isQED

Extraction & Application of Sparse L^{-1}RC Models

Byron Krauter
IBM Microelectronics

krauter@us.ibm.com

isQED

Background

- IBM is developing packaging tools for
  - Chip-package power distribution analysis
  - Package signal integrity analysis
- Goal is provide complete package electrical analysis in shortest TAT possible
- Modeling approach is linear simulation using sparse 3D lumped circuit models
  - Sparse L^{-1}
  - Distributed BEM capacitance (which is similar to L^{-1})
- Used internally today

Other Team Members

- Michael Beattie
- Hui Zheng
- Anirudh Devgan
- Sani Nassif

Outline

- Chip-Package Design Issues
  - Power distribution
  - I/O timing & Signal Integrity
- Package Modeling Issues & Requirements
- Sparse 3D Circuit Modeling
  - Sparse inductance models
  - Distributed BEM capacitance extraction
  - Package analysis components
- Package Analysis Applications Using Sparse 3D Models
  - Signal integrity analysis
  - Full package loop inductance analysis
  - Package power distribution models
  - Package decoupling capacitor optimization

isQED

Chip-Package Design Issues

isQED

Chip-Package Design Issues

- Power Distribution
  - DC losses increasing with scaling
    - Now causing gradients at chip/package interface
    - Total power increases when low spots limit chip performance
  - AC losses increasing with clock gating
    - Impacts both on-chip & off-chip performance
- I/O Timing & Signal Integrity
  - Faster clocks reduce skew tolerance for synchronous interfaces
  - Reduced bit times require complete path analysis
    - Frequency dependent losses
    - Return path discontinuities
  - Better chip-to-board models are needed but accurate, simple lumped models are too difficult to synthesize

Why DC Losses Are Increasing

- Power densities constant or growing
- At lower supply voltages -> higher currents
- With modest impedance improvements

$$V_{noise} = Z * I = Z * P / V_{supply}$$

$$V_{noise}/V_{supply} = Z * P / (V_{supply})^2$$

- Hot spots producing gradients at package level
  - Power & performance issues due to single point voltage sense & regulation

Are AC Losses Increasing?

- Clock gating necessary to limit power
- Gating produces larger power steps, ΔP, at lower voltages
- AC noise increases if Z doesn't also reduce

$$V_{noise} = Z * \Delta I = Z * \Delta P / V_{supply}$$

$$V_{noise}/V_{supply} = Z * \Delta P / (V_{supply})^2$$

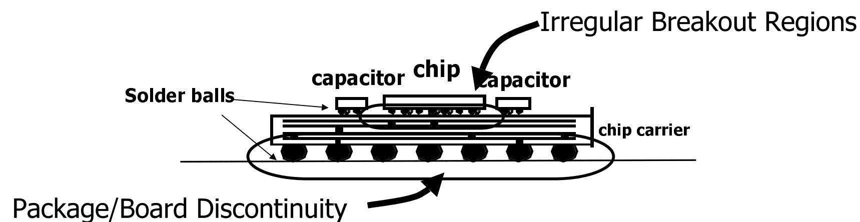
- DC losses however are mitigating AC noise
  - Constant background power damps AC noise
  - Non-linear subthreshold leakage damps better!

Package Modeling Issues & Requirements

ISQED 2005 - Tutorial II March 21, 2005

Package Role & Modeling Consequences

- Most packages are chip carriers & space transformers
- Emphasis on cost & signal count makes package regions irregular & 2D models inaccurate
- Conventional 3D inductance models are too large & dense for traditional simulators to handle



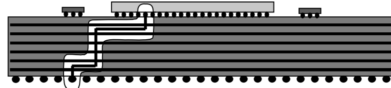
- Return paths widen at chip & board interfaces & are sometimes discontinuous inside package

Power Distribution Modeling Requirements

- Global accuracy is paramount
  - Because step response analysis requires a complete chip & package plus some board approximation
- But local accuracy is needed too
  - Because local gating depletes local regions when local chip decoupling isn't large enough
    - Decoupling capacitors have an effective range
  - Local DC power damps local transient noise
    - Need model detail down to the C4 level

Signal Integrity Modeling Requirements

- Signal integrity analysis (if the part has any chance of working) is a localized problem



- But 3D models are needed because package & board interfaces & package irregularities make 2D models inaccurate
- Frequency (or time) domain scattering models enable efficient system timing & noise analysis

Time Domain Scattering Parameters

- Given by the inverse Fourier transform of frequency domain scattering (S) parameters
- Also obtained by applying an impulse to DUT

- Transient simulators don't use S(f) parameters directly
  - Convert to rational approximations to use with recursive convolution methods
  - Or finite impulse responses (i.e. time domain scattering parameters) to use with direct convolution methods
- IBM's FastLine uses time domain scattering parameters
  - Scattering parameters derived applying triangular pulse to DUT

Sparse 3D Circuit Modeling

Overview

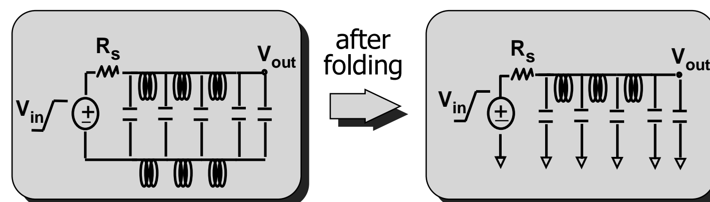
- Potentials & circuit solutions
- Sparse inductance models
- Distributed BEM capacitance extraction
- Package analysis components
  - scan-line shapes processing
  - windowed $RL^{-1}C$ extraction
  - linear simulation for final analysis or MOR

Potentials & Circuit Solutions

- Circuit operation depends on local potential differences in approximately charge & flux neutral local regions

$$\int \rho \, dv \approx 0 \qquad \int j \, ds \approx 0$$

- Equivalent circuits often mask this fact



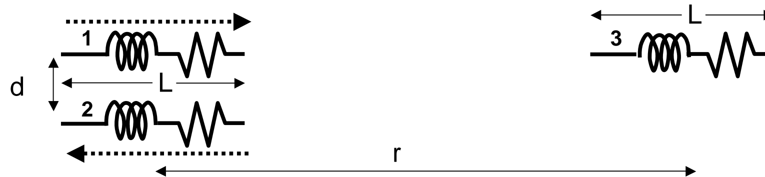
Sparse Inductance Models

Sparse Inductance Models

- L doesn't matter, but L-M does
  - Circuit analysis is concerned with local potential differences (e.g. L-M)
  - Local regions are approximately charge & current flux neutral making far-field mutual couplings negligible
- Sparse L models preserve L-M locally
  - For example (L-constant) - (M-constant) = L- M
- Sparse L models discard far-field couplings

Sparse Inductance Models

- Nearby return currents create approximately canceling inductive couplings

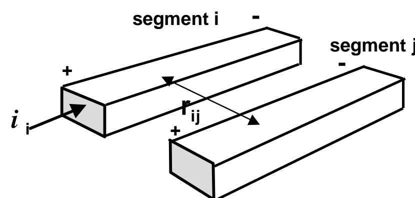


$$M_{13} \approx M_{23} \approx \frac{\mu_O}{4\pi} \frac{L^2}{r}$$

$$M_{13} - M_{23} \approx \frac{\mu_O}{4\pi} \frac{L^2}{r} \left(\frac{d^2}{r^2} \right)$$

