simulation waveform
- **Step 2**: Identify modules with potential coverage risk
- **Step 3**: Augment waveform to cover the worst case



Assessment flow of semi-vector-based analysis

Basic idea

- How to measure the toggle statistics? → Calculate toggle correlation
- How to augment the waveform to worst case? → Build simultaneous toggle
- How to prevent pessimistic augmenting? → Keep similar toggle correlation



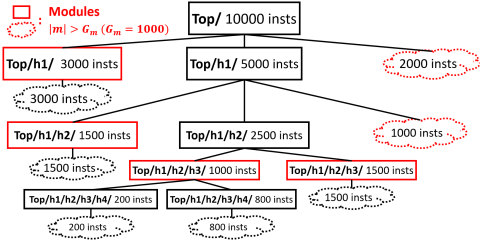
Step 1: Analysis toggle statistics in simulation waveform

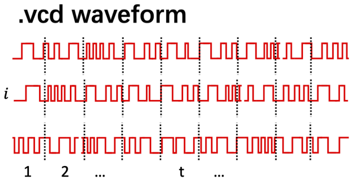
Modules quantifying

- Number of instances in a modern design can be vast
  → Analysis hierarchical instances (H-insts) rather than the flattened instances

Waveform quantifying

- Long waveform and sparse toggle events → Quantify toggle events into time slots





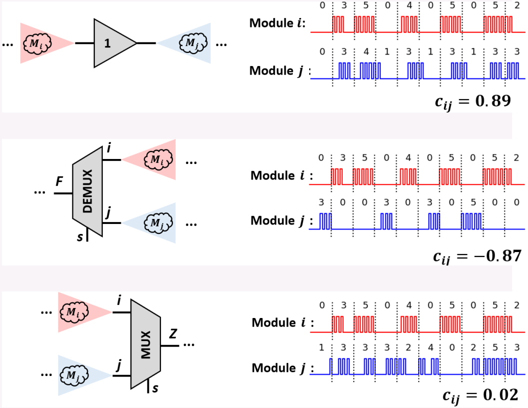
$$\mathbf{T}_i = [T_i^1 \quad T_i^2 \quad \cdots \quad T_i^t \quad \cdots \quad T_i^{\mathcal{T}}]$$

Step 1: Analysis toggle statistics in simulation waveform

Modules toggle activity correlation (MTAC)

- MTAC c_{ij} calculated based on the Pearson correlation coefficient
- c_{ij} measures the dependence relationship of toggle activity between modules
- Modules correlation graph $\boldsymbol{G}\{\boldsymbol{V},\boldsymbol{E}\} = K_n$, with modules correlation matrix $\boldsymbol{\mathcal{A}}_{\boldsymbol{G}}[i,j]_{i \neq j} = c_{ij} = c_{ji}$

$$c_{ij} = \frac{\mathrm{cov}(\mathbf{T}_i, \mathbf{T}_j)}{\sigma_{\mathbf{T}_i}\sigma_{\mathbf{T}_j}}$$



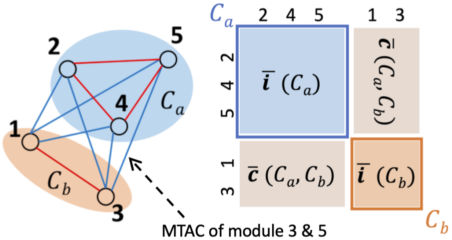
Step 2: Identify modules with potential coverage risk

Identify functionally independent modules clusters

- Correlation-independent modules with low absolute MTAC value
- Find clusters set $\boldsymbol{C} = \{C_1, \cdots, C_N\}$: correlation graph segmentation with c_ϵ
  $\bar{\mathbf{c}}(C_a, C_b) \le c_\epsilon \bar{\mathbf{i}}(C_a) \ge c_\epsilon,\ \bar{\mathbf{i}}(C_b) \ge c_\epsilon \forall C_a \in \mathbf{C}, \forall C_b \in \mathbf{C}, a \neq b$
- $\bar{\mathbf{c}}$ and $\bar{\mathbf{i}}$ defined similarly to the graph *cut* [6]

$$\bar{\mathbf{c}}(C_a, C_b) = \frac{1}{|C_a||C_b|} \sum_{i \in C_a, j \in C_b} |c_{ij}|$$

$$\bar{\mathbf{i}}(C_a) = \frac{2}{|C_a|^2 - |C_a|} \sum_{p,q \in C_a, p \neq q} |c_{pq}|$$

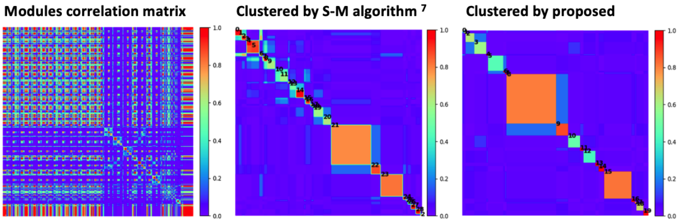


[6] Z. Wu et al. "An optimal graph theoretic approach to data clustering: Theory and its application to image segmentation" 1993

Step 2: Identify modules with potential coverage risk

Clustering algorithm: two-stage flow

- Extended from the S-M algorithm[7]
- SPLIT stage: recursively breaks the modules to satisfy $\bar{\mathbf{c}}(C_a, C_b) \leq c_\epsilon$
- MERGE stage: recursively merges pairs of clusters to satisfy $\bar{\mathbf{i}}(C_a) \geq c_\epsilon$

