Congestion-Aware Power Grid Optimization for 3D circuits Using MIM and CMOS Decoupling Capacitors

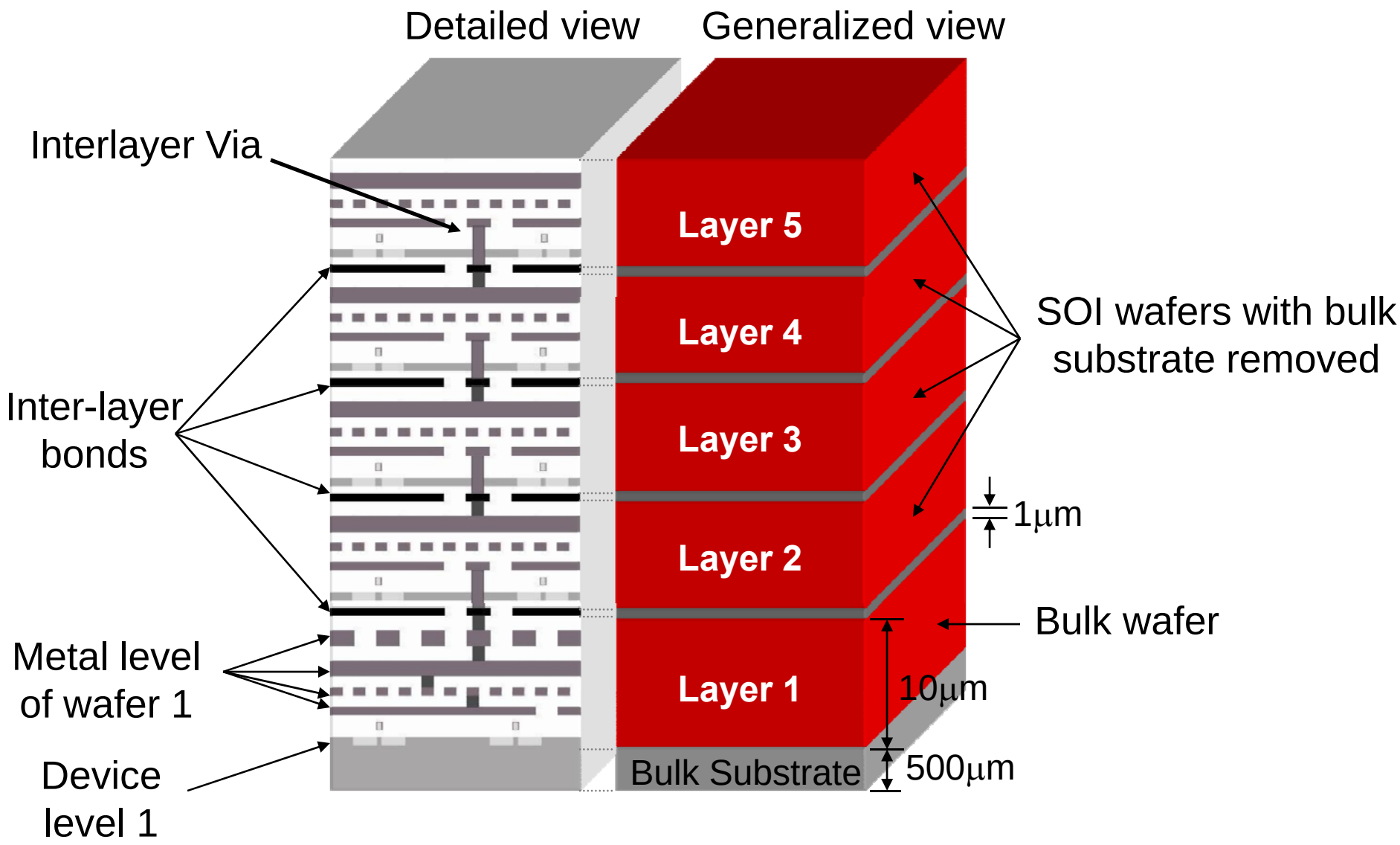
Pingqiang Zhou
Karthikk Sridharan
Sachin S. Sapatnekar

University of Minnesota

Outline

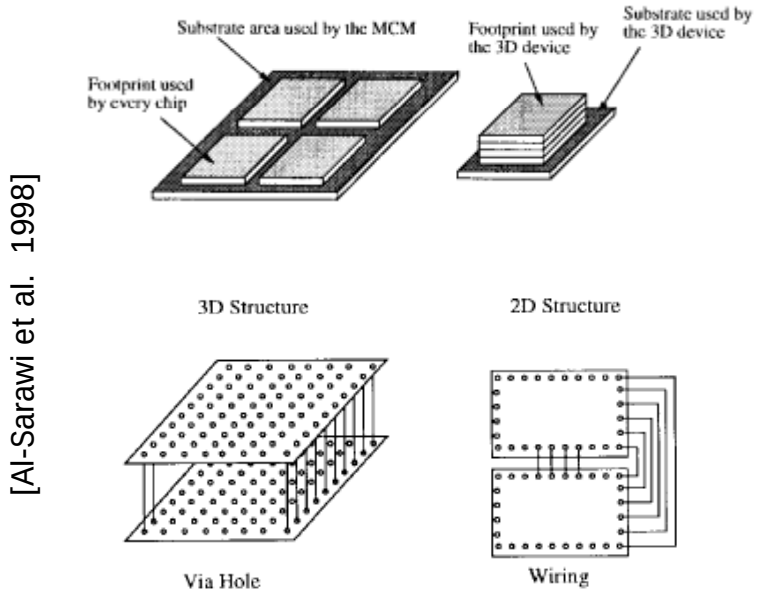
- Motivation
- A new CAD solution to 3D power grid optimization
- Experimental results
- Summary

