Interconnect-centric design and analysis for electrical integrity in systems-on-a-chip

Dennis Sylvester, Department of Electrical and Computer Science, University of Michigan, Ann Arbor, MI
Kenneth Shepard, Department of Electrical Engineering, Columbia University, New York, NY

November 9, 2000  ICCAD 2000
Outline

- Noise intro
- Noise methodology
- Clock distribution
- Interconnect analysis for noise
- Inductance
- Substrate noise, power supply noise
- Process variation
- SOI
- Measurement characterization
Noise intro
What is noise?

Physical noise sources

- Thermal noise. Open-circuit mean-square voltage is:
  \[ \langle V_n^2 \rangle = 4kTB \]

- Flicker or 1/f noise. In MOS devices, due to interface traps.
- Shot noise. For a pn junction, the mean-square noise current is:
  \[ \langle i_n^2 \rangle = 2qBI \]

These sources are random; that is, they cannot be predicted at any time even if the past values are known.
What is noise?

Man-made (or environmental) noise sources

- Are deterministic although a stochastic model may be appropriate in some cases.
- Can be orders of magnitude larger than physical noise sources.
- For CMOS digital circuits, noise is any deviation in the analog voltage from the nominal supply or ground rails when the node should represent a stable logic ‘0’ or ‘1’ value.
Where does noise come from?

- A digital IC is a very noisy place
  - Signals switching from Gnd->Vdd, Vdd->Gnd couple to other quiet or switching nets through unintended “couplings”
- Analog *always* vulnerable
- Noise immunity used to win out over noisiness for digital circuits
- What are these “couplings”?
Scaling Trends

Other issues:
• Increasing off currents
• Reduced Vdd/Vt

November 9, 2000
ICCAD 2000
Why are noise problems getting worse?

More noise-sensitive listeners

- Requirements for mixing analog and digital together on the same chip.
- Performance requirements are pushing designs toward the use of more “high-gain” digital circuits, which are more sensitive to noise.
Why are noise problems getting worse?

Louder talkers

- Interconnect scaling
  - More levels of interconnect packed closer together.
  - Metal thicknesses remain constant because of resistance
  - In the scaling form 1.8 um to 0.9 um pitch, line-to-line capacitance increases from 46% to 68% of full-loaded self-capacitance.

- Faster slew times
- Higher current density demands
Noise impact on functionality

Logic signal of interest has settled down when coupling interactions disturb the voltage representing the logic value.
Noise can make you go faster or slower. Noise effects are the same, just against switching rather than static signals.

At the same time consider simultaneous switching effects.
Mixed-signal noise issues

Channel Noise

Digital Noise Coupling

Device Noise

Phase Noise

RF Filter

LNA

IF Filter

ADC

Base-Band Processing

Digital Noise Coupling

Mixed Signal Wireless IC

November 9, 2000

ICCAD 2000
Effects of Substrate Noise on ADC

Simultaneous switching of the output buffers creates substrate noise that corrupts the A2D conversion.

8 bit Video A2D Converter

Expected response

ADC Codes

Samples

November 9, 2000

ICCAD 2000
Noise immunity of digital systems

- Analog voltages are used to define logic ‘0’ and ‘1’
- Noise causes these analog voltages to vary but the system will still function as long as the values fall into one of the valid logic levels.
- What constitutes a valid logic level depends on the node of the circuit and the time-domain characteristics of the noise.
Classification system

Noise is any deviation in the analog voltage from the nominal supply or ground rails when the node should represent a stable logic ‘0’ or ‘1’ value

- $V_H$ noise.
- $V_{H^*}$ noise.
- $V_L$ noise.
- $V_{L^*}$ noise.
Non-full-rail circuits

Both introduce noise to subsequent stages because of reduced logic-high voltage level.
Coupling through substrate: substrate noise

Substrate and n-wells capacitively coupled to device nodes across reverse-biased pn junctions.
Coupling through FETs: charge-sharing noise

Charge-sharing noise applies to skewed-evaluate circuits and is produced by charge redistribution with internal nodes of the circuit.
Coupling through FETs: Miller noise
Coupling through interconnect: crosstalk noise

Voltage induced on an evaluation node due to capacitive coupling to a switching node of another net.
Power supply noise can be categorized into two types:

- **DC IR drop.** Variations in the DC power supply levels on chip due to IR drops in the power/ground distribution.
- **Delta-I noise.** The simultaneous switching of off-chip drivers and internal circuits causes periodic variations in the supply and ground rails on chip due to inductance on the chip and package supply and ground wires.
Power supply noise

The following figures show the power and ground variations measured on-chip for two different clock frequencies for a custom CMOS microprocessor.
Stray minority carrier interaction

- Minority carrier back-injection into the substrate. This is the same current source as CMOS latch-up. This is becoming much less of a problem at scaled supply voltages because of the need to bootstrap more than 0.6 V.
- Ionizing radiation
  - Cosmic rays
  - Alpha particles produced by radioactive decay of lead in C4 package technology
Key point

- As circuit designs (or device designers), we focus primarily on the devices.
- Noise issues force us to really put the focus on the wires, the substrate, and the entire electrical environment of the chip (package, etc) in addition to the devices.
Noise methodology
Noise Themes

- Ad hoc rules of thumb being supplemented with increasingly sophisticated electrical analysis
- Board-level issues being brought onto the chip, where things are much more complicated.
- Quiet the talker or deafen the listener.
- For digital circuits, we’re considering “analog” noise affecting digital operation abstraction.
Design methodology for noise for digital circuits

- Dynamic simulation is not possible on a large-scale
- One approach is to have design rules which attempt to limit the amount of coupling interaction or improve the noise immunity of the listener
  - very hard to generalize to different circuit topologies and different technologies
  - tend to worst-case things a lot
  - fundamentally disregards the trade-off between noise margin and performance (tuning of widths, lengths, and Vts)
  - never guaranteed to catch everything!
Common “circuit-checking” rules

• Circuit design rules
  – Disallow single NFET or PFET pass gates because of the $V_t$ voltage drop.
  – Disallow pass gates at the ends of long wires.
  – Disallow long wire runs feeding domino gate inputs
  – Disallow high beta static circuits feeding low beta static circuits, or low beta static circuits feeding high beta static circuits
  – Routing rules or driver strength rules
  – Rules on the use of low-$V_t$ devices
  – Use of differential circuits to mitigate common-mode coupling and power-supply noise
  – Requiring use of babysit devices on internal nodes
An example

Sizing requirement: Half-latches must be used for dynamic gates and be sized to 1/10 of the tree strength.

How do you know how much noise an evaluation node can tolerate?
Mixed-signal designers have tried the same rules approach

• Apply differential operation everywhere (build immunity to common-mode noise)
• Distribute digital signals and clocks in complementary form if possible
• Critical operations should be performed after clock transitions (after the substrate settles)
• Minimize wire-bond inductance
• Use PMOS differential inputs in op amps since the well can be tied to their common source
• Use guard rings (nwell or substrate ties)
Guard rings can be bad

- For example if $Z_s < Z_{pd}$ adding guard rings increases noise at the analog circuit
Determining noise immunity of digital listeners

- Because noise was not traditionally a concern for digital integrated circuits, noise immunity analysis has not (until recently) been given the attention it deserves
- Noise margin metrics must consider noise within the bandwidth of the receiving circuits (AC noise margins)
Essential stability requirement

**Principle**: Functional failures due to noise are fundamentally due to a latch falsely changing state. Latches can be either static, bistable feedback configurations of restoring logic gates or dynamic nodes, which are acting as latches, storing state by virtue of the charge on an evaluation node.
Essential stability requirement

Noise not large enough to cause disruption!
Essential stability requirement

Noise large enough to cause disruption, but latch might not have time to recover before clock edge drops.

OK?
Essential stability requirement

Variant noise fail!
Essential stability requirement

First point that things went wrong!
Essential stability requirement

Invariant noise fail!
Possibilities

• Track and time false latch switching and signal failures only when values can be caught in latch
  + Flags only the “real” fails
    – Requires correctly “timing” the glitches to the latches (with consideration of process variations)
• Track signal values only when cause latch to switch
  – Failures criterion very sensitive to process variations
• Localize failure at the gate of “amplification”
  – Conservative metric
    + More robust against process variation (better yield)
    + Localized source of failure
Fundamental feature of digital circuits

- The inherent noise immunity of digital systems comes from high-gain restoring logic gates with very nonlinear voltage transfer characteristics.

But technology trends have tipped the scales!
Noise stability
On-chip time-domain measurements

- Being able to diagnose noise problems on chip is going to increasingly require test circuits to sample time-domain waveforms. Subsampling techniques are key!
On-chip time-domain measurements

Sampler

clk#

clk

en

signal
Measuring noise on-chip

- Substrate noise measurement
- Coupling noise measurement
- Time-domain techniques
Noise graph traversal

Partitioning at FET gates
Noise graph traversal

Propagate noise
Check noise stability
Pick up circuit noise

Partitioning at FET gates

Use expanded power-supply rails

A B C
Noise graph traversal

Propagate noise
Check noise stability
Pick up circuit noise

Pick up interconnect noise

Partitioning at FET gates

Checked stability with compressed power-supply rails

A
B
C

A
B
C

November 9, 2000
ICCAD 2000
Superposition in action

Calculating noise at node E

November 9, 2000

ICCAD 2000
What information can we give to static noise analysis to do better?