Sparse Inductance Models

- 3D partial inductances are derived using the 3D Green's function for a point source



$$L_{ij} = \frac{\mu_0}{4\pi} \iint \frac{\vec{dl_i} \cdot \vec{dl_j}}{r_{ij}}$$

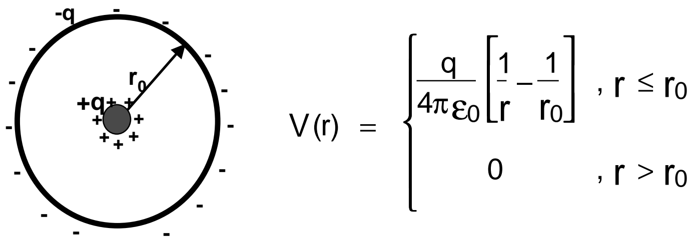
- Partial inductances are "monopole" in nature and decay very slowly (logarithmically for long wires)
- Sparse inductance models are "dipole" in nature and decay much more rapidly

Sparse Inductance Models

- Equipotential shells (e.g. shift & truncate)
  - User knowledge required (approximate loop size)
  - User defined return radius determines L & M shifts
  - Used just like partial inductances
- Truncated L^{-1}
  - User knowledge still required (approximate loop size)
  - User defined truncation window (like return radius above) determines L & M shifts
  - Most efficient sparse method in simulation (when used with nodal analysis)
  - L^{-1} circuit element not currently handled in most Spice
  - Similar to capacitance, but different
    - Positive off-diagonals due to diagonal wiring & KCL
    - Complex models are not diagonally dominant

Background: Spherical Capacitor

- Consider two concentric spheres with equal and opposite charges



$$V(r) = \begin{cases} \dfrac{q}{4\pi\varepsilon_0}\left[\dfrac{1}{r} - \dfrac{1}{r_0}\right] & , r \le r_0 \\ 0 & , r > r_0 \end{cases}$$

- Potential differences inside the capacitor are unchanged by the outer shell radius

Sparse L: Shift & Truncate

- Compute partial inductances using "spherical capacitor" like magnetic vector potentials

$$\vec{A}(r) = \begin{cases} \frac{\vec{i}\mu_0}{4\pi}\left[\frac{1}{r}-\frac{1}{r_0}\right] & , r \le r_0 \\ 0 & , r > r_0 \end{cases}$$

- Yields sparse (shift & truncate) approximation of original matrix

$$\mathcal{L}_{ij} = L_{ij} - \frac{\mu_0 \vec{l}_i \cdot \vec{l}_j}{4\pi\, r_0} \quad , r_{ij} < r_0$$

$$\mathcal{L}_{ij} = 0 \quad , r_{ij} > r_0$$

Sparse L^{-1}

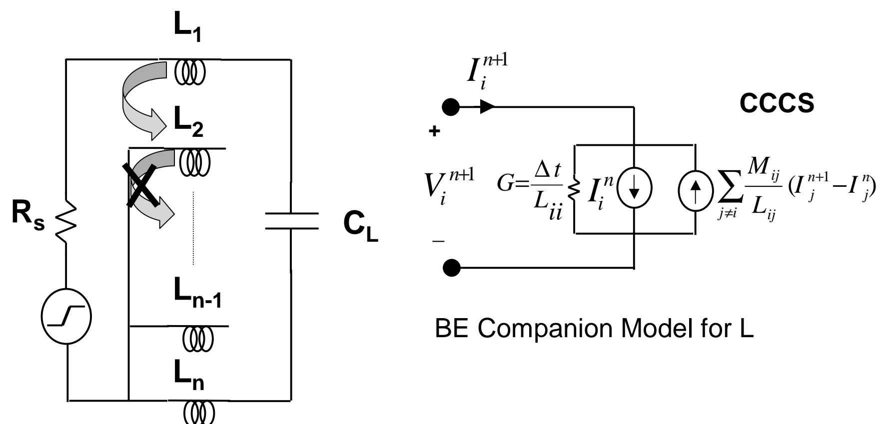
- Introduced by Devgan *et al*
  - Partial inductance matrices are inverted & truncated
- Many behaviors are similar to capacitance
  - Coupling decays rapidly due to shielding
  - Floating conductors matter (unlike L & P formulations)
- But directionality & KCL make it different
  - Diagonal wiring naturally creates positive off diagonals
  - L^{-1} is not diagonally dominant
- Shifts L & M, but L-M is locally accurate
- Retains more magnetic coupling than sparse L
- Circuit equations are symmetric with L^{-1}RC & Norton sources

Sparse L vs. Sparse L^{-1} Models

- The two methods behave similarly
  - Inverse of truncated L^{-1} looks similar to shift & truncate L
- But truncated L^{-1} provides an indirect coupling mechanism that's missing in sparse L models
  - Inverse of truncated L^{-1} never goes to zero
  - Coupling decays exponentially outside window
- For the same sparsity, this indirect coupling makes L^{-1} more accurate