3D IC Design

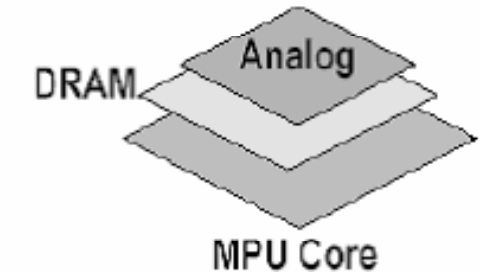

Detailed view
Generalized view
Interlayer Via
Inter-layer bonds
Metal level of wafer 1
Device level 1
Layer 5
Layer 4
Layer 3
Layer 2
Layer 1
Bulk Substrate
SOI wafers with bulk substrate removed
1μm
Bulk wafer
10μm
500μm

3D Integration: Driving forces

- **Improved global interconnect performance**
- **Reduce footprint / improve packing density**
- **Mixed-signal integration**

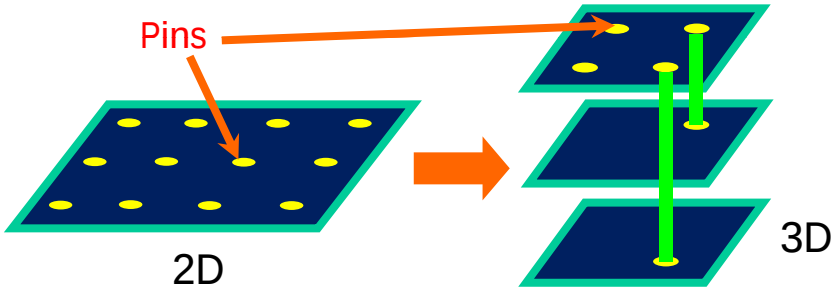


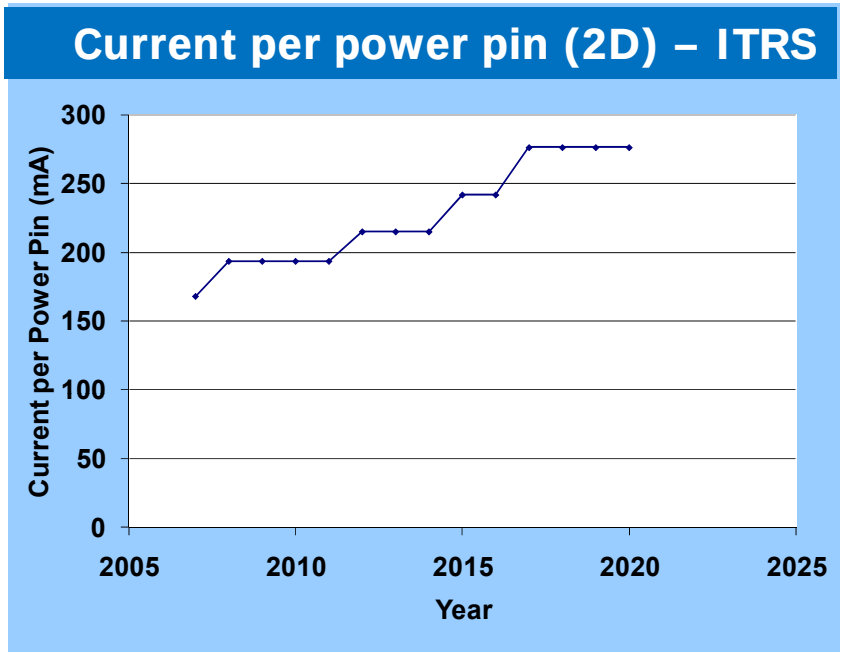
[Al-Sarawi et al. 1998]



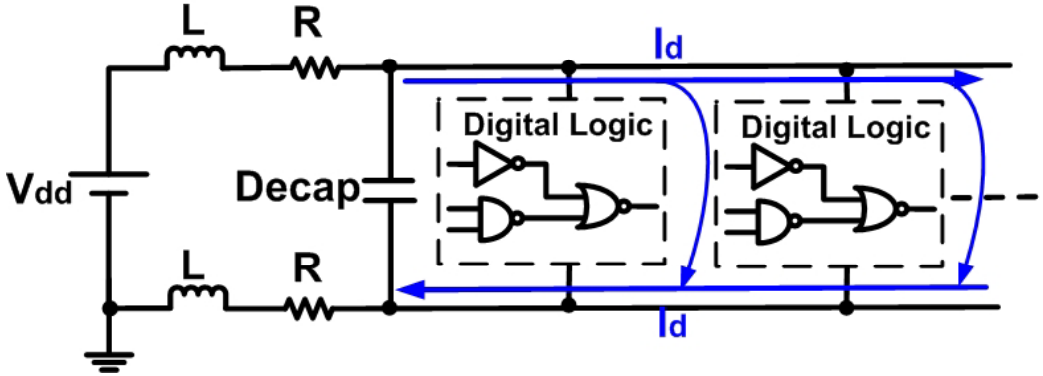
Power Supply Integrity in 3D

- Higher current density, faster current transients worsen supply noise
- Greater challenge in 3D due to via resistance, limited number of supply pins