Information available from logic and timing analysis

- Static logic constraints
- Hazard-free (or switching) logic constraints
- Timing orthogonality
Timing orthogonality
Timing orthogonality

November 9, 2000

ICCAD 2000
Simple, simple approach

Miller Effect

Nominal
Simple, simple approach

Miller Effect

Worst-case
Simple, simple approach

Miller Effect

Best-case
Nonmonotonic waveforms and aggressor alignment
Noise on delay effects

Perform grounded capacitance global timing

Freeze arrival times and slews at all drivers

Using frozen timing data for secondary nets, find effective load models and receiver delays and slew
Static Noise Analysis

Two-level hierarchy

- Macro
- Global

Macro analysis

- Timing abstracts
- Noise abstracts
Global Harmony Architecture

- Full chip
  - Global
    - Layout
    - Schematic
  - Extraction
- Macro
  - Layout
  - Schematic
  - Extraction
- Static timing analyzer
- DCL timing abstracts
- Harmony
- Noise assertions
- Noise abstracts
- DCL Binary Dynamic Table
- Interconnect Subrule
- Model order reduction
- Global Harmony
Interaction between macro and global

Abstracts:
- Last-stage linearizations
- Noise tolerances
- Pin capacitances
- Propagated noise from output

Completely linear global problem!
Global Noise Analysis

- Superposition is used. Each coupled noise source is independently calculated. These are then added with aligned peaks.

- Secondary net drivers are held by their “resistances” when not switching.

Macro 1 -> Interconnect model -> Macro 2

Macro 3 -> Interconnect model -> Macro 4

November 9, 2000, ICCAD 2000
Cell-level noise

0.15um process, 1.5-V supply
22791-cell design
15354 nets with significant coupling

November 9, 2000
ICCAD 2000
Cell-level characterization

\[ \beta(V_{in} - V_T)^2 \text{ for } V_{in} > V_T \]
\[ 0 \text{ for } V_{in} < V_T \]
“Canonical” waveform

\[ v_{in}(t) = V_{dc} + k e^{-p_1 t} - k e^{-p_2 t} \]
"Canonical" model

Cell model

\[ \beta (V_{in} - V_T)^2 \text{ for } V_{in} > V_T \]
\[ 0 \text{ for } V_{in} < V_T \]

Easy analytic solution for noise and sensitivity.

November 9, 2000

ICCAD 2000
The fit

Characterization done for each noise type.
Noise avoidance in standard-cell designs

- Keep wire lengths short with controlled placement
- Tune drivers to interconnect
- Break long wires with repeaters
Interconnect analysis for noise
Interconnect crosstalk analysis

- Capacitive coupling interactions have been “traditionally” considered in static noise analysis
- Inductance (and inductive coupling) is now becoming a menacing problem on-chip
Net Complexes

- For each (primary) net, build a complex of (secondary) nets with significant coupling to the primary net.
- Introduces some “redundancy”
Coupling noise estimation is an important part of deciding which aggressors to keep as named aggressors and which to lump together as a virtual attacker.
Multiport Macromodels

- With the increased importance of coupling in timing and delay analysis, multiple drivers must frequently be considered in the analysis of a piece of the interconnect network.
- Multiport models can take the form of admittance, impedance, hybrid, or scattering parameters.
- Admittance has the advantage of being compatible with SPICE-like simulators.

\[ i = Yv \]
\[ v = Zi \]
Passivity

- Passivity is a statement that the network only stores and dissipates energy. It cannot actively generate it.
- Passivity implies certain properties on the admittance.

\[ \sum_{p=1}^{n} \int_{-\infty}^{t} v_k i_k \, dt' \geq 0 \]

\[ i) \quad Z(s^*) = Z^*(s) \]

\[ ii) \quad z^T (Z(s) + Z^T(s^*)) z \geq 0 \text{ for all complex values of } s \text{ with } \text{Re}(s) > 0 \text{ and for any complex } z \]
Multiport Driving-Point Admittance

Admittance generally preferred over impedance because of consistency with MNA.

Practical Issues

- Capacitances at the ports give a singular G matrix.
- Admittance is zero at dc (no dc path to ground).

\[
C\dot{x} = -Gx + Bv
\]

\[
i = B^T x + Dv + E\dot{v}
\]

\[
Y(s) = B^T (G + sC)^{-1} B + D + sE
\]

G and C are symmetric, positive-definite.
Multiport Implicit Reduction

Multiport Reduction Techniques

- Block Arnoldi
- Block Lanczos
- Block Symmetric Lanczos
- PRIMA

Explicit techniques (e.g., AWE) could also be used, but these would require column at a time evaluation.

\[
\begin{bmatrix}
  y_{11} & \cdots & \cdots & y_{1N} \\
  \vdots & \ddots & \ddots & \vdots \\
  \vdots & \ddots & \ddots & \vdots \\
  y_{N1} & \cdots & \cdots & y_{NN}
\end{bmatrix}
\]
Block Arnoldi

Block Arnoldi

\[ K_q = \text{span}\{R, AR, A^2R, \ldots, A^{N-1}R\} \]

\[ Y(s) = B^T (G + sC)^{-1} B + D + sE \]

\[ A = -G^{-1}C \quad R = G^{-1}B \]

\[ Y(s) = B^T (I - sA)^{-1} R + D + sE \]

**Practical issue:** Deflation may be necessary.

Neither passivity nor stability assured!

Change of variable \( \bar{x} = C^{1/2}x \) can be used prior to Arnoldi to guarantee stability.

November 9, 2000

ICCAD 2000
Block Lanczos

Block Lanczos

\[ K_q = \text{span}\{ \mathbf{R}, \mathbf{A}\mathbf{R}, \mathbf{A}^2\mathbf{R}, \ldots, \mathbf{A}^{\left\lfloor \frac{q}{N} \right\rfloor} \mathbf{R} \} \]

\[ K_q' = \text{span}\{ \mathbf{B}, \mathbf{A}^T \mathbf{B}, (\mathbf{A}^T)^2 \mathbf{B}, \ldots, (\mathbf{A}^T)^{\left\lfloor \frac{q}{N} \right\rfloor} \mathbf{B} \} \]

\[ \mathbf{Y}(s) = \mathbf{B}^T (\mathbf{G} + s\mathbf{C})^{-1} \mathbf{B} + \mathbf{D} + s\mathbf{E} \]

\[ \mathbf{A} = -\mathbf{G}^{-1} \mathbf{C} \quad \mathbf{R} = \mathbf{G}^{-1} \mathbf{B} \]

\[ \mathbf{Y}(s) = \mathbf{B}^T (\mathbf{I} - s\mathbf{A})^{-1} \mathbf{R} + \mathbf{D} + s\mathbf{E} \] \[ \Rightarrow \mathbf{Y}(s) = \mathbf{B}^T \mathbf{V}_q (\mathbf{I} - s\mathbf{T}_q)^{-1} \mathbf{W}_q^T \mathbf{R} + \mathbf{D} + s\mathbf{E} \]

**Practical issue**: Breakdown possible!

Neither passivity nor stability assured!
Block Symmetric Lanczos

\[ Y(s) = B^T (G + sC)^{-1} B + D + sE \]

\[ Y(s) = B^T V_q (I - sH_q)^{-1} V_q^T B + D + sE \]

\[ H_q = V_q^T G^{-1} C G^{-1} V_q \]

\[ V_q^T G^{-1} V_q = I \]

Passivity assured for RC circuits.
Using the matrix of Arnoldi vectors, just do:

\[ \tilde{C} = V^T C V^T \]
\[ \tilde{G} = V^T G V^T \]
\[ \tilde{B} = V^T B \]

\[ Y(s) = \tilde{B}^T (\tilde{G} + s\tilde{C})^{-1} \tilde{B} + D + sE \]

Guaranteed passivity for RLC circuits.
Practical Issues

• Watch the number of ports
  – Net complexes
  – Hybrid transfer/driving-point functions
  – Partitioning
Hybrid transfer/admittance
multiports

- For fixed, “linearizable” loads, as when driving MOS gates, it is convenient to reduce the number of driving ports
- At these “tap points”, we instead calculate a transfer function from each of the ports.

\[
\begin{bmatrix}
  i_{\text{port}} \\
  v_{\text{tap}}
\end{bmatrix} = \begin{bmatrix}
  Y & H
\end{bmatrix} \begin{bmatrix}
  v_{\text{port}}
\end{bmatrix}
\]
# Summary of Multiport Techniques

<table>
<thead>
<tr>
<th></th>
<th>Moments</th>
<th>Passive RC?</th>
<th>Passive RLC?</th>
<th>Number of state matrices</th>
</tr>
</thead>
<tbody>
<tr>
<td>Block PVL</td>
<td>2q</td>
<td>No</td>
<td>No</td>
<td>1</td>
</tr>
<tr>
<td>Block syPVL</td>
<td>2q</td>
<td>Yes</td>
<td>No</td>
<td>1</td>
</tr>
<tr>
<td>Block Arnoldi</td>
<td>q</td>
<td>No</td>
<td>No</td>
<td>1</td>
</tr>
<tr>
<td>Prima</td>
<td>q</td>
<td>Yes</td>
<td>Yes</td>
<td>2</td>
</tr>
</tbody>
</table>
Recursive Convolution

This is an efficient technique for convolving an input waveform with a transfer function. **Constant cost at each time point.**

\[
v_{\text{receiver}}(t) = \int_{0}^{t} H(t - \tau)v_{\text{driver}}(\tau)d\tau
\]

\[
v_{\text{driver}}(t) = \sum_{r} \left[ d_{r}^{\prime} + e_{k}^{r}t + f_{k}^{r} e^{p_{r}(t-t_{k})}\right]
\]

\[
d_{k}^{r} = d_{k-1}^{r} + \left(\frac{-k_{r}}{p_{r}} + \frac{k_{r}t_{k}}{p_{r}}\right)dm_{k-1}
\]

\[
e_{k}^{r} = e_{k-1}^{r} + \left(\frac{-k_{r}}{p_{r}}\right)dm_{k-1}
\]

\[
f_{k}^{r} = \left( f_{k-1}e^{p_{r}(t_{k}-t_{k-1})} + \frac{k_{r}}{p_{r}} dm_{k-1} \right)
\]
Recursive Convolution and Circuit Simulation

When applied to multiport driving point admittances, the recursive convolution approach can provide a convenient stencil to incorporate reduced-order models into the MNA formulation of SPICE-like circuit simulators.
Direct stamp

- Can directly stamp the reduced G, C, B, and L matrices into SPICE
- Involves voltage-controlled current sources and current-controlled current sources.
Example net

5 mm

November 9, 2000

ICCAD 2000
Inductance
Why inductance is important

Line delay

(Line length)

\[(\frac{RC}{2})\]

\[(\frac{LC}{2})^{1/2}\]

Fastest slew time

November 9, 2000

ICCAD 2000
Why inductance is important

\[ \text{Line delay} = \frac{(RC)}{2} \]

\[ \frac{(LC)^{1/2}}{2} \]

Fastest slew time

Line length
Why inductance is important

This analysis may require an “effective” inductance that considers simultaneous switching.