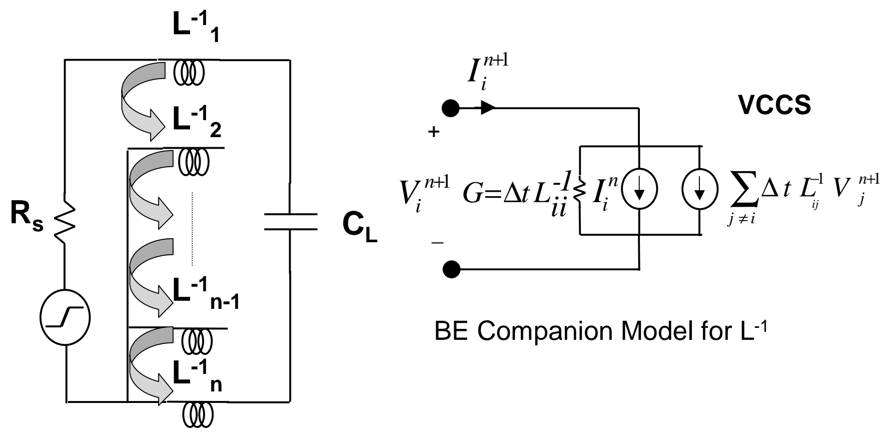
Direct Coupling Only in L Models

- What if return path is not within shift-truncate window

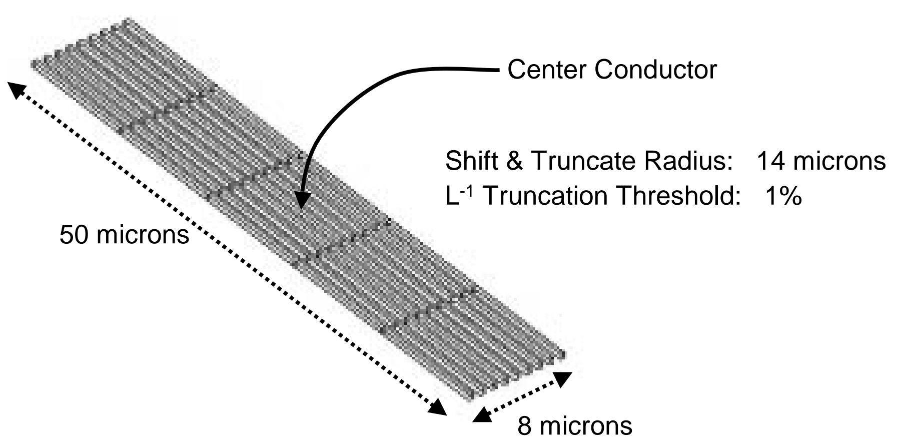


Some Indirect Coupling in L^{-1} Models

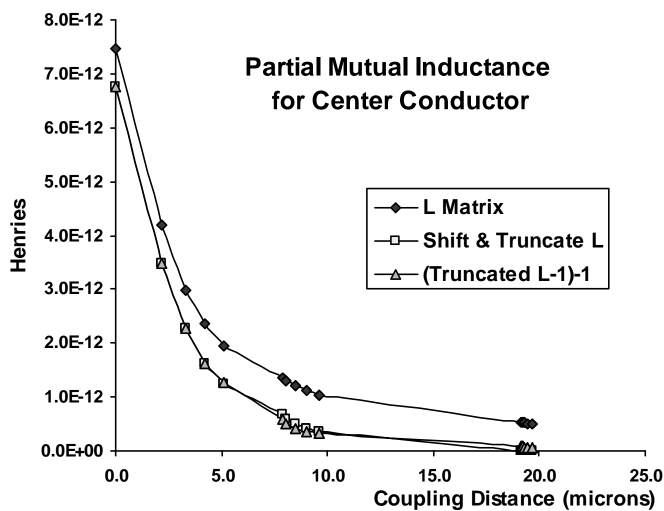
- What if return path is not within L^{-1} extraction window



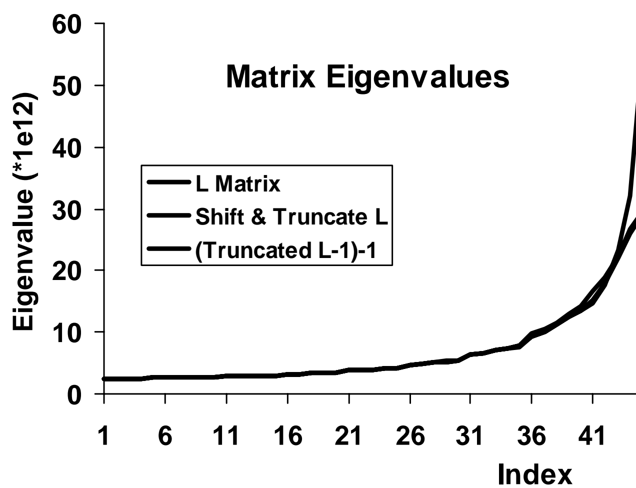
L vs Sparse L vs (Sparse L^{-1})$^{-1}$



L vs Sparse L vs (Sparse L^{-1})$^{-1}$



L vs Sparse L vs (Sparse L^{-1})$^{-1}$



Sparsity vs Accuracy

- Matrix densities for similar accuracy are different
- In last example
  - Shift & Truncate L
    - 44 non-zero coupling terms for center conductor
    - 24% sparse for entire 45x45 matrix
  - Truncated L-1
    - 10 non-zero coupling terms for center conductor
    - 79% sparse for entire 45x45 matrix
- In larger problems the density advantage is about 2:1
  - For example, a 2% dense shift and truncate L matrix would have the same accuracy as a 1% dense truncated L^{-1} matrix.

Simulation Efficiency

- Simulation reduces to linear nodal analysis when using only L^{-1}RC & Norton sources
- Nodal analysis circuit equations are symmetric positive definite (s.p.d)
- s.p.d equations can be solved via more efficient Cholesky factorization
  - No pivoting is required for Cholesky factorization
  - Powerful pre-factorization ordering algorithms can significantly reduce potential fill-ins
  - Fewer fill-ins reduce computation time and memory requirements

L based Transient Simulation

- Differential equations

$$\begin{bmatrix} G & A_L \\ -A_L^T & 0 \end{bmatrix} \begin{bmatrix} V_n \\ I_L \end{bmatrix} + \begin{bmatrix} C & 0 \\ 0 & L \end{bmatrix} \begin{bmatrix} \dot{V}_n \\ \dot{I}_L \end{bmatrix} = \begin{bmatrix} I_C \\ 0 \end{bmatrix}$$

- Inner loop equations for Backward Euler integration

$$\begin{bmatrix} G + \frac{C}{\Delta t} & A_L \\ -A_L^T & \frac{L}{\Delta t} \end{bmatrix} \begin{bmatrix} V_n(t+\Delta t) \\ I_L(t+\Delta t) \end{bmatrix} = RHS$$

L^{-1} based Transient Simulation

- Inner-loop Linear System (BE)

$$[G + C/\Delta t + A_L \; L^{-1} \cdot \Delta t \; A_L^T] \, [V_n(t+\Delta t)] = RHS$$

- Clock tree/mesh with 1.8 million elements (no coupling)

	L, Thevenin	L-inverse, Norton with Cholesky	L-inverse, Norton without Cholesky	L-inverse, Thevenin
Memory [GByte]	1.82	1.10	1.34	1.54
Sim. Time [min]	68	34	57	54
Equations	1,202,782	929,115	929,115	929,116
Non zeroes	3,982,287	2,887,621	2,887,621	2,887,623

Ongoing Issues with Truncated L^{-1}

- Return paths need to be included
  - Truncation must be based on return distances & not percentages
  - Dual windows needed to minimize fringe couplings to infinity
    - Compute L^{-1} in larger window
    - Truncate to a smaller window
- Floating L^{-1} elements cannot be discarded
  - Conductors present in extraction should be present in simulation
- L^{-1} models are not diagonally dominant for complex structures

Distributed BEM Capacitance Extraction

Distributed BEM Capacitance Extraction

- Similar to L^{-1} extraction
  - L^{-1} ignores KCL on a net basis & distributed BEM capacitance extraction considers non-physical surface potentials
- Produces spatially distributed coupling models
  - segment-to-segment couplings
  - no distribution heuristics

Not Like Regular BEM Cap. Extraction

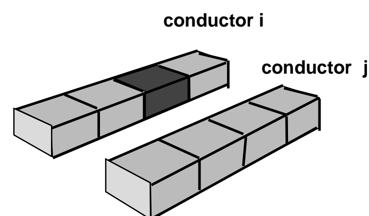
- Discontinuous surface potentials (ground all but one segment)
- Locally solve for each segment

$$C = A^T (P)^{-1} A$$

- Yields spatially distributed C model with self coupling terms

→ 0 Volts

→ 1 Volt



Package Analysis Components

Shapes Processing

- Initial
  - add chip & board fixtures
  - merge & eliminate overlapping shapes
  - discretize lines & planes
  - create node numbers
- Extraction Specific
  - create overlapping blocks for parallel extraction
  - create neighborhood lists
- Analysis Specific
  - create signal channels for signal integrity analysis
  - create block regions for power distribution model creation

Extraction

- Parallelized full package L^{-1} & C extraction
  - Windowed & scan line based
  - Dual windows to minimize fringe couplings to infinity
    - Larger extraction window
    - Smaller truncation window
  - L & P hash tables
  - L^{-1} & C for N blocks stored in 2N binary files