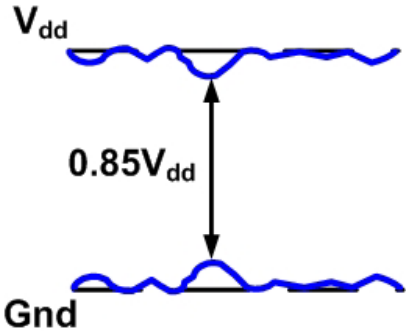




Power bottleneck: a major problem for 3D



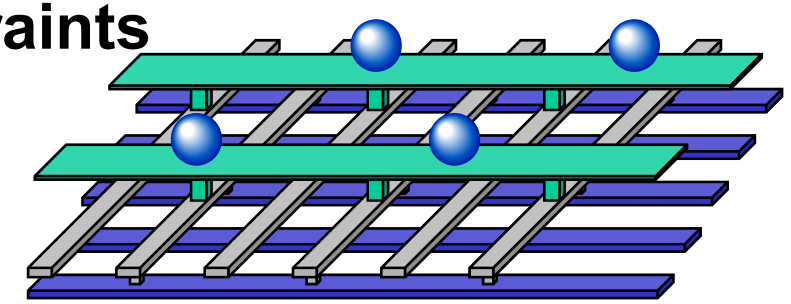
3D Circuit



[IBM]

Traditional power delivery

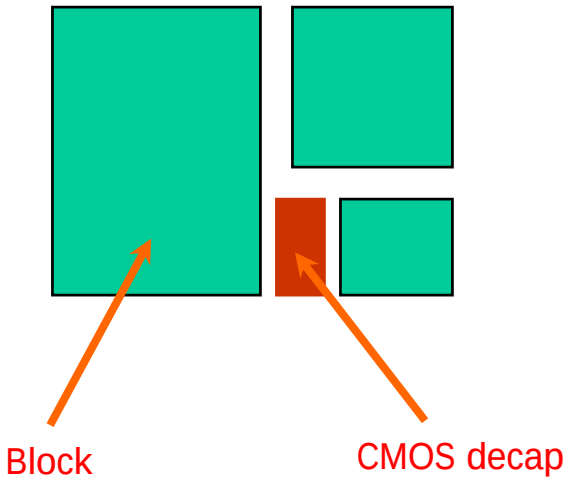
- **Requirements**
  - V_{dd}, GND signals should be at correct levels (low V drop)
  - Electromigration constraints
    - Current density must never exceed a specification
    - For each wire, $I_i/w_i < J_{spec}$
  - dI/dt constraints
    - Need to manage dI/dt to reduce inductive effects
- **Techniques for meeting constraints**
  - Widening wires
  - Using appropriate topologies
  - Adding decoupling capacitances



- **Already challenged for 2D technologies**
  - Reliable power delivery hard
  - Decaps get leaky
- **New CAD approaches necessary**

Decoupling capacitances (decaps)

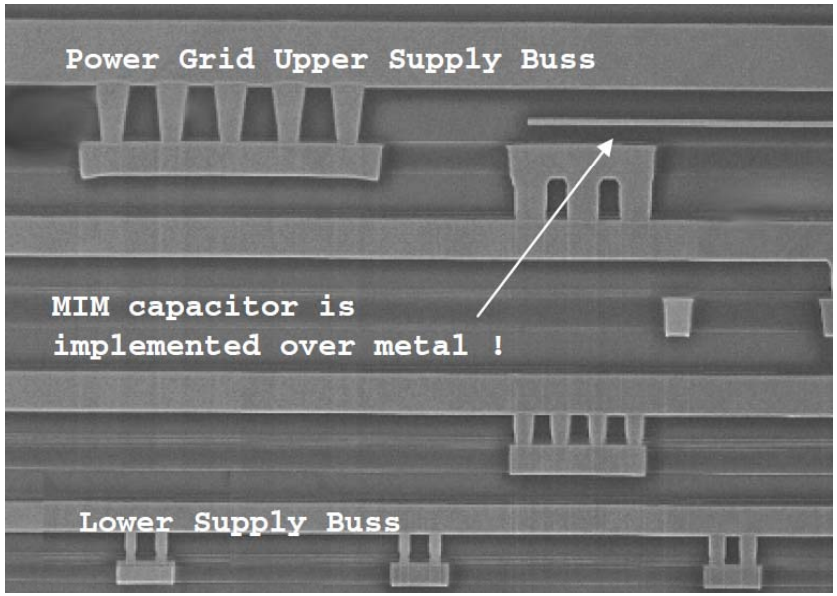
- The most powerful method to reduce transient noise
- Conventional decap technology: CMOS decap
- New considerations for CMOS decaps in 3D
  - Compete for area on device layer with landing pads of 3D vias
  - May increase footprint size
  - Get more leaky, due to T-leakage feedback