November 9, 2000  ICCAD 2000
Inductance versus capacitance

• Capacitance
  – Locality problem easy… electric field lines “suck up” to nearest neighbor conductors
  – Local calculation hard… hence all the effort in “accuracy”

• Inductance
  – Locality problem hard… magnetic field lines are not local; current returns can be complex
  – Local calculation easy… no strong geometry dependence… analytic formulae work very well.
Typical on-chip power distributions

No ground plane!

November 9, 2000  ICCAD 2000
Observations

• At the frequencies at which $R$ becomes frequency-dependent, $L$ dominates.
• At the frequencies at which $L$ dominates, current returns have “collapsed.”
• Power and ground lines are always available as low-impedance current returns at high frequencies.
Partial inductance and return-limited inductance

Current return at infinity.
Return-limited inductance extraction

• Need to determine which mutual inductances to discard and wish to use the power-ground network as an “always-available” current return.

• To do this, we:
  – Use the power-ground distribution to divide the interconnect into disjoint interaction regions. Mutual inductances between interaction regions are discarded.
  – Power-ground wires within the interaction region act as a “distributed ground plane”.

• A set of geometry-based matrix decomposition rules guide the interaction region definition (halo rules).
Interaction region

signal
Interaction region

- Power/ground
- Signal
Interaction region

- Interaction region
- Power/ground
- Signal
Assura RLCX architecture

GDSII → LVS → Edge file → Device files → Resistance extraction (rex) → Edge file with resistor cuts → Resistance files → Capacitance extraction → Capacitance file

Resistance files → Spice netlist generator → Spice files

Interaction region file → regcal → Microinductance file → wirecal → Inductance file
Inductance in the design flow

Noise

• Detailed SPEF format must be enhanced to include mutual inductances.

• Multiport interconnect macromodels must support inductance for noise (glitch) and noise-on-delay analysis
  – Have already done this with CadMOS’ PacifIC and CeltIC noise tools.
If B is quiet, acts as an ac ground. Total cap on A is \( C_1 + C_2 + C_3 + C_4 \).
Inductance in the coupling noise problem

\[ w > R/L \]

Inductance of A determined by loops through P/G lines.
How much B reduces inductance depends on how “easily” return currents find their way into B.
Inductance in the coupling noise problem

If A is static low and B is switching low->high, on the far end of A get VL noise.
If A is switching low->high and B is switching low->high, A switching sees lower effective C.
If A is switching high->low and B is switching low->high, A switching sees higher effective C.
Inductance in the coupling noise problem

If A is static low and B is switching low->high, the far end of A gets “VL*” (and VL) noise.
If A is switching low->high and B is switching low->high, A switching sees higher effective L.
If A is switching high->low and B is switching low->high, A switching sees lower effective L.
Example
Example (con’t)

Keep bit 7 quiet and switch the other bits.

Worst VH noise occurs for bits 6,8 switching high to low and other bits switching from low to high.

Worst VH* noise occurs for bits 6,8 switching low to high and other bits switching from high to low.
Substrate extraction

\[ \varepsilon \frac{\partial}{\partial t} (\nabla \cdot E) + \frac{1}{\rho} \nabla \cdot E = 0 \]

Ignores magnetic field

Two approaches to solving this:

- Differential form: finite difference
- Integral form: boundary element
Finite difference

Substrate is broken into a number of rectangular volumes, resulting in a mesh circuit consisting of nodes interconnected by branches of resistors and capacitors in parallel. May also discretize into more “abnormally” shaped volumes.

Capacitors can be ignored up to several GHz (dielectric relaxation time).

Very discretization dependent. Extension of PEEC models to substrate.
Boundary element

- Resistance-only case
- Source and observation point as ports on the surface of the substrate
- 3D problem reduced to 2D

\[ \phi(r) = \int_{S} J(r')G(r, r')da \]
The most significant coupling paths to critical transistors are modeled with higher accuracy.
Heavily doped and lightly doped substrates

Lightly-doped
- p-type 0.1 Ohm-cm
- p-sub 20 Ohm-cm
  
Heavily-doped
- p-type 1 Ohm-cm
- p-type 15 Ohm-cm
- p+sub 1 mOhm-cm

November 9, 2000
ICCAD 2000
How noise couples into the substrate

- Coupling from the digital power supply
- Coupling from switching source-drain nodes
- Impact ionization in the MOSFET channel
Triple well process

Also trench isolation and SOI.
Kelvin Grounding

November 9, 2000  ICCAD 2000
Techniques for minimizing substrate interaction
Substrate Extraction: Modeling the Substrate

• Determine locations on substrate that connect to circuit (devices, well taps, substrate contacts)
• Extract 3D RC model of substrate
  – BiCMOS, twin-well, triple-well, standard CMOS, SiGe, deep trench

November 9, 2000
ICCAD 2000
Substrate Extraction: Modeling Wells / Interconnect

- Extract RC model for wells
Substrate analysis

AVDD

Rs

Cp

Rp

Substrate

AVDD

Rs

Cp

Rp

Insert Substrate Noise

November 9, 2000

ICCAD 2000
Substrate noise waveforms

Noise on sensitive instance
Major Noise Contributors
Substrate as high-frequency current return

• Magnetic coupling to the substrate (in addition to capacitive coupling). Effects interconnect response as well as substrate noise (emerging issue)

• Has an effect already for RF passives (spiral inductors)
Substrate noise detector

V_{bias} \quad pFET source follower, isolated from substrate by nwell connected to a p+ substrate tap.
Power supply noise
Power supply network topologies

Routed networks

- More common in technologies with limited routing resources (3-4 layers) and moderate power requirements
- Trunks distribute power to major blocks (and there may be a grid locally)
- Common in analog/mixed-signal design when multiple power/grounds are frequently required.
Power supply network topologies

Routed networks

+ Uses little wiring resources
+ Easy to distribute multiple power/ground
- Power may run long distances from periphery to center of chip (IR drops)
Power supply network topologies

Gridded networks

- Very common in high-performance (high-power) digital integrated circuit designs (e.g., microprocessors)
- Common in technologies with large number of metal layers and C4 bonding technology
- Grids reduce the effective resistance of the power-ground distribution and also help to reduce the on-chip inductances
- Chews up a significant portion of metal resources
Power supply network topologies

C4 technology

- Pads distributed across the chip with solder bumps
- Solder bumps placed on each pad location
- The chip is then flip-chip mounted onto the package
- Very low inductance pad connections

Bonding wire: 5 nH typical inductance
Flip-chip: < 1 nH of inductance,
< 1 pF of capacitance
Power supply network topologies

Power planes

- Grids taken to the extreme: two dedicated planes for power and ground (e.g., DEC Alpha)
- For very high current demands, though not as necessary with C4 technology
- Holes cut through to make connections
- Very expensive in wiring resources, but really reduces inductances and coupling capacitance
Hierarchical analysis

Divide the power distribution into “macro” and “global.” At the interface between the macro and global power distributions (usually on a via level), define a set of power points.

November 9, 2000

ICCAD 2000
Modelling the transient current demands of the switching digital

Current sources at the power points

I_{peak}  I_{ave}  T_{cycle}  I_{peak}  I_{ave}
IR drop hot spots

S/390 Microprocessor Design

November 9, 2000

ICCAD 2000
Decoupling capacitance “hierarchy”

<table>
<thead>
<tr>
<th>Location and type</th>
<th>Frequency range</th>
</tr>
</thead>
<tbody>
<tr>
<td>On-chip explicit FET capacitors and powergrid metal capacitance</td>
<td>10 GHz</td>
</tr>
<tr>
<td>Non-switching circuits and nwell capacitance</td>
<td>1 GHz</td>
</tr>
<tr>
<td>Surface-mount capacitors in chip package or MCM</td>
<td>100 MHz</td>
</tr>
<tr>
<td>Circuit-board level ceramic capacitors</td>
<td>10 MHz</td>
</tr>
<tr>
<td>External power-supply aluminum electrolytic capacitors</td>
<td>0.1 MHz</td>
</tr>
</tbody>
</table>

Frequency range of usefulness of given level determined by smaller of \( \frac{1}{2}pRC \) or \( \frac{1}{2}p\sqrt{LC} \).
Decoupling capacitance

Nwell capacitors

Time constant typically between 250ps and 500ps

Nonswitching circuits

November 9, 2000

ICCAD 2000
Power supply resonances

- In high performance microprocessors, there are typically resonances, $1/2\cdot\pi\cdot\sqrt{LC}$, at 1/10 the clock frequency.
- These can become problematic if they are high Q resonances, since some of the core energy will inevitably infringe onto the tank resonant range.
Low-noise digital logic

- One approach to quieting the talker: apply current-mode logic, particularly for large on-chip capacitances. Dissipate static power, but low swing and if under high-frequency switching activity (serialized) a net power win.
SOI
Noise in SOI

Buried oxide

N+ N+ P N+ gate
Noise in SOI

- Parasitic bipolar effect
- Body-effected threshold voltage variations
- Reduced “grounded” component of interconnect capacitance
Advantages of SOI?