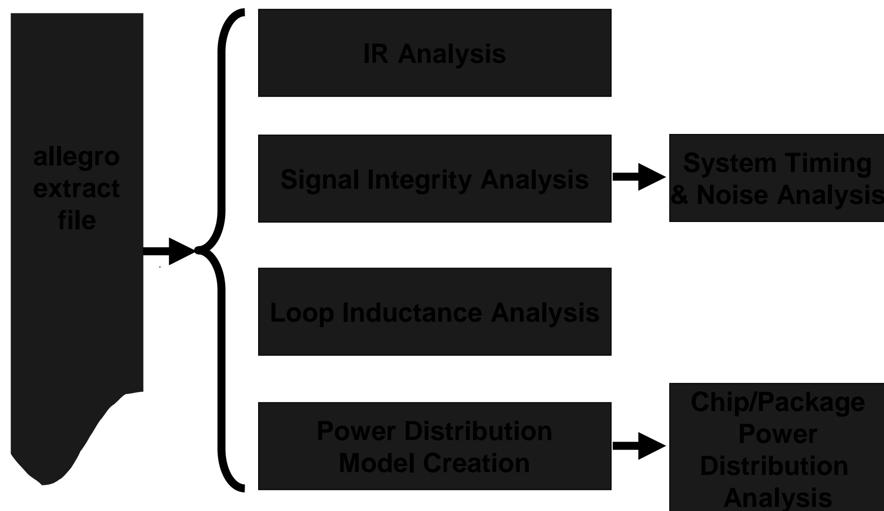
Netlisting & Simulation

- Parallelized netlist generation
  - Create application specific netlists
    - using L^{-1} & C values from binary files
    - application specific shape regions (e.g. channel or block regions)
  - Append boundary & simulation conditions
- Linear simulation
  - Fixed time step transient analysis
  - Sparse Cholesky solver

Package Analysis Applications

ISQED 2005 - Tutorial II March 21, 2005

Applications of shapes->extract->linear simulation



ISQED 2005 - Tutorial II 43 March 21, 2005

isQED

Signal Integrity Analysis

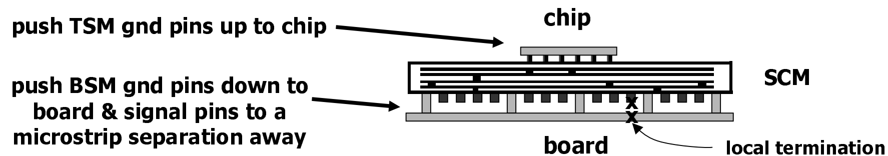
ISQED 2005 - Tutorial II

March 21, 2005

isQED

Signal Integrity Analysis Flow

- Common all power supplies & add "chip & board planes" to allegro extract file

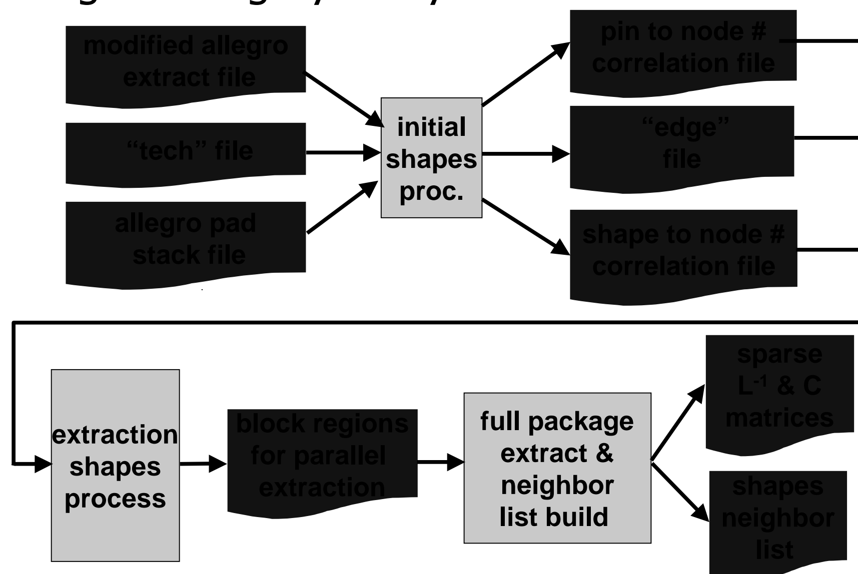


- Initial Shapes Processing
  - merge & eliminate overlapping shapes
  - discretize lines & planes
  - create node numbers

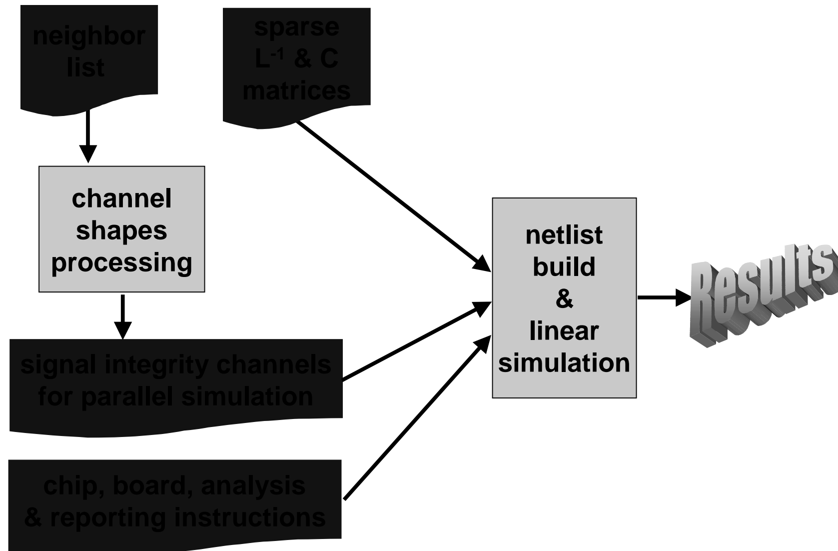
Signal Integrity Analysis Flow

- Divide package into N block regions for extraction
- Parallelized L^{-1} & C extraction & neighbor list creation
  - extract L^{-1} & C for N blocks & store in 2N binary files
  - store neighbor lists in N binary files
- Signal integrity channel formation
  - create shape-only extraction blocks & simulation channels independent of extraction parameters
- Parallelized netlist build & linear simulation
  - assemble L^{-1} & C values for each signal integrity channel
  - append boundary & simulation conditions
  - simulate & process results

Signal Integrity Analysis Flow

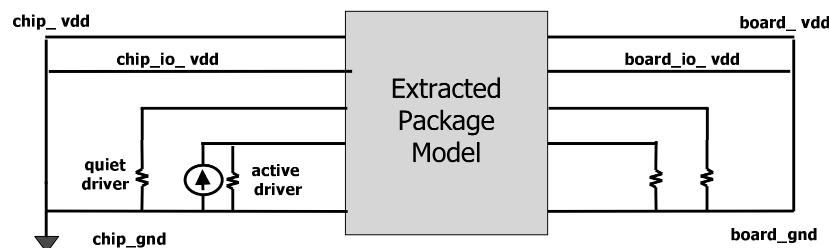


Signal Integrity Analysis Flow

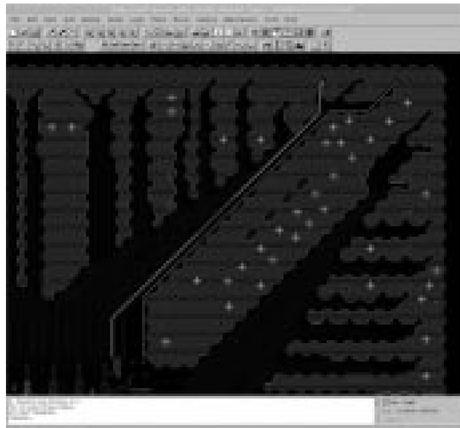


Signal Integrity Boundary Conditions

- Linear analysis with 50 Ω Norton driver models
- All chip supplies shorted & all board supplies shorted
- Near and far end termination with Zo=50 Ω
- Single point ground model => local differential measurements
- Apply
  - Saturated ramp to find far end and saturated near end noise
  - Triangular pulse to create time domain scattering models