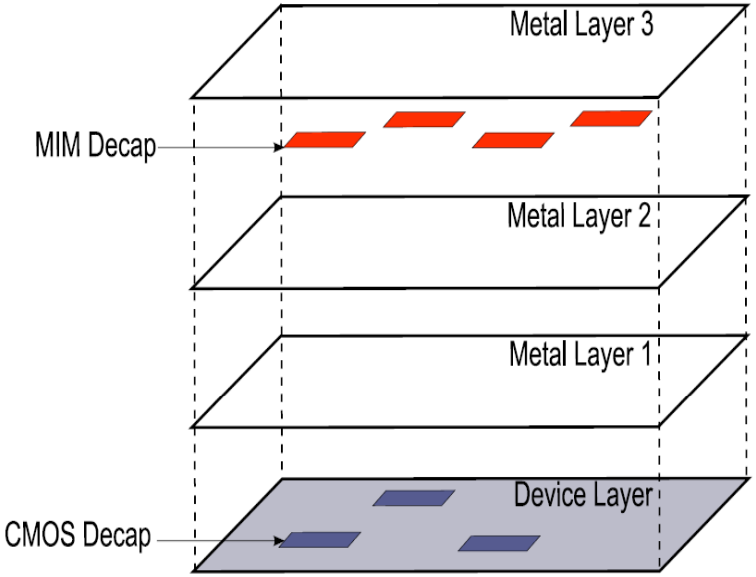


Any other option?

MIM decaps

[Roberts 2005]



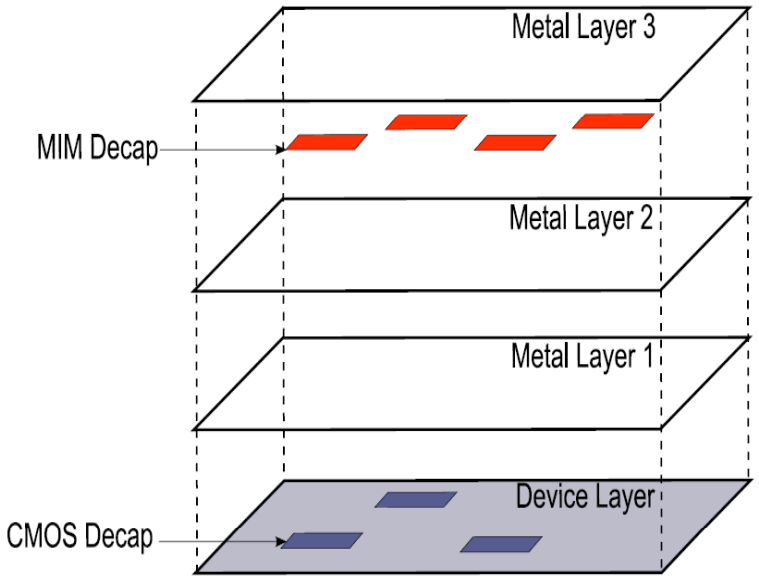


Decap	*Capacitance density (fF/μm²)	*Leakage density (A/cm²)	Congestion
CMOS	17.3	1.45e-4	—
MIM	8.0	3.2e-8	routing blockage