- Reduced source/drain diffusion capacitance
- Improved stack performance
- Some reduction in wire capacitance
- Gives advantage to certain circuit families (e.g., passgates)
- Why not just put body contacts everywhere?

Body-affected $V_t$ shift less

Internal node discharges faster

November 9, 2000

ICCAD 2000
SOI: Parasitic bipolar

Try to fix with S/D extensions and retrograde doping, but problem gets worse with length scaling.

Requires special sensitization in static noise analysis.

November 9, 2000

ICCAD 2000
Parasitic bipolar leakage
Threshold voltage variation

This FET is much more sensitive to VL noise

No special sensitization for static noise analysis required
Body voltage determined by...

- Capacitive coupling of gate, source, drain, and body
Body voltage determined by...

- Capacitive coupling of gate, source, drain, and body
- Diode currents at source-body and drain-body junctions
Body voltage determined by...

- Capacitive coupling of gate, source, drain, and body
- Diode currents at source-body and drain-body junctions

reverse biased

Time scale: $\gg$ cycle time

November 9, 2000
Body voltage determined by...

• Capacitive coupling of gate, source, drain, and body
• Diode currents at source-body and drain-body junctions
• Impact ionization currents

hole collection
charges body

Time scale: >> cycle time
State diagram view of switching

\[ G = H \]
\[ S/D = L/L \]

\[ G = L \]
\[ S/D = L/H \]
5a
State diagram view of switching

NFET Body Voltage

G=H
S/D=L/L

G=L
S/D=L/H

5a
State diagram view of switching

NFET Body Voltage

capacitive coupling

G=H S/D=L/L

G=L S/D=L/H

5a

November 9, 2000

 ICCAD 2000
State diagram view of switching

G=H  
S/D=L/L

G=L  
S/D=L/H

NFET  
Body  
Voltage

0 V  
0.437 V

diode leakage  
capacitive coupling
Body voltage variation

• Separated consideration of capacitive coupling from dc mechanisms, which influence floating body charge

• Floating body charge gives the device memory. Use the reference body voltage to represent this charge.

• Four techniques for reference body voltage estimation based on different switching assumptions or knowledge.
Body-voltage estimation

- Full-uncertainty analysis
- Modified-accessibility analysis
- Accessibility analysis
- Detailed analysis
Full uncertainty

No knowledge of switching behavior. Circuit could sit in any state for an indeterminate amount of time.
Full uncertainty estimation

g s d
H L L 1
H H H 2
L L L 3
L H H 4
L L H 5
H H L 6

states 0 0.5 1.0 1.5 2.0 2.5

Body Voltage
Full uncertainty estimation

Body Voltage

special consideration

states

November 9, 2000
Full uncertainty estimation
Full uncertainty estimation

<table>
<thead>
<tr>
<th>States</th>
<th>Voltage Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>H L L</td>
<td>1.75</td>
</tr>
<tr>
<td>H H H</td>
<td>DC</td>
</tr>
<tr>
<td>L L L</td>
<td>DC</td>
</tr>
<tr>
<td>L H H</td>
<td>DC</td>
</tr>
<tr>
<td>L L H</td>
<td>-0.29</td>
</tr>
<tr>
<td>H H L</td>
<td>1.0</td>
</tr>
</tbody>
</table>

Body Voltage States

0.0 0 0.5 1.0 1.5 2.0 2.5

November 9, 2000
Accessibility analysis

Circuit under steady switching activity so that every accessible state of every FET is visited with “reasonable” frequency.
Accessibility estimation

Accessibility states:

1. H L L L (DC 0.6 forward-bias)
2. H H H H (DC 3.1 forward-bias)
3. L L L L (DC)
4. L H H H (DC 3.1)
5. L L H H (DC 0.6 forward-bias)
6. H H L L (DC)

Body Voltage states:

0.0 0.5 1.0 1.5 2.0 2.5

November 9, 2000
Accessibility estimation

states | 0  | 0.5 | 1.0 | 1.5 | 2.0 | 2.5 | Body Voltage in state 2

November 9, 2000
Accessibility estimation

<table>
<thead>
<tr>
<th>States</th>
<th>Body Voltage</th>
</tr>
</thead>
<tbody>
<tr>
<td>L L L</td>
<td>H H L</td>
</tr>
<tr>
<td>H L L</td>
<td>1.75</td>
</tr>
<tr>
<td>H H H</td>
<td>2.47</td>
</tr>
<tr>
<td>L L L</td>
<td>DC</td>
</tr>
<tr>
<td>H L L</td>
<td>DC</td>
</tr>
<tr>
<td>L L H</td>
<td>-0.29</td>
</tr>
<tr>
<td>H H L</td>
<td>0.75 1.0</td>
</tr>
</tbody>
</table>
Modified accessibility analysis

• Compromise between full uncertainty and accessibility analysis
• Tag certain nets as “active,” which means that you guarantee that these nets switch reasonably frequently.
• All you need to get this to work in transistor-level analysis tools or for cell characterization tools:
  – Precharacterized displacement curves
  – Precharacterized V-zero, V-forward reference state values
• For rules, these have to be conditional on the active net tagging.
Modified accessibility analysis

Body-voltage maximum case

For full uncertainty, I’ll assume that the FET sits in the state $G=L/S, D=H$ for the entire “prelude.”

For modified accessibility analysis, I’m forced to visit (at least) $G=L, S=L, D=H$. 
Static noise analysis

<table>
<thead>
<tr>
<th>Transistor</th>
<th>Actual body voltage (state 5)</th>
<th>Estimated max body voltage (state 5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1 for case 1</td>
<td>1.0 V</td>
<td>1.2 V</td>
</tr>
<tr>
<td>M3 for case 2</td>
<td>0.05 V</td>
<td>0.349 V</td>
</tr>
<tr>
<td>M5 for case 3</td>
<td>0.720 V</td>
<td>0.746 V</td>
</tr>
<tr>
<td>M5 for case 4</td>
<td>0.11 V</td>
<td>0.349 V</td>
</tr>
</tbody>
</table>
Static noise analysis
References

Power-supply noise

References

Inductance

References

SOI

References

On-chip time-domain measurement circuitry

References

Substrate noise

Clock Distribution
Clock Distribution Outline

- Metrics/Goals
- Technology Trends and their impact on clocking
- Types of clock distribution networks
- Active de-skewing
Function of clock distribution network

• Synchronize millions (billions) of separate elements
  – Within a time scale on order of ~10ps
  – At distances spanning 2-4 cm
    • Ratio of synchronizing distance to element size on order of $10^5$
  – Reference: light travels <1 cm in 10ps
Metrics/Goals

• Besides basic connectivity, what makes a clock network good or bad?
  – Skew
  – Jitter
  – Power
  – Area
  – Slew rates

• Descriptions follow
Clock Skew

- The most “high-profile” of the clock network metrics
- Defined as: Maximum difference in arrival times of clock signal to any 2 latches/FF’s fed by the network

Skew = max | t₁ – t₂ |

Fig. From Zarkesh-Ha
Clock Skew

• Causes:
  – Designed (unavoidable) variations – mismatch in buffer load sizes, interconnect lengths
  – Process variation – process spread across die yielding different $L_{\text{eff}}$, $T_{\text{ox}}$, etc. values
  – Temperature gradients – changes MOSFET performance across die
  – IR voltage drop in power supply – changes MOSFET performance across die

• Note: Delay from clock generator to fan-out points (clock latency) is not important by itself
  – BUT: increased latency leads to larger skew for same amount of relative variation
Clock Skew

• Effect:
  – Eats into timing budget
  – Needs to be considered for maximum (setup) and minimum (hold) path timings

Ref: Simplex website
Skew Histogram, IBM µP

# of paths

ps

Ref: IBM, JSSCC, 11/99

November 9, 2000

 ICCAD 2000
Jitter

- Clock network delay uncertainty
- From one clock cycle to the next, the period is not exactly the same each time
- Maximum difference in phase of clock between any two periods is jitter

![Diagram](image.png)

**NOTES:** JITTER $J_1 = t_2 - t_1$.
JITTER $J_2 = t_3 - t_2$. 
Jitter

• Caused by variations in clock period that result from:
  – Phased-lock loop (PLL) oscillation frequency
  – Various noise sources affecting clock generation and distribution
    • Ex. Power supply noise which dynamically alters the drive strength of intermediate buffer stages
Jitter Impact on Timing Budget

- Needs to be considered in maximum path timing (setup)
- Typically on the order of 50ps in high-end microprocessors
Jitter Measurement

Jitter can be reduced by minimizing power supply noise (IR and $L^*di/dt$).