Signal Integrity Channel Example

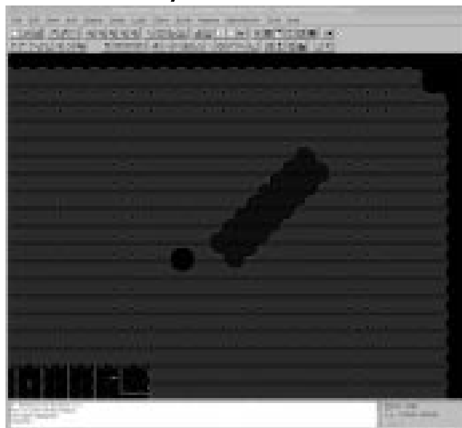


Allegro Layout

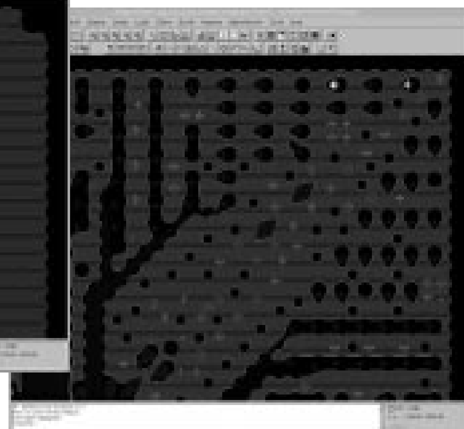
Victim Net Highlighted

Signal Integrity Channel Example

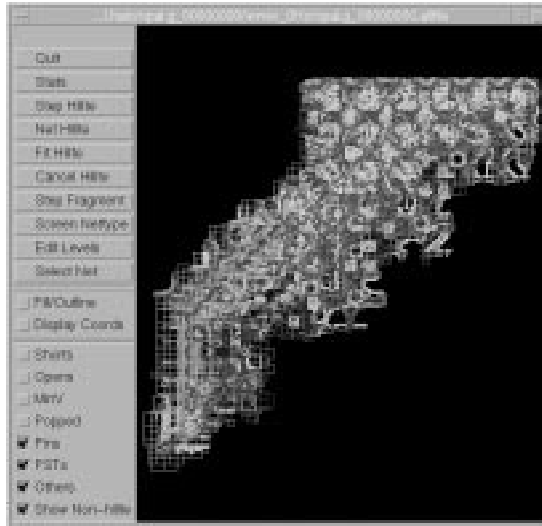
Layer Above



Layer Below

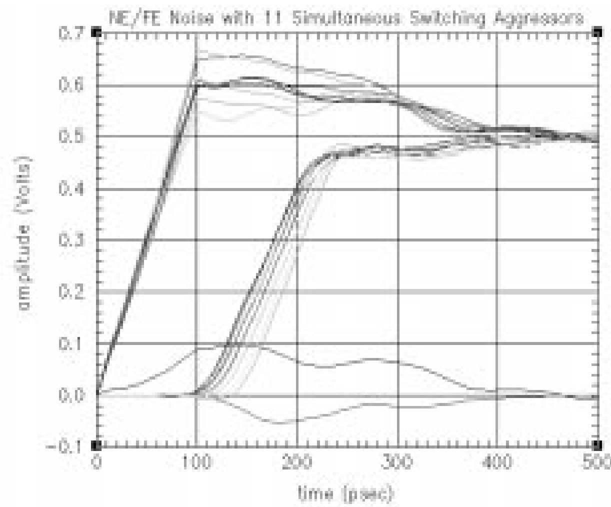


Signal Integrity Channel Example



Channel Shapes
All Levels Displayed

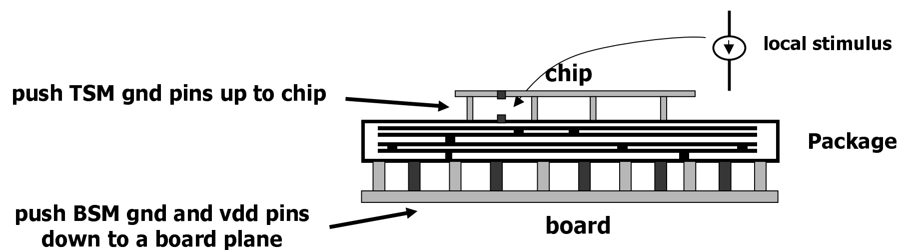
Far & Near End Noise Example



Full Package Loop Inductance Analysis

Analysis Setup

- Add "chip & board planes"
  - More realistic current return paths
  - Natural averaging effect for multiple return paths
- VDD-centric analysis: short all top gnd pins to the chip plane (below) and local stimuli for vdd pins
- GND-centric analysis: short all top vdd pins to the chip plane and local stimuli for gnd pins



3s3p3s Package Example Statistics

- Package Size : 28mm X 28mm
- Chip Size : 7.5mm X 7.5mm
- Layers : 10
- Top Pins : 102 VDD + 155 GND
- Bottom Pins : 24 VDD + 64 GND
- # of shapes : ~ 274,000
- R extraction time : 156 sec
- L extraction time : 24 min on 9 CPUs

Loop resistance calculation

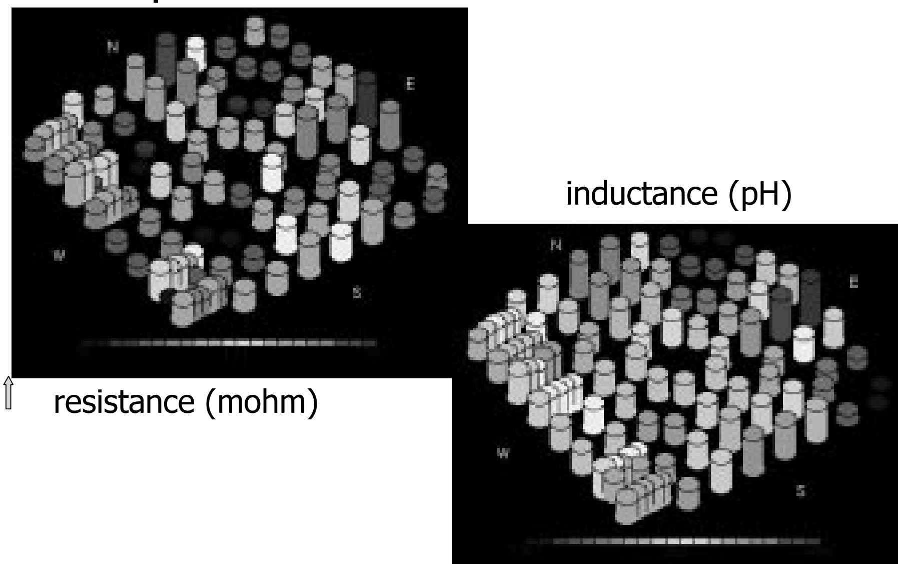
- Do a resistance-only extraction on the package
- Use a 1-amp current source as the local stimulus for each top VDD (VDD-centric) or GND (GND-centric) pin
- Do a DC analysis and the voltage across each top termination is the average resistance.