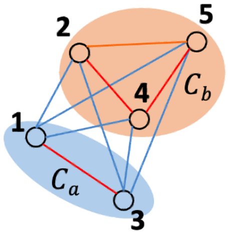


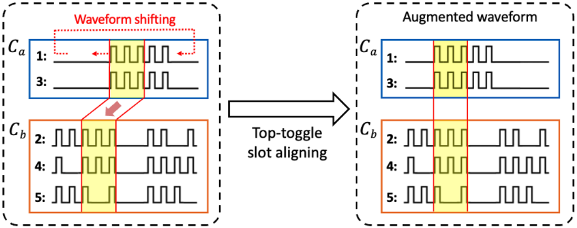
[7] J. Shi et al. "Normalized cuts and image segmentation" 2000

Step 3: Augment waveform to cover the worst case

Worst-case approximation by waveform augmenting

- Assuming correlation-independent modules switching simultaneously
- Aligning the highest switching slots by waveform shifting process





Experimental Results

Experimental results and discussion

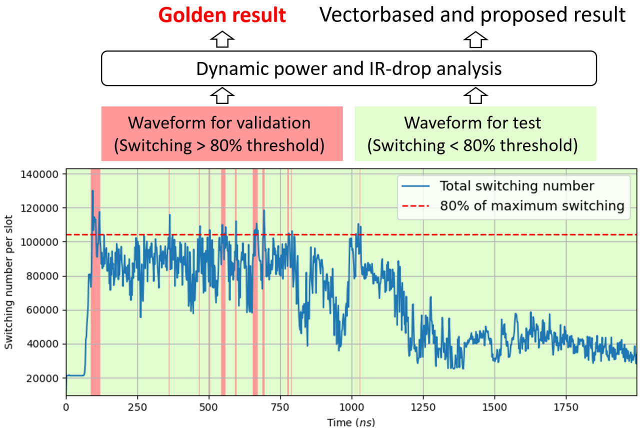
Validated on ARM CPU design blocks

- Design blocks type: CPU core, SoC core, cache, interconnection controller
- Low toggle correlation disagreement between the original waveform and the augmented waveform: corr.error $= \frac{2}{n^2-n}\sum_{ij}(c^o_{ij} - c^a_{ij})^2$

	Type	Instances number	Modules number	Clusters number	Augmented corr. error
1	CPU-core	600000+	200+	7	0.0058
2	CPU-core	1000000+	100+	6	0.0007
3	CPU-core	1500000+	1000+	3	0.0009
4	Cache	1000000+	500+	13	0.0081
5	SoC-core	3000000+	1500+	15	0.0029
6	Soc-core	1500000+	1000+	13	0.0018
7	Inter	1000000+	1000+	3	0.0030
8	Inter	1000000+	1000+	3	0.0024

Validation & test Waveform

- Validation waveform: segments with top 20% total switching count → **Concealed**
- Test waveform: remaining parts → **Input of experiments**



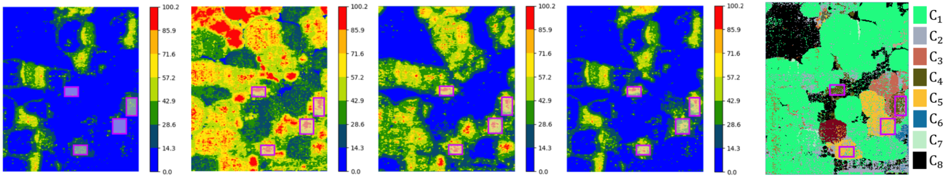
Worst-case dynamic power and IR-drop estimation results

Worst-case value estimation

- Proposed semi-vector-based assessment flow yields more reasonable results
- Peak power error: **proposed 2.93%** (V-based 8.98% & V-less 99.12%)
- Worst dynamic IR error: **proposed 7.62%** (V-based 19.05% & V-less 50.61%)

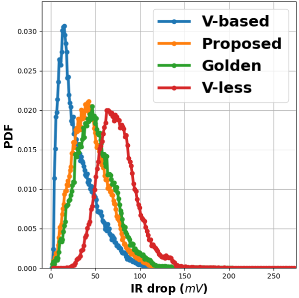
	Worst peak-power (mW)				Worst dyn-IR-drop (mV)				Runtime (Minutes)		
	Golden	V-based	V-less	Proposed	Golden	V-based	V-less	Proposed	V-based	V-less	Proposed
1	1372	1308	2905	**1352**	157.8	131.9	208.9	**143.4**	128	52	136
2	3175	2984	4333	**3254**	191.8	160.6	165.8	**189.6**	201	93	242
3	4371	4113	7983	**4468**	203.0	193.0	261.8	**210.6**	258	107	283
4	1981	1704	3428	**1928**	151.5	100.6	247.6	**147.1**	111	81	124
5	2435	2199	4489	**2449**	167.2	158.3	274.7	**173.3**	436	172	488
6	455	356	1411	**407**	119.6	108.1	191.0	**127.1**	161	130	179
7	2536	2380	5020	**2589**	149.2	92.3	282.3	**174.4**	322	170	357
8	2433	2352	4809	**2464**	159.0	114.2	244.4	**187.0**	281	144	321

Worst-case dynamic power and IR-drop estimation results



Vector-based dynamic IR | Vectorless dynamic IR | Proposed dynamic IR | Ground truth | Correlation-independent clusters

- Vector-based result failed to identify all IR-drop weak regions (highlighted in purple boxes)
- Vector-based mismatch regions appeared in the intersection of clusters
- Vectorless result exhibited overly pessimistic
- Proposed result is more accurate in both weak regions estimation and PDF of IR value



Conclusion

Conclusion

- Semi-vector-based assessment is proposed
- Functionally independent modules are through the analysis of modules toggle activity correlation
- Waveform is augmented by building simultaneous toggle of functionally independent modules
- Worst-case coverage of power and dynamic IR-drop assessment is improved

THANK YOU!