Peak-to-peak period jitter = 32ps

Ref: Intel, ISSCC00
Clock Power

• Power consumption in clocks due to:
  – Clock drivers
  – Long interconnections
  – Large clock loads – all clocked elements (latches, FF’s) are driven

• Different components dominate
  – Depending on type of clock network used
  – Ex. Grid – huge pre-drivers & wire cap. drown out load cap.
Clocks: Power-Hungry

Not only is the clock capacitance large, it switches every cycle!

$$P = \alpha C V_{dd}^2 f$$
Low Power Clocking Techniques

Gated clocks

• Prevent switching in areas of the chip not being used
• Easier in static designs

Edge-triggered flops in ARM rather than transparent latches in Alpha

• Reduced load on clock for each latch/flop as well as eliminated spurious power-consuming transitions during flow-through of latches
Clock Distribution Metric: Area

- Clock networks consume silicon area (clock drivers, PLL, etc.) and routing area
- Routing area is most vital
- Top-level metals are used to reduce RC delays
  - These levels are precious resources (unscaled)
  - Power routing, clock routing, key global signals
- By minimizing area used, we also reduce wiring capacitance & power
- Typical #’s: Intel Itanium – 4% of M4/5 used in clock routing
Slew Rates

• To maintain signal integrity and latch performance, minimum slew rates are required
• Too slow – clock is more susceptible to noise, latches are slowed down, eats into timing budget
• Too fast – burning too much power, overdesigned network, enhanced ground bounce
• Rule-of-thumb: $T_{\text{rise}}$ and $T_{\text{fall}}$ of clock are each between 10-20% of clock period (10% - aggressive target)
  – 1 GHz clock; $T_{\text{rise}} = T_{\text{fall}} = 100-200\text{ps}$
Slew Rates

- Latch set-up times are dependent on clock input slew rates (eats into timing budget)
- Short-circuit power grows with larger slew rates
  - This can be significant for large clock drivers

\[ T_{\text{setup}} = 200 + 0.33 T_{\text{slew}} \ (ps) \]
Clock Distribution Example

Alpha 21264 clock distribution -- grid + H-tree approach

Power = 32% of total

Wire usage = 3% of metals 3 & 4

4 major clock quadrants, each with a large driver connected to local grid structures
Alpha 21264 Skew Map

GCLK Skew
(at Vdd/2 Crossings)

GCLK Rise Times
(20% to 80% Extrapolated to 0% to 100%)

November 9, 2000
Ref: Compaq, ASP-DAC00

ICCAD 2000
Technology Trends: Timing

- Clock period dropping fast, so skew must follow accordingly
- Slew rates must scale with cycle time also
- Jitter – PLL’s get better with CMOS scaling but other sources of noise increase
  - Power supply noise more important
  - Switching-dependent temperature gradients
Technology Trends: New Interconnect Materials

Copper reduces RC slew degradation and potential skew

Low-k dielectrics decrease clock power and improve latency/skew/slew rates

Ref: IBM, JSSCC, 11/98

ICCAD 2000
Technology Trends: Power

- Heavily pipelined designs ⇒ more latches, more capacitive load for clock
- Larger chips ⇒ more wirelength needed to cover the entire die
- Complexity ⇒ more functionality and devices means more clocked elements
- Dynamic logic ⇒ more clocked elements
Meeting skew requirements is relatively easy….
With an unlimited power budget!

Ex. Wide wires – minimize RC product but increase total C
Ex. Size up drivers – limit latency (translates to skew) but buffer cap. jumps
SoC power requirements

- SoC’s have more stringent power limitations due to packaging constraints
  - Plastic packaging, power ~ 2-3 W
- This pushes the skew-power tradeoff towards higher skew
- Intolerable considering the drive for high performance
  - SoC’s are good candidates for power-friendly skew-reducing “tricks”
Clock Network Design Example

- Clock distributed from center of chip and periodically buffered; typical three levels of distribution
  - global clock (L4 or above)
  - area clock (L3, L2)
  - local clock (L1, L0)

- Clock noise sensitivity must be reduced
  - IR drop (or supply) noise
    - differential routing can reduce noise sensitivity of clock buffers
  - capacitive coupling noise
    - clock routing in top layers with shields on either side and above/below
Clock Tree Topology

• Clock Tree
  – PLL connects to a cluster of 4 Global clock buffers
    • need to balance wire routing to global buffers to reduce skew
  – Global buffers drive 3 levels of clock tree
    • L3 and L2 are area clock buffers
    • global-area buffers need to implemented as cascaded stages of
      INVs to reduce input load and increase output drive strength
      (w/ ratio = 2)
  – L1: Local clock buffer
  – Local clock buffer (L0) will drive locally clocked elements
  – Load attached at L0: gate load + wire load
  – Fanout (of load) at all output buffers ~ 4-6
    • skew should be reasonable (100-200ps)
Local Clock Domains

- Fanout limited to reduce skew and maintain slew rates
- Too small a fanout gives diminishing performance returns
- Increased number of local buffers increases power and area (more upstream buffers too)

Local clock domain size drops quickly with higher performance requirements

Ref: IBM website, Carrig
Global Clock Buffer Structure

- Differential clock lines distributed to global clock buffers
Clock Management

• **Mini-Block level Clock**
  - Count clock nodes per Std. Block
    - total load (gate + wire)
  - Determine local clock tree levels/size
  - Estimate size of area clock buffer
  - Reserve space for clock buffers and clock wires/shields
  - apply balanced clock routing

• **Top-level Clock**
  - Add clock grid topology for each Std. block
  - Estimate PLL to local buf. delays for all Std.blocks
  - Determine worst case delay
  - Add buffer-chains to align delays
  - Consider electromigration for high-activity, heavily-loaded wires
  - Add shielding inside, if necessary
  - Top-level balanced clock routing

Courtesy: S. Muddu
Network Types: Grid

- Gridded clock distribution was common on earlier DEC Alpha microprocessors
- Advantages:
  - Skew determined by grid density and not overly sensitive to load position
  - Clock signals are available everywhere
  - Tolerant to process variations
  - Usually yields extremely low skew values
Grid Disadvantages

- Huge amounts of wiring & power
  - Wire cap large
  - Strong drivers needed – pre-driver cap large
  - Routing area large
- To minimize all these penalties, make grid pitch coarser
  - Skew gets worse
  - Losing the main advantage
- Don’t overdesign – let the skew be as large as tolerable
- Still – grids seem non-feasible for SoC’s
Network Types: Tree

- **Original H-tree (Bakoglu)**
  - One large central driver
  - Recursive H-style structure to match wirelengths
  - Halve wire width at branching points to reduce reflections

Fig. courtesy of Zarkesh-Ha
H-Tree Problems

• Drawback to original tree concept
  – slew degradation along long RC paths
  – unrealistically large central driver
    • Clock drivers can create large temperature gradients (ex. Alpha 21064 ~30° C)
  – non-uniform load distribution

• Inherently non-scalable (wire resistance skyrockets)

• Solution to some problems
  – Introduce intermediate buffers along the way
  – Specifically at branching points
Buffered Clock Tree

L2

Drives all clock loads within its region

PLL

WGBuf

NGBuf

SGBuf

EGBuf

L3

Other regions of the chip

November 9, 2000

ICCAD 2000
Buffered H-tree

• Advantages
  – Ideally zero-skew
  – Can be low power (depending on skew requirements)
  – Low area (silicon and wiring)
  – CAD tool friendly (regular)

• Disadvantages
  – Sensitive to process variations
  – Local clocking loads are inherently non-uniform
Balancing a Tree

Some techniques:

a) Introduce dummy loads

b) Snaking of wirelength to match delays

Con: Routing area often more valuable than Silicon

November 9, 2000

ICCAD 2000
RLC effects in clock network

Inclusion of inductive effects in global clock tree shows:

- Line delay increased & signal slopes sharper
- Effective gate delay decreased

<table>
<thead>
<tr>
<th>RLC and Return Path Extraction</th>
<th>Line Delay</th>
<th>Gate Delay</th>
</tr>
</thead>
<tbody>
<tr>
<td>Addition of return path resistance</td>
<td>+ 27 %</td>
<td>- 3 %</td>
</tr>
<tr>
<td>Addition of inductance</td>
<td>+ 8 %</td>
<td>-10 %</td>
</tr>
<tr>
<td>Difference from RC Model</td>
<td>+ 35 %</td>
<td>- 13 %</td>
</tr>
</tbody>
</table>

Not just line inductance but return path resistance is important!

Ref: Intel, ISSCC00
Clock Integrity

- Shield everywhere
  - Laterally and above/below
  - Provide *current return paths* and eliminate coupled noise effects (both C and L)
Clock Integrity

- di/dt for clock drivers can be enormous
  - All clocks should be switching at the same instant
- Potential for L*di/dt noise on power supply
- Explicit decoupling capacitance has been taken as solution to this problem
- Thin gate oxides used in silicon white space to create large (100+ nF) capacitance to supply charge
- Alpha 21264 required additional decoupling capacitance at package level to limit switching noise (!) – future trend...
Clock Shielding

- How much does shielding help? Or reference planes?
- Is it worth the area penalty?

Ref: Compaq, ASP-DAC00
Impact of Reference Planes in Power Distribution

Conditions not given!

Aluminum

November 9, 2000
Ref: Compaq, ASP-DAC00

ICCAD 2000
Clock Grid Simulations / Reference Planes

Copper wiring allows for smaller wires, finer grid pitch, lower power

Eliminating need for reference planes

Ref: Restle, VLSI00
Reduce Self Inductance

Dedicated Ground Planes

G

W1

C

G

Wg

Dedicated G.P.