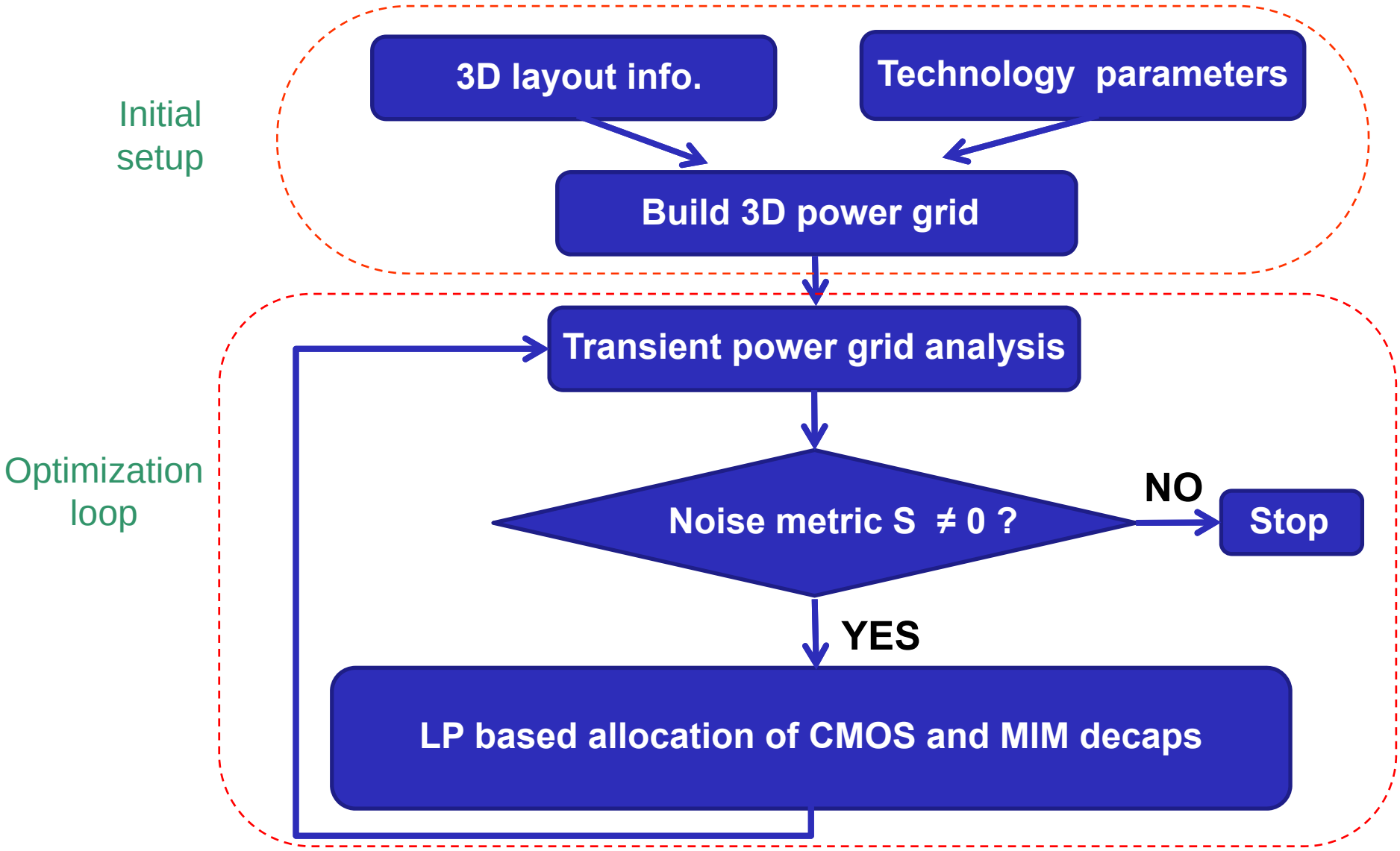
* Numbers deduced from Roberts et al., IEDM05 and PTM simulations

Our Contributions

- **Apply newer decap technology - MIM decap**
- **Develop CAD solutions for inserting both MIM and CMOS decaps:**
  - Sequence of linear progamming based problem formulation
  - Linearized noise model based on adjoint sensitivity analysis
  - 3D congestion analysis and linear congestion model