Loop inductance calculation

- Do a L^{-1} only extraction on the package
- Use a 1 amp/second current source as the local stimulus for each top VDD (VDD-centric) or GND (GND-centric) pin
- Do a DC analysis and the voltage across each top termination is the average loop inductance per C4.

$$[A_L \; L^{-1} \cdot \Delta t \; A_L^T]\,[V_n] = \Delta i_n$$

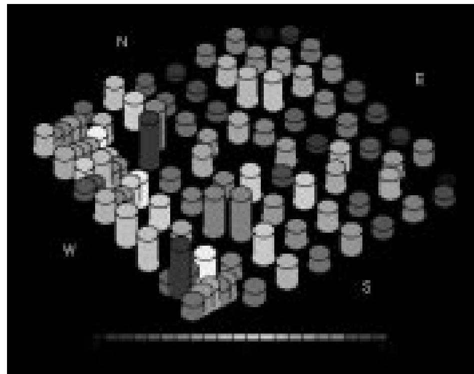
Vdd profiles



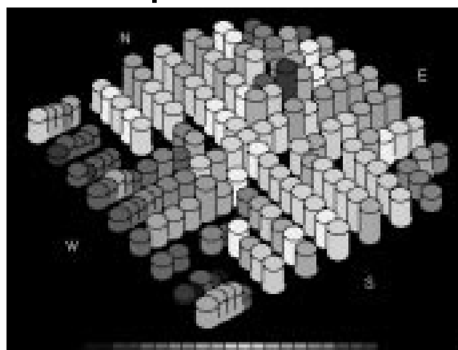
L/R profile for VDD

- However, L profile is not simply a scaled version of R profile

L/R
(pH/mohm)

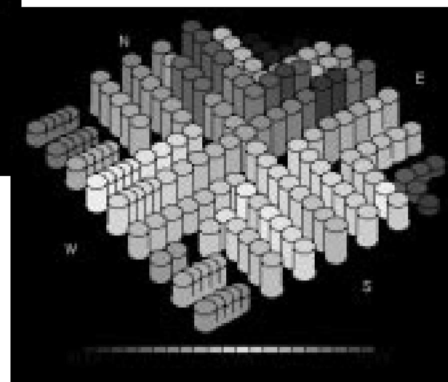


Gnd profiles



resistance (mohm)

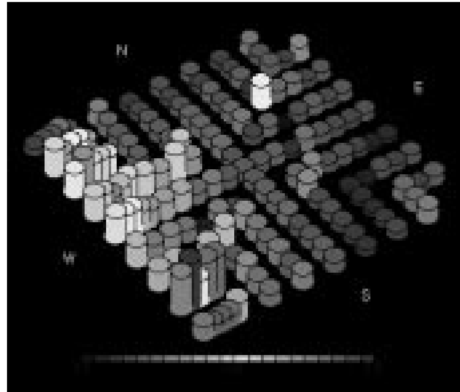
inductance (pH)



L/R profile for GND

- However, L profile is not simply a scaled version of R profile

L/R
(pH/mohm)

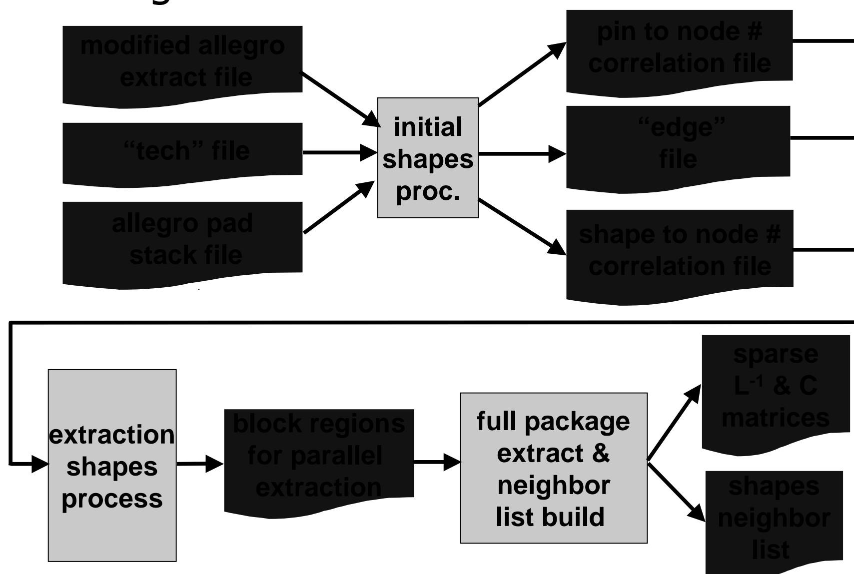


Package Power Distribution Models

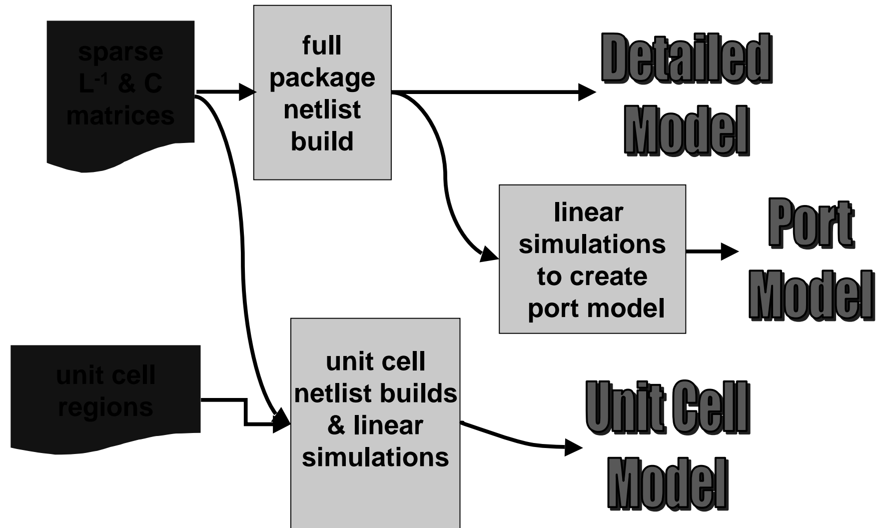
Power Distribution Package Models

- Detailed model
  - Sparse L^{-1}RC model for all power/ground shapes
  - Basis for simplified package models
- Port model
  - Small but dense model
- Unit cell circuit model
  - Subdivide package into cells
  - Solve for x,y,& z cell impedances
  - Reassemble cell subcircuits
  - Small & sparse model