Ref: Massoud, DAC98
Guard Traces vs. Reference Plane

- Below 5 GHz, guard traces appear better

Ref: Massoud, DAC98
ICCAD 2000
LF Current Distribution

- Current Flow in Dedicated G.P.
- Current Flow in Guard Traces
- Smaller Current Loops
- Smaller Inductance
- Current Spreads Through outer Bigger Current Loops

Ref: Massoud, DAC98
HF Current Distribution

Current Flow in Dedicated G.P.  
Current Flow in Guard Traces

Current Concentrates Underneath Signal Line  Smaller Current loops, and Smaller Inductance

Top View

Ref: Massoud, DAC98
Reduce Self Inductance

Interdigitating the Signal Line

- Inductance Decreases by 43%
  - Resistance Not Increased
  - Capacitance Increases by 27% (bad!)
  - Allocated Space Increases by 11% (tolerable for extreme cases)

- This example – 10 µm lines
  Narrower (more common) lines will exhibit less L reductions

Ref: Massoud, DAC98
Interdigitating Signal Lines

3 Lines

L2 = L1 / 4  
C2 = 1.6 C1  
R2 = 1.3 R1

11 Lines

ωL
R = 1.07

ωL
R = 0.20

Decreased by a factor of 5.4

RC analysis

Ref: Massoud, DAC98

November 9, 2000

ICCAD 2000
Shielding Overview

- Overall, most global signals will benefit more from neighboring shield wires than other approaches
  - Reference planes
  - Interdigitation

- Clock signals – always shielded
  - Better question – when do other signal wires need to be shielded?
Inductance: Signal Shielding

- Double spacing to return path -- increased ringing and delay
- By 50-nm, most global wires will need to be shielded with nearby ground/\(V_{dd}\) lines (within a few pitches)
- Strong drivers may lead to overshoot/ringing effects
- Delay optimal buffers may be detrimental to performance due to inductive effects
- Use of shield wires take up global routing resources
Shield Wire Usage

- Min. pitch 0.25 μm wiring not prone to (self) inductive effects
- At 0.05 μm, min. pitch global wires need return path within 3 spacings
  - Wider wires require closer shield wires
- Unsealed flat wiring used (W = 2 μm)
High-level View of Tree Network
Hierarchical Clock Distribution

- As with many other large-scale problems, exploiting hierarchy can present new solutions
- Establish distinct global and regional/local distribution networks
- Each can be of a different type!
  - Not all or nothing anymore
- Mix and match clock network types to maximize advantages, minimize disadvantages
Network of choice in high-performance

- Globally – Tree
- Why?
- Power requirements are reduced compared to global grid
  - Smaller routing requirements, frees up global tracks
- Trees are easily balanced at the *global* level
  - Keeps global skew low (with minimal process variation)
Network of Choice

- Locally – Grid
- Why?
- Smaller grid distribution area allows for coarser grid pitch
  - Lower power in interconnect
  - Lower power in pre-drivers
  - Routing area reduced
- Local skew is kept very small
- Easy access to clock by simply connecting to grid
Scaling of Distribution Networks

- **Buffered H-trees**
  - Regular, low-power, acceptable skew
  - Scalable although # of sub-blocks will rise with shrinking timing budget
- **Grid**
  - As chips get larger, so do grids
  - Power and routing area penalties increase
Clock Distribution: Simple Models

H-tree block size can be determined based on allowable (local) skew

\[ T_{\text{skew}} = 0.377 \left( 1.1^2 R_w C_w L^2 + 1.1 R_w L C_{\text{latches}} \right) \]
\[ + 0.693 \left( 1.1^2 C_w L R_{\text{dev}} \right) \]

10% variation in wire and device resistance, capacitance assumed

\( C_{\text{latches}} \) is estimated based on a heuristic to calculate # of latches

Depends on logic depth and local clocking domain area

Latches are distributed along wire

Comparisons with HSPICE results give <10% error for \( L_{\text{max}} \)

November 9, 2000
**High-Speed Clock Distribution?**

Is the buffered H-tree scalable?

Skew $\sim 10\%$ of cycle time
- local skew $\sim 5\%$

Using unscaled global wires, a buffered H-tree with 256 regions delivers $<10$ ps of local skew ($L_{\text{cluster}} = 1.7$ mm)

As long as we can increase the # of sub-blocks, the H-tree seems OK

BUT this analysis ignores global variations

Across a large die, these variations could dominate

OR other techniques could limit global variations…
Global Skew Reduction: Package Approach

- Most clock network latency occurs at global level (largest distances spanned)
- Latency $\propto$ Skew
- With reverse scaling, routing low-RC signals at global level becomes more difficult & area-consuming
Clock Distribution: A Global Alternative

- RC of package-level wiring is up to 4 orders of magnitude smaller than on-chip
- Global skew greatly reduced
- Lower capacitance yields lower power dissipation
- Opens up more global routing tracks
- Results are not yet conclusive

⇒ Incorporate *global* clock distribution into the package
⇒ Flip-chip packaging allows for high density, low parasitic access from substrate to IC
Reducing Skew – Circuit Technique

- Active de-skewing circuit (Intel) can be used to compensate for:
  - Mismatched loads in different clock regions
  - Processing variations
  - Temperature/voltage gradients (long-term)

- Also can easily allot skew for time borrowing

- Main idea:
  - Compare 2 sections of clock with phase detector and use variable delay lines to compensate for differences
Overall View: Clock with Deskew
Compare the reference clock to the feedback clock

If $\Delta t$ is same for 4 consecutive periods, adjust delay line to compensate (done at global/regional level)

Ref: Intel, ISSCC00
Reference Clock

• What is a reference clock?
  • Why not use this network for distribution???

• Skew due to global process variation and regional load mismatch is replaced by reference clock skew

• Ideally – ref. clock skew is << global/regional skew

• Ref. clock covers smaller area, precisely balanced, has predictable loading
Delay Line Specifics

Deskew range should be greater than the expected skew in the clock network (w/o deskew)

The control register connects passive loads to buffer outputs

Ref: Intel, ISSCC00
Variable Clock Buffers

Last stage of delay line uses variable size NMOS/PMOS to further control skew due to devices and load mismatches

- However, passive loads are less sensitive to environmental variations
De-skewing results, Intel

Projected max skew without deskew mechanism = 110ps

Max skew with deskew mechanism = 28ps

Regional Clock

Distribution Skew (ps)
Clock Summary

• Key tradeoff – clock skew vs. power
• Hierarchical clock distribution
  – Global tree distribution
  – Local grid distribution
• Active circuit skew compensation techniques are promising to further limit global skew
• Scaled designs – trees and grids get denser/finer, need automation for design and low power approaches (package level distribution?)
Process Variation
Economics of Yield

- Enormous cost of mega fabs ($2B +)
- Cycle of under and over manufacturing capacity
  - Full price for every die when under capacity
  - Silicon price in over capacity
- Performance distribution
  - Mean at design point - ordinary parts
  - High (low) tail - high (low) performance parts
  - Common in µP’s and DSP’s
- We want to shift mean rightwards (higher performance)
  - Very high return in high volume manufacturing

From Kahng, DAC99
Yield

• Functional
  – Chip doesn’t work
  – Short and open circuits in metal levels
  – Electromigration failure

• Parametric
  – Chips run at different speeds
  – Binning of parts, sell at different prices
  – Crosstalk noise, ILD variation, $I_{dsat}$ variation ($L_{eff}$, $T_{ox}$, $V_{th}$)

• Parametric yield loss is becoming dominant over defect-based yield loss as processing conditions are improved
Types of Variation

• Random
  – Modeling consists of approximating the random effect by a normal distribution
  – Knowing mean and $\sigma$, use statistical approaches (Monte Carlo, worst-case) to account
  – Example: random dopant fluctuations which impact device $V_{th}$

• Systematic
  – This type of effect should be studied and modeled deterministically to allow for design with variation in mind
  – Includes environmental variations such as IR drop, thermal gradients, dynamic delay
More Categories of Variation

- **Between-die**
  - Across the wafer
  - Larger length scale (~8 inch) gives rise to larger potential process-induced variation
  - Example: Thermal gradient in furnace leads to variation in $T_{ox}$ across the wafer

- **Intra-die**
  - Each device on the chip is affected differently
  - Length scale < 2 cm, magnitude of variation is smaller than between-die
  - But impact of variation is greater!
  - Example: Proximity effects where minimum pitch features exhibit different width bias than isolated features
Functional Yield: Electromigration

Migration of metal atoms in conductors which pass large DC current \textit{densities}.

Signal wires have higher immunity due to AC nature of current.

Wires must be sized so that average current density is kept below a critical value.

\[ I_{\text{signal}} = \alpha f C_{\text{load}} V_{\text{DD}} \]

Soln: Copper provides 10-100X longer lifetime at same current density vs. Aluminum (IBM).
Functional Yield: Electromigration

- Vias are susceptible to electromigration
- Current crowding effects lead to enhanced current density in certain parts of the via
- Arrays of parallel vias should be used for high current inter-level connections
Interconnect Reliability Scaling

- New low-k materials have worse thermal properties than SiO$_2$
- Global wiring is more susceptible to thermal effects (self-heating) due to larger separation from substrate
- Polyimides yield ~30% lower allowable current density in 0.1 $\mu$m global wiring
- Self-heating effects lead to worsened electromigration reliability (even in Cu) since the metal temperature is increased over the local ambient temperature

Table gives max. allowable peak current density (in MA/cm$^2$)

Ref: Banerjee, DAC99
Environmental Variations

- Temperature gradients (esp. due to large, active clock drivers) lead to changes in MOSFET drive current, wire resistance, electromigration lifetime, etc.