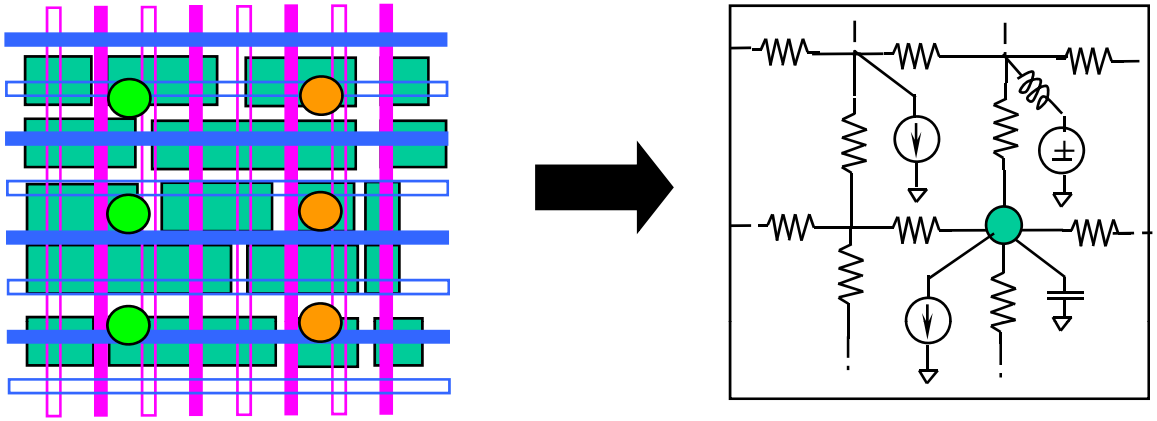


Overall Algorithm Flow



Power network modeling and analysis

- **Power Network Modeling**



- **Modified Nodal Analysis**

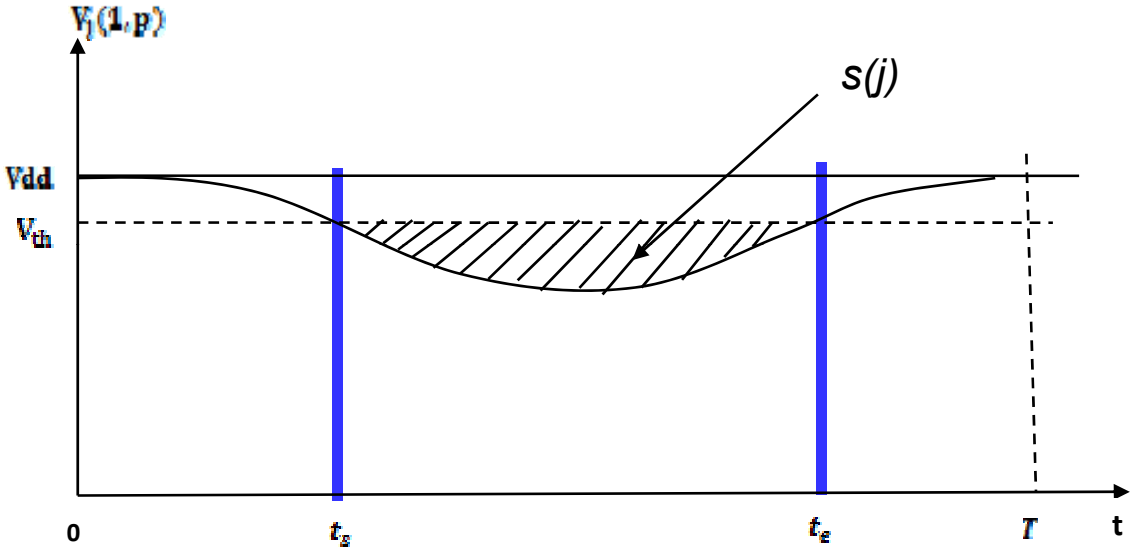
$$G\,x(t) + C\,x'(t) = b(t)$$

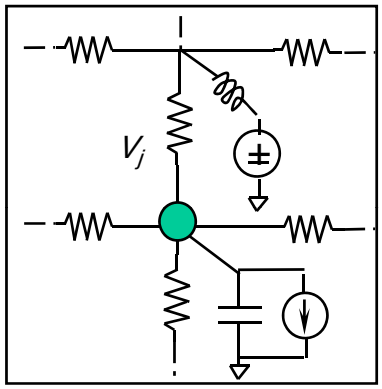
- x(t): time varying vector of voltages and currents
- b(t): time varying vector of independent current sources

- **Adjoint Sensitivity Analysis**
  - Based on Tellegen's theorem: the instantaneous power in any circuit is zero
  - An approach to calculate the sensitivity of one objective function w.r.t all the parameters in the circuit

Power Noise Metric *S*

- **Noise: optimize the integral of noise violation over time**





Waveform of node *j* on VDD grid

$$\begin{cases} S(j) = \int_{t_s}^{t_e} (v_{th} - v_j(1,p))dt \\ S = \sum_j S(j) \end{cases}$$

Decap optimization: problem formulation

Noise metric

Leakage power

$$\text{minimize} \quad \alpha\, S(x_k, y_k) + (1-\alpha)\, P(x_k, y_k)$$

subject to

Decap resource constraint

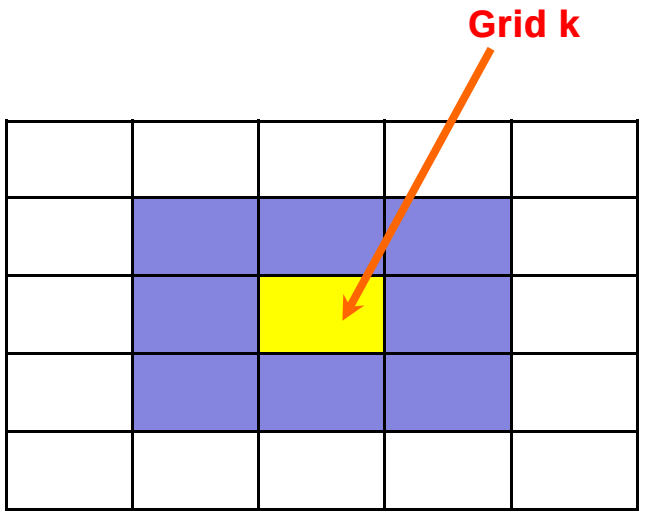
$$0 \le x_k \le C^k_{CMOS}$$

Congestion constraint

$$0 \le y_k \le C^k_{MIM}$$

Congestion in grid k ≤ 1

- x_k : CMOS decap added to grid *k*
- y_k : MIM decap added to grid *k*



Nonlinear optimization problem!

Sequence of linear programs: formulation

- **Objective**

$$\min \quad \alpha \Delta S + (1-\alpha) \Delta P$$

  - $\Delta S = \sum_k (a_k \Delta x_k + b_k \Delta y_k)$ = change of violation area S
  - $\Delta P = \sum_k (c_k \Delta x_k + d_k \Delta y_k)$ = change in leakage
  - Δx_k : Newly added CMOS decap to grid k
  - Δy_k : Newly added MIM decap to grid k