Package Model Creation



Package Model Creation

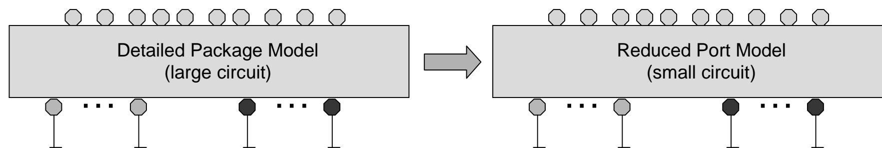


Detailed Model Extraction Statistics

Design	Case 1	Case 2
Package Size	26 mm x 26 mm	34 mm x 34 mm
Chip Size	12 mm x 12 mm	15 mm x 15 mm
Layers	12 (3+6+3)	12 (3+6+3)
Top Pins	168 VDD + 148 VD2 + 313 GND	255 VDD + 184 VDD2 + 72 VDD3 + 38 VDD9 + 508 GND
Bottom Pins	28 VDD + 55 VD2 + 80 GND	60 VDD + 53 VDD2 + 20 VDD3 + 10 VDD9 + 240 GND
Shapes	~ 400,000	~ 800,000
Nodes	~ 200,000	~ 370,000
RL extraction time	2 h 35 min (9 parallel)	2 h 22 min (16 parallel)
Extracted Model	R: 400 K self: 234 K mutual: 1.9 M	R: 800 K self: 444 K mutual: 4.3 M

Need for Model Reduction

- Don't care about the responses on the internal nodes (~ 100K's) in the detailed package model
- Need to find a reduced port model that approximates the characteristics at the top/bottom pins (~ 100's)



Simulation Performance Comparison

Design	Package Type	Memory (GB)	CPU Time
Case 1	Detailed	7.6	1h 26min
	Port Model	1.2	16 min
Case 2	Detailed	15.6	3h 28min
	Port Model	2.5	33 min

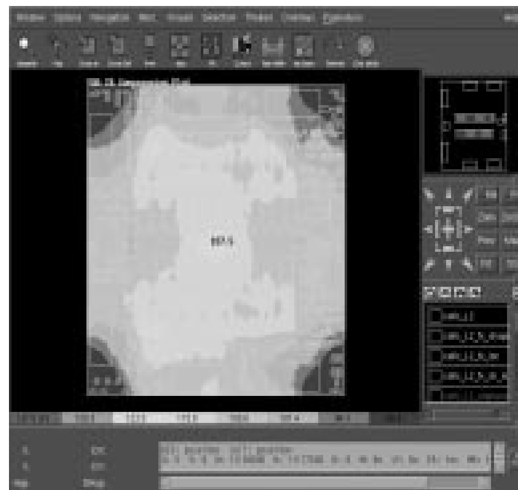
Simulation time: 50 ns (time step: 50 ps)

Number of time steps: 1000

Worst Voltage Drop Comparison

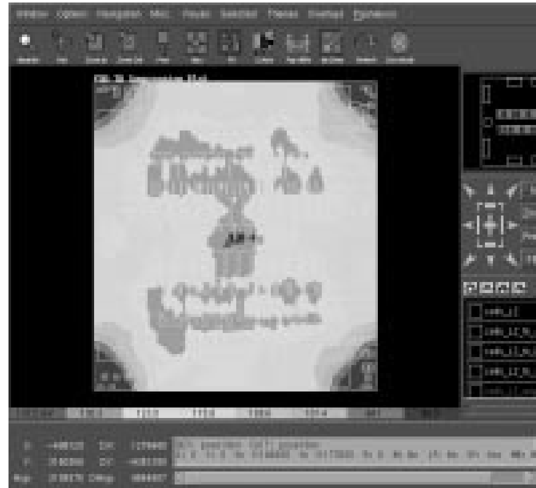
Design	Package Type	Max (mv)	Min (mv)
Case 1	Extracted	127.5	63.2
	Port Model	138.4	69.5
Case 2	Extracted	281.0	153.1
	Port Model	288.4	148.9

Detailed Model (Case 1)



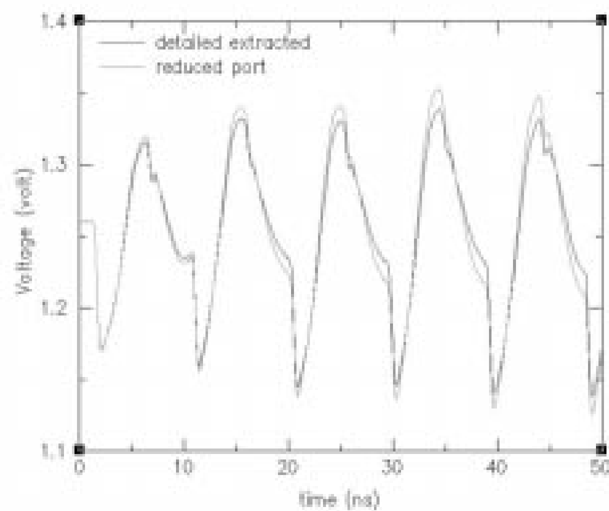
Max: 127.5 mV (5350000, 5450000)
Min: 63.2 mV (550000, 150000)

Port Model (Case 1)



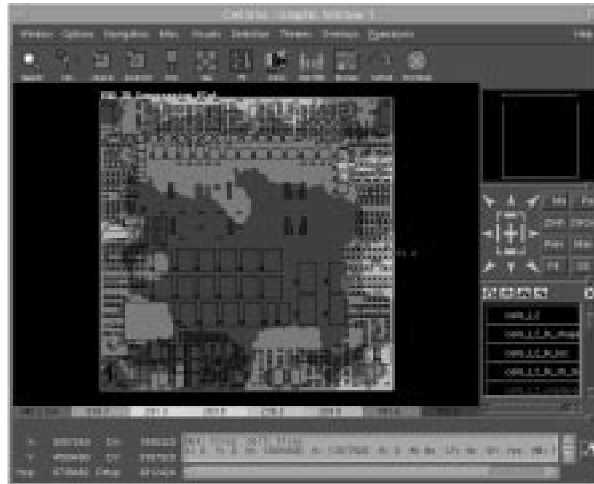
Max: 138.4 mV (5350000, 5450000)
Min: 69.5mV (550000, 150000)

Transient Vdd-Gnd (Case 1)



isQED

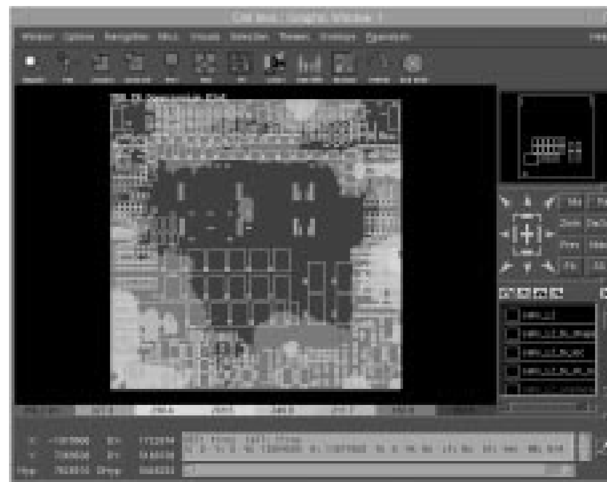
Detailed Model (Case 2)



Max: 281.0 mV (13750000, 6350000)
Min: 153.1 mV (10550000, 7350000)