Ref: Compaq, ASP-DAC00
Environmental Variation: Supply Voltage

\[ \tau_d \propto \frac{C_L}{L^{0.5} \cdot T_{ox}^{0.5}} \left( \frac{V_{th}}{V_{dd}} \right)^{0.3} \left( \frac{1}{W_n} + \frac{2.2}{W_p} \right)^{1.3} \]

10% reduction in \( V_{dd} \) for \( V_{th}/V_{dd} = 0.25 \) yields 9.2% rise in delay

\( V_{th}/V_{dd} = 0.3, \) rise = 18.4%!
Environmental Variation: Supply Voltage

- Voltage changes on $V_{dd}/GND$ lines propagate to signal lines
  - may cause functional failures
  - enhanced leakage current (exponential function of input voltage)
  - reduced noise margins, more susceptible to noise glitches

Fig. courtesy of Motorola
Power Distribution: IR Drop

Grid structure yields low IR drops but wirebonding constrains power to be supplied from chip periphery

Middle of die sees large IR drops due to $D_c/2$ maximum wirelength

Top layer voltage drop is given by:

$$V_{top} = I_{top}R_{top} = J_{avg} \frac{D_c}{2} P_{top} \bullet R_{int} \frac{D_c}{2} = \frac{I_{chip}}{8} P_{top} R_{int}$$

With flip-chip, worst-case resistive path drops from $D_c/2$ to $P_{bump}$

(bump pad pitch, ~200 µm)

Compared to IBM S/390 (flip-chip), expression (max) = 32 mV, experiment (avg) = 23 mV
## Environmental Variation: Supply Voltage

<table>
<thead>
<tr>
<th>Total Vdrop due to IR:</th>
<th>Wirebond: (measured @40A)</th>
<th>Flip-chip: (estimated @40A)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pin -&gt; bond flange:</td>
<td>-150 mV</td>
<td>N/A</td>
</tr>
<tr>
<td>-&gt; chip edge:</td>
<td>-60 mV</td>
<td>-90 mV *</td>
</tr>
<tr>
<td>-&gt; chip center:</td>
<td>-110 mV</td>
<td>-35 mV</td>
</tr>
<tr>
<td>Peak AC Noise:</td>
<td>~ 75 mV</td>
<td>~ 150 mV</td>
</tr>
<tr>
<td>Total Combined Vdrop:</td>
<td>320 mV – 395 mV</td>
<td>125 mV – 275 mV</td>
</tr>
<tr>
<td>Measured Freq. Gain</td>
<td>8-10% over WB</td>
<td></td>
</tr>
</tbody>
</table>

Ref: Compaq, ISSCC00
Impact of neighboring signal activity

Intel 1GHz Coppermine – 50MHz drop in timing due to capacitive crosstalk effects

November 9, 2000
ICCAD 2000
Ref: Intel, ISSCC00
Noise Immune Layout Fabric

This layout style trades off area for:

- **Noise immunity** (both C and L)
- **Minimizes variations** (CMP)
- **Predictable**
- **Easy layout**
- **Simplifies power distribution**

**Major area penalty (>60%)**

Ref: Khatri, DAC99

November 9, 2000

ICCAD 2000
Systematic Process Variation Example

ILD thickness variation

- Function of underlying pattern density

Delay and noise strongly depend on ILD thickness

Fig from Mehrotra, DAC00
Systematic Variation

Design/route with variation as a constraint using empirical variability models

- What we should be working towards

A better approach

Layout using models

Extraction
**V\textsubscript{th} Variation**

V\textsubscript{th} variation has a (WL)	extsuperscript{-0.5} dependency

Smaller devices exhibit absolute larger variation

V\textsubscript{th} variation strongly affects gate delay, subthreshold leakage

Both of which are extremely important in modern designs

Ref: Eisele, Trans. VLSI 97
Major Manufacturing Problem: Intra-Chip $L_{gate}$ Variation

- Roadmap for Semiconductors: one of biggest challenges in lithography is $L_{gate}$ control
- Scaling worsens stepper lens aberration
  - Intra-chip $L_{gate}$ variability increased
- NTRS: optical lithography still crucial
- Need to study it
  - For modeling within CAD flow
  - For yield and performance improvement
Circuit Performance / Yield Perspective: Why to Study Intra-Chip $L_{\text{gate}}$ Variation

- The gate $L_{\text{gate}}$ impacts the yield-circuit performance tradeoff
  - Smaller $L_{\text{gate}}$ produces faster circuits
  - Smaller $L_{\text{gate}}$ produces leakier transistors

- Gate $L_{\text{gate}}$ variation degrades both circuit performance and yield

- Better models needed to predict effect of $L_{\text{gate}}$ variability on circuit performance and yield

November 9, 2000

Orshansky, ICCAD00
Gate Classification by Local Layout Patterns

- Spatial $L_{gate}$ variability depends on local layout patterns
  - Need to characterize different configurations separately

- Gates are classified by:
  - **A) Orientation**
    - (vertical vs. horizontal)
  - **B) Distance to neighbors**
    - (proximity effect)
  - **C) Left vs. right neighbor**
    - (coma effect)

Edge distances:
- $1 = $dense $\rightarrow 5 = $isolated

Orshansky, ICCAD00
Spatial $L_{gate}$ Maps for Different Gate Categories

- All spatial maps are statistically significant
- Mask-level gate $L_{gate}$ correction is feasible

November 9, 2000
Orshansky, ICCAD00
Ring Oscillator Speed Gradient

- Delay of 151-stage NAND ring oscillator simulated
- 14% speed variation across chip!
- Delay map consistent with \( L_{\text{gate}} \) maps
- Chip timing properties depend on location within field

November 9, 2000

Orshansky, ICCAD00
Spatial $L_{\text{gate}}$ Variation Directly Affects Clock Skew

- Spatial $L_{\text{gate}}$ non-uniformity impacts global chip properties, e.g. clock skew
- Simulated global skew for 16 cluster H-tree clock a 5x5mm$^2$ chip
- Maximum skew is 74ps. This is 8% of the clock cycle
- Cumulative effect of $L_{\text{gate}}$ spatial variability on circuit timing may be as high as $\sim$25%!
  - Need models to predict performance degradation early

Orshansky, ICCAD00
Back-end Variation: Motivation

• Modeling of interconnect effects is a hot topic
• Modeling the \textit{variation} of interconnect is not
  – Why?
  – If a large part of path delay is due to interconnect, uncertainty in back-end geometry translates directly to uncertainty in circuit speed
• Early in the design phase we need:
  – \textbf{Speed} (allow for design optimization, etc.)
  – \textbf{Realistic} process spreads (pessimism leads to overconstraints)
Back-End Process Variation

CMP can lead to ± 20% variation in ILD thickness

Linewidth variation greatly impacts RC line parameters

Industrial 0.35 μm process

November 9, 2000

 ICCAD 2000
Common Approaches to Back-End Statistical Modeling

- Often neglected
- Skew-corner method
  - Set all parameters to worst-case simultaneously
  - Extremely pessimistic
- Overdesign

All interconnect is viewed as nominal

\[ JP = \prod_{i=1}^{N} p(i) \]

5 independent parameters

\[ JP = 3.71 \times 10^{-15} \]
Methodology Overview

- Monte Carlo approach combined with analytical models
- Only one SPICE run is required for a given circuit configuration
- 2000 parameter sets yield <1% error in mean and <3% error in standard deviation (90% confidence)
Advantages of this Approach

- **Speed**
  - Runtime 3-4 orders of magnitude less than full simulations

- **Flexibility**
  - Designers won’t be constrained by overly pessimistic estimates of circuit performance variation
  - Works with a distribution, not single value -- allows for arbitrary performance figures to be found (2σ, 95%)

- **Accuracy**
  - Yields a much more realistic 3-σ point than skew-corner
Choosing Analytical Models: Capacitance

- Interconnect geometry of interest is shown to left
- Need to obtain electrical characteristics first (R and C)

Empirical C expressions to allow for:
- Differentiation of capacitance terms (coupling, ground)
- Distinct values of upper and lower ILD thickness
Crosstalk Noise Model

A complete model accounting for:

- Line resistance
- Finite input rise time
- Non-linear resistance of aggressor gate
Analytical Models Match Simulation

Models show good fit to 2-D simulation and SPICE

Correlation > 0.97
Fitting parameters for device resistance in delay expressions
Calculated Process Spreads

Crosstalk noise analysis

Tighter pitches meet noise constraints

Delay analysis

Skew-corner
Simulations
Analytical

$3\sigma / \text{mean} (%)$

Wasted design space

Pitch ($\mu m$)
## Results

33 to 63% reduction in performance uncertainty compared to skew-corner

<table>
<thead>
<tr>
<th>Pitch (µm)</th>
<th>% improvement over skew-corner ( (1 - \sigma_{\text{realistic}} / \sigma_{\text{skew-corner}}) )</th>
<th>Delay</th>
<th>Rise Time</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>% Error</td>
<td>SIMS</td>
</tr>
<tr>
<td>1</td>
<td>0.6</td>
<td>54.9</td>
<td>48.2</td>
</tr>
<tr>
<td>3</td>
<td>1.9</td>
<td>61.5</td>
<td>56.8</td>
</tr>
<tr>
<td>5</td>
<td>1.8</td>
<td>62.8</td>
<td>56.2</td>
</tr>
<tr>
<td>10</td>
<td>0.1</td>
<td>57.8</td>
<td>52.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Pitch (µm)</th>
<th>% Error</th>
<th>SIMS</th>
<th>Analytical</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7.2</td>
<td>39.7</td>
<td>34.8</td>
</tr>
<tr>
<td>1.4</td>
<td>7.5</td>
<td>42.8</td>
<td>35.4</td>
</tr>
<tr>
<td>1.8</td>
<td>7.0</td>
<td>46.1</td>
<td>35.6</td>
</tr>
<tr>
<td>2.2</td>
<td>7.3</td>
<td>45.3</td>
<td>32.8</td>
</tr>
</tbody>
</table>

3σ_{\text{realistic}} \hspace{1cm} 3σ_{\text{skew-corner}}
Clock Network for Skew Analysis