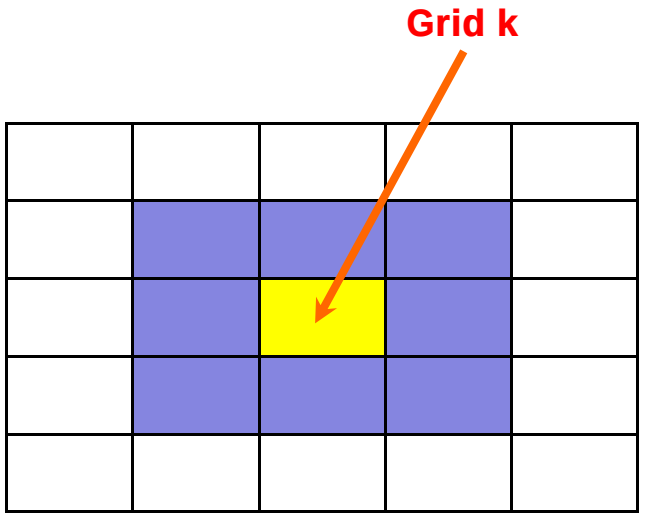
- **Constraints**
  - **Congestion constraint**

$$\Delta Cong_k \leq \gamma \cdot Cong_k$$

  - **Decap resource constraint**

$$0 \leq \Delta x_k \leq \min\{\Delta_{CMOS}, C^k_{CMOS}\}$$

$$0 \leq \Delta y_k \leq \min\{\Delta_{MIM}, C^k_{MIM}\}$$



Congestion Analysis and Linear Model

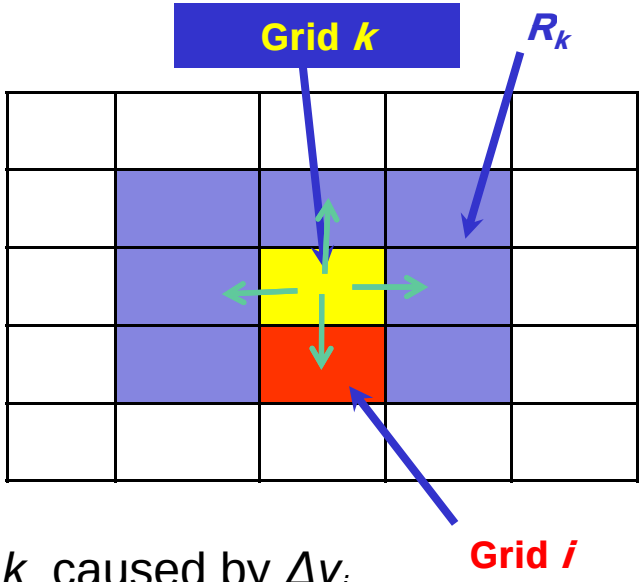
- **3D congestion analysis**
  - Extension of "Estimation routing congestion using probabilistic analysis" , [Lou, et al. TCAD'02].

$$\begin{cases} F(p,q,r) = F(p-1,q,r) + F(p,q-1,r) + F(p,q,r-1) \\ F(p,1,1) = F(1,q,1) = F(1,1,r) = 1 \end{cases}$$

- **Linear congestion model**

$$\Delta Cong_k = \left(\sum_{i \in R_k, i \neq k} \frac{\Delta W_{k,i}}{Cap_k}\right) + \lambda_k \cdot \Delta y_k$$

$$= \sum_{i \in R_k} (\lambda_i \cdot \Delta y_i)$$



- Δy_i : the small MIM decap added to grid *i*
- $\Delta w_{k,i}$: the # of routes moved out of grid *i* to grid *k* caused by Δy_i
- Cap_i : the capacity of grid k

[Details in the paper]

Experimental Setup

- **90nm technology node**
- **6 metal layers for each 2D tier**
- **Supply voltage: 1.2 V**
- **Voltage drop threshold: 0.12 V (10%)**
- **3D Benchmarks**

Ckt	# Nodes	Worst V droop (V)	# nodes with noise violations	Violation Area S (V · ns)
Ibm123	18,634	0.135	3330	13.739
Ibm05	12,026	0.122	1359	72.260
ibm08	17,030	0.125	3191	41.305
ibm10	29,262	0.159	5935	91.286
ibm18	75,042	0.163	6392	108.649