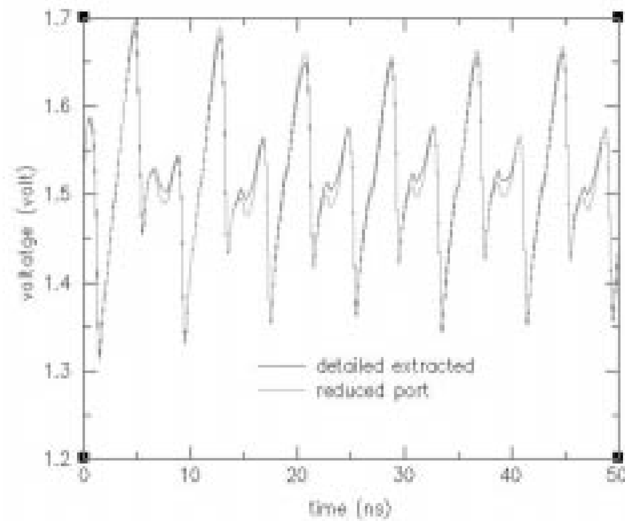
isQED

Port Model (Case 2)



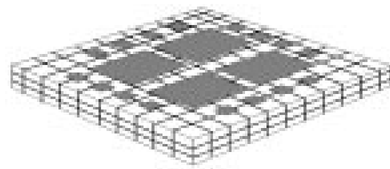
Max: 288.4 mV (13750000, 6350000)
Min: 148.9 mV (9650000, 7450000)

Transient Vdd-Gnd (Case 2)



Building Unit Cell Models

- Subdivide package into unit cells



- Compute x,y & z impedances across every cell

- Reassemble with simpler cell models

isQED

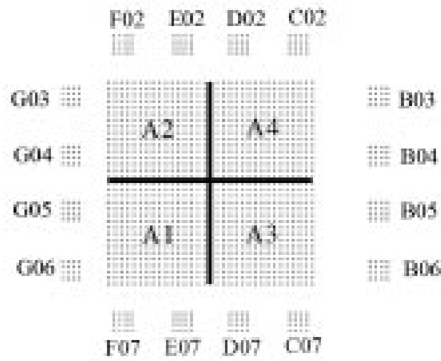
Package Decoupling Capacitor Optimization

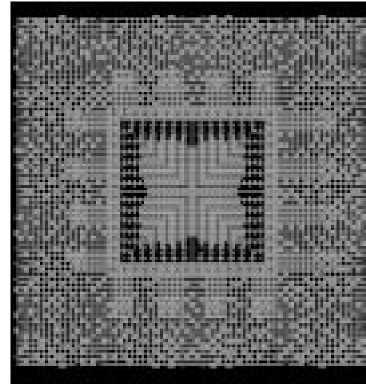
isQED

Package Decoupling Capacitor Optimization*

- Based on simulated annealing on a reduced order package model (e.g. Prima adapted for $L^{-1}RC$)
- Results:
  - Mixing package capacitor types helps minimize impedance by mixing time constants
    - Reduces resonant impedance
  - Mixing package capacitor types helps minimize cost
    - Cheaper capacitors have larger RC time constants
  - Non-uniform switching should be decoupled with non-uniform capacitor placement & usage

* H. Zheng, B. Krauter, L. Pileggi, On-Package Decoupling Optimization with Package Macromodels, Int'l Custom Integrated Circuits Conference, Sept. 2003.





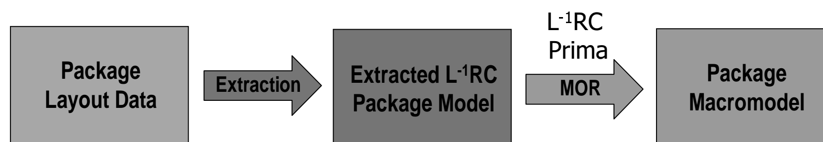
Chip-package interface (top layer)

Vdd/Gnd mesh (middle layers)

Ceramic Package for Optimization Study

Extraction & MOR Statistics

- Size: 30.6mm X 30.6mm
- # of layers: 14
- # of decaps: 16
- # of segments (VDD and GND): 76,400
- # of mutuals: 537,000 (full system: 2.9e9)
- Extraction runtime: 1 hour 40 minutes (typical number of segments in a local extraction: 150)
- MOR runtime: 664 seconds (projected into 105x105 system: 21 ports, order of 5 for each port)

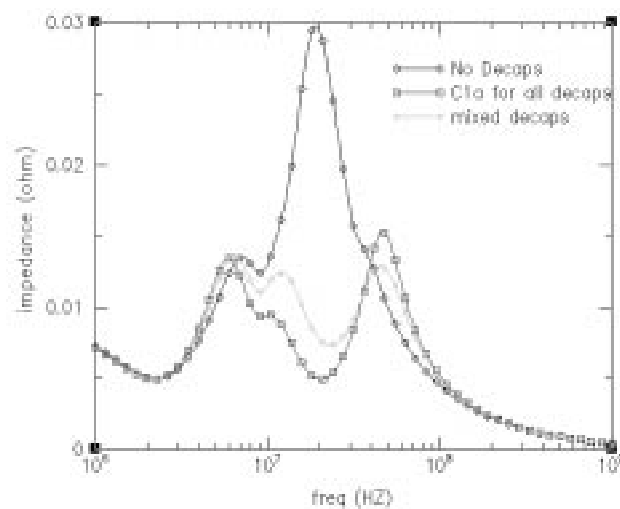


Capacitor Parasitics & Costs

- A capacitor with large ESC, low ESR & low ESL is more expensive

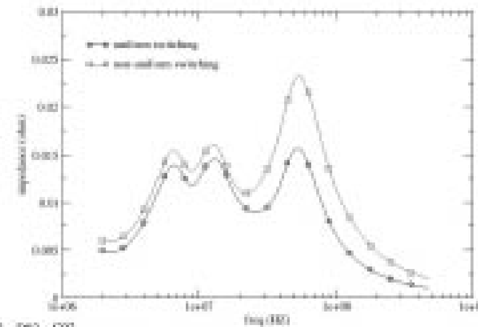
	C1d	C1c	C1b	C1a	No C
ESC (nf)	50	100	50	100	--
ESR (mohm)	60	60	30	30	--
ESL (pH)	100	100	40	40	--
Relative Cost	1	2	2	4	0

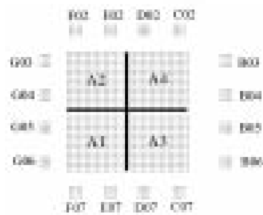
Mixing Decoupling Capacitors



- C1a for all decaps:
  - Cost: 64
  - Peak Z: 15.3 mohm
- Mixed decap choices:
  - Cost: 27
  - Peak Z: 13.7 mohm

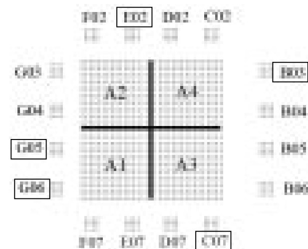
Non-uniform Switching





- Uniform switching:
  - $W_i = 0.25$
  - Peak Z: 15.7 mohm
- Non-uniform switching:
  - $W_1 = 0.5$, $W_2 = W_3 = 0.20$, $W_4 = 0.1$
  - Peak Z: 23.2 mohm

Decap Deployment



- Non-uniform switching:
  - $W_1 = 0.5$, $W_2 = W_3 = 0.20$, $W_4 = 0.1$

	C1d	C1c	C1b	C1a	No C	Peak Z (mohm)	Decap Cost
No Decaps					All	32.1	0
All C1d	All					23.2	16
All C1a				All		25.0	64
Optimized	B03, E02, G05, G06	C07			The rest	21.1	6