Experimental Results

- **Comparison of three optimization strategies**

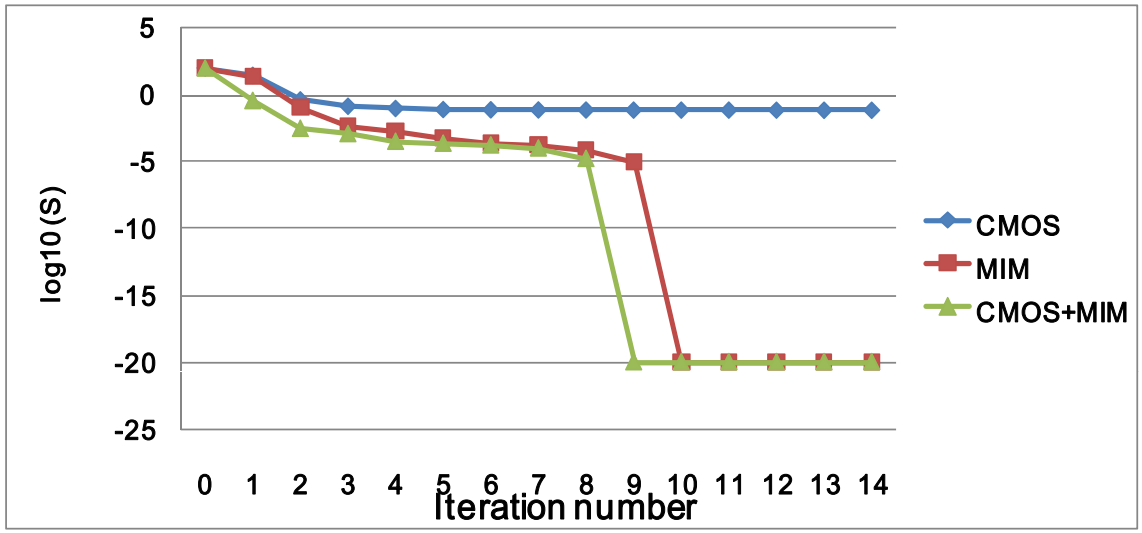
Ckt	CMOS only				MIM only			CMOS + MIM			
	VNs	S (V · ns)	Lkg (mA)	Decap (pF)	maxC (%)	avgC (%)	Decap (pF)	Lkg (mA)	maxC (%)	avgC (%)	Decap (pF)
Ibm123	368	0.023	2.1	564	15.8	3.9	607	1.1	8.4	1.7	628
Ibm05	24	0.049	2.7	480	19.7	1.7	550	2.1	0.0	1.2	546
ibm08	31	0.010	1.2	313	30.6	1.5	768	0.6	0.0	0.9	774
ibm10	351	0.182	1.6	417	10.6	5.9	511	0.9	4.5	2.5	520
ibm18	130	0.071	2.7	698	39.5	5.3	812	1.4	7.0	3.6	826

- VNs: number of violating nodes
- Lkg: leakage current
- maxC: maximum increment of congestion
- avgC: average increment of congestion

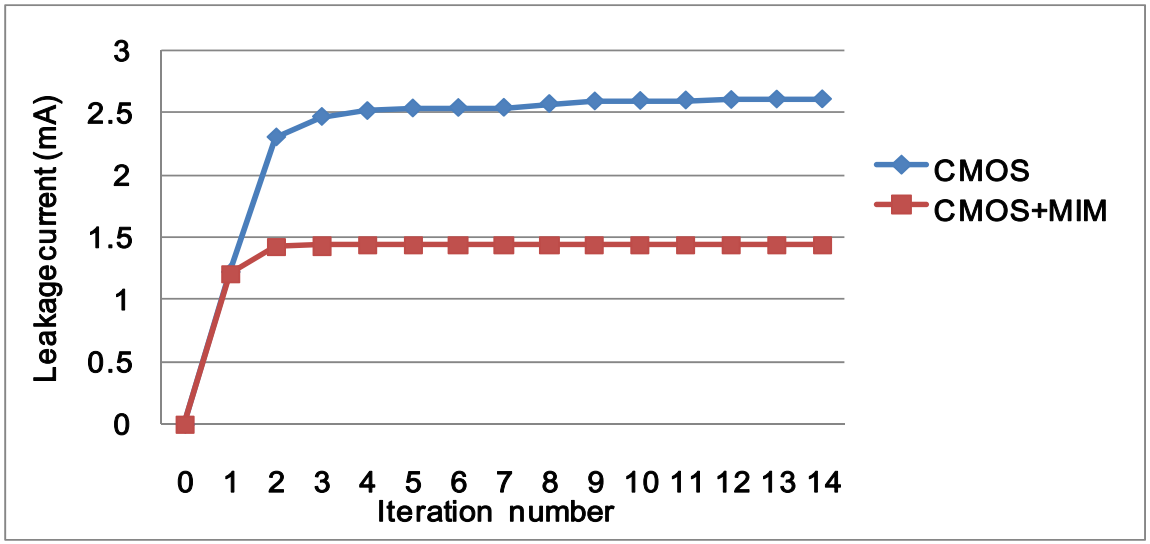
CMOS+MIM can achieve a good tradeoff between leakage power and routing congestion

Experimental results: ibm18 (cont.)

- Violation Area



- Leakage



Experimental Results (cont.)

- **Optimization results of power grid densities**

Cases	Power Grid Density	#Nodes	# nodes with noise violations	Worst V droop (V)	Violation Area S (V · ns)	Decap (pF)	Lkg (mA)	maxC (%)	avgC (%)	Time (s)
Case1	Normal	18634	3330	0.135	13.739	628	1.1	8.35	1.66	42.6
Case2	Denser	36433	4210	0.126	2.615	488	0.6	31.27	4.75	45.0
Case3	Densest	72114	4671	0.124	1.482	229	0.3	58.41	7.62	53.1

- Lkg: leakage current
- maxC: maximum increment of congestion
- avgC: average increment of congestion

1. **Denser power grid $\Rightarrow$ smaller voltage droop**
2. **Denser power grid $\Rightarrow$ increased congestion**

Summary

- **Power delivery into a 3D chip is a critical problem for next-generation designs**
- **MIM decap is an efficient option for 3D power grid optimization**
- **A LP based decap allocation approach using both MIM and CMOS decaps**
- **Our algorithm can also be used to solve the 2D power grid optimization problem**

Thank You!