Buffered H-tree design

Analyzing global network only

Delay variation of buffers due to input rise time is accounted for in this analysis

Device variation (e.g. current drive) is not considered
Overly pessimistic clock skew estimation may result in significant design overhauls

- Delaying chip time to market

Additional branches in clock tree increase difference between stochastic and skew-corner approaches (latency ↑)
Another Clock Skew Analysis

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Clock Skew Coeff. [ps/1%]</th>
<th>% of Variations</th>
<th>Clock Skew Comp. [ps]</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T_{ILD}$</td>
<td>1.04</td>
<td>3%</td>
<td>3.12</td>
</tr>
<tr>
<td>$H_{int}$</td>
<td>1.04</td>
<td>3%</td>
<td>3.12</td>
</tr>
<tr>
<td>$V_T$</td>
<td>0.11</td>
<td>5%</td>
<td>0.55</td>
</tr>
<tr>
<td>$L_{eff}$</td>
<td>0.53</td>
<td>5%</td>
<td>2.65</td>
</tr>
<tr>
<td>$t_{ox}$</td>
<td>0.53</td>
<td>1.2%</td>
<td>0.64</td>
</tr>
<tr>
<td>$V_{DD}$</td>
<td>0.64</td>
<td>10%</td>
<td>6.40</td>
</tr>
<tr>
<td>$C_L$</td>
<td>0.53</td>
<td>20%</td>
<td>10.6</td>
</tr>
<tr>
<td>$T$</td>
<td>0.51</td>
<td>8%</td>
<td>4.08</td>
</tr>
<tr>
<td><strong>Internal Clock Skew [ps]</strong></td>
<td></td>
<td></td>
<td><strong>61.7</strong></td>
</tr>
<tr>
<td><strong>Total Clock Skew [ps]</strong></td>
<td></td>
<td></td>
<td><strong>92.9</strong></td>
</tr>
</tbody>
</table>

Similar H-tree structure

Internal skew – due to variation among latch locations in smallest sub-block

Global skew (due to mismatch) dominated by:
- Non-uniform clock sink distribution
- IR drop in power supply

Ref: Zarkesh-Ha, CICC99
RSF Generation

From calculated performance distributions, response surface functions (RSF) can be generated

$$X = a_0 + a_1 W + a_2 W^2 + a_3 H_1 + a_4 H_2 + a_5 T + a_6 K$$

RSF’s create a *direct link* from back-end process parameters (e.g. linewidth) to circuit performance (e.g. delay, crosstalk)
2 Uses for RSF’s

- 3-σ process corner prediction
- For a known 3-σ performance value, we can find a set of technology parameters that yield 3-σ performance
- Based on the criteria of maximum likelihood (maximize joint probability of the set)

Impact analysis
- Sensitivity is calculated as

\[
\frac{\partial X}{\partial \lambda}
\]

X is the performance metric, \( \lambda \) is a technology parameter (W,T)
- Impact is defined as product of sensitivity and process spread (6σ) as determined from manufacturing line
Linewidth variation is most significant for delay variation at small pitches, ILD at large pitches.

Line thickness has least impact on noise due to tighter process control. ILD variation just as important as linewidth in noise ($C_c/C_{total}$).
Measurement / Characterization
What’s the Point?

• Why not just simulate everything?
• While simulation is indispensable, we need measurement to **validate** the simulation tools/engines/models
  – New relevant phenomena (e.g. inductance)
  – Process variability
  – New materials
Modeling Assumptions

- Interconnect capacitance measurement technique (CBCM)
- Check approximation of dense array as ground plane
  - Difference b/w CBCM & 3D due to systematic process variation (pattern density) not considered in sims
Statistical Variation

- Dedicated test chips can be used to extract systematic back-end process variation
  - Due to CMP dishing, erosion, pattern density, CD proximity effect
- Pattern dependence drives these variations
- **Models must be built from empirical data**
- ILD variation due to underlying pattern density is biggest effect

Nakagawa, IEDM97
Statistical Variation

Closed-form expressions accounting for systematic variation are useful for:

- Setting constraints on routers to ensure manufacturability
- Calibration of full-chip RC extraction
- Provide more realistic bounds on 3-σ RC variation

$$T = T_0 + K_t(D - D_0)$$

Ref: Nakagawa, IEDM97
Noise Estimation & Measurement

- Delay $\Rightarrow$ Noise
- Early noise estimation will require simple models expressed in terms of known quantities
- These models will need to be silicon-verified

![Graph showing $V_{\text{max}} / V_{\text{dd}}$ vs. Relative Victim Driver Strength]

Aggressor Strength = 7
Metal 4, L = 6 mm
Minimum Pitch (2.4 $\mu$m)

November 9, 2000  ICCAD 2000
Measuring Clock Skew with Flip-chip packaging

E-beam probing from front side

Laser probing from back side

Ref: Intel, ISSCC00
Measuring Clock Skew

Intel – optical probing through back-side of flip-chip microprocessor

<10 ps resolution
fast

IBM – PICA, collect infrared light emitted from MOS drains and rebuild waveforms

< 10 ps resolution
slow

Itanium clock waveforms
Measuring Clock Skew

Clock latency from central driver to local re-generators

Skew between intermediate clock buffers

PICA, IBM, ISSCC99
Waveform Sampling

0.18 μm microprocessor waveforms

Helpful for:

- Silicon debug
- Parametric yield enhancement
Field Solvers/3D Effects

- Tuning 3D field solvers to actual capacitance values is needed due to cladding, trapezoidal conductors, rounded edges, etc.
- More accuracy in 3D cell characterization yields better:
  - Look-up tables for full-chip 2D extraction
  - Pre-layout capacitance estimates (wireload models) -- reduce guardbanding
Charge-Based Capacitance Measurement (CBCM)

- Ultra fine resolution (< 0.1 fF)
- Measurement setup is easy -- DC ammeter
- Totally on-chip
- No reference capacitors needed
- Results can be used:
  - to check validity of 2-D & 3-D interconnect simulators
  - in circuit layout tool for extraction
  - process monitoring (small size allows for scribe line implementation)
Test Structure Design

(1) \[ I = (C + C') V_{dd} f \]

(2) \[ I' = C' V_{dd} f \]

\[ I_{net} = I - I' \]

\[ I_{net} = C V_{dd} f \]

- Device mismatch: \( C' \) of (1) and (2) are not perfectly equal
- Mismatch could result from \( C_{gs} \) and \( C_{gd} \) (small \( W \) -> small error)

November 9, 2000

ICCAD 2000
Resolution

- Measurement at different frequencies and supply voltages allows for robust self-checking extraction.

- A 2.25 $\mu$m$^2$ metal 2 to metal 1 overlap was measured at 0.44 fF

November 9, 2000

ICCAD 2000
Variation

- 3-level metal, 0.5 µm process, 25 dies

- Capacitance variation can be at least as large as $I_{dsat}$ variation

- Higher level metals typically exhibit larger variation in capacitances

<table>
<thead>
<tr>
<th>Sigma / mean $(\sigma/\mu)$</th>
<th>M1 / poly</th>
<th>M2 / M1</th>
<th>M3 / M2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C_{area}$</td>
<td>1.9%</td>
<td>6.9%</td>
<td>3.9%</td>
</tr>
<tr>
<td>$C_{fringe}$</td>
<td>2.7%</td>
<td>4.3%</td>
<td>16%</td>
</tr>
</tbody>
</table>
Interconnect Dominated Ring Oscillator (IDROS)

- Implement CBCM measurement data into layout extraction tool
- Speed accurate to within 6% of measured data
- 68% of speed variation was due to capacitance fluctuation
- Only 32% due to device variation
- These numbers are highly layout dependent

31 Stages
\[ W_n = 10.2 \, \mu m \]
\[ W_p = 19.8 \, \mu m \]
\[ L_{eff} = 0.5 \, \mu m \]

Metal 2
\[ W = 1.8 \, \mu m \]
\[ S = 1.8 \, \mu m \]

Metal 1
Inductance/High Speed Effects

- Frequency dependent
  - Current return paths
  - RLC line parameters
- Inductance is not length scalable ($l \log l$)
- No elegant way to measure on-chip inductances (akin to CBCM)
- TDR and S-parameter are difficult to translate to real circuit performance
  - Don’t reflect actual on-chip environment

Kleveland, IEDM99

L(f), L(geometry)

How to estimate pre-layout?

Need better measurement techniques for model development (noise, delay)
Inductance rising with increased distance to return path

This test structure includes other potential return paths so L is reduced over “conventional” value.
As $f$ rises, impedance is inductance dominated.

Nearest return path is sought, reducing $L$.

MOSFET driver state makes little difference.
Translation to Circuit Performance

RLC parameters are used in simulation to extract expected delay/waveform behavior.
Indirect RLC performance measurement

When $V_{dd}$ is increased, the “ledge” will occur at different % to $V_{dd}$

A ring oscillator with loads connected at beginning of line sees the source waveform

At a certain $V_{dd}$, the oscillation freq. jumps
Direct Circuit/Interconnect Performance Measurement

• Increased coupling noise in DSM era:
  – higher operating frequency of LSI
  – larger interconnect coupling due to scaling
• Accuracy of the interconnect models, tools, and design guidelines are critical
• Need to keep models calibrated and updated
• Focus:
  – Peak height: reliability, dynamic logic
  – Impact on delay: timing margin and design rules
• **Direct** measurement techniques maximize accuracy
Prior Work and Improvements

Time domain technique using Wide-Range-Comparator (WRC)

Includes:

- Novel and accurate method of finding the noise peak
- Characterization of the noise impact on delay
- Variable arrival times of aggressor/victim signals
- Effect of driver strengths of aggressor/victim lines

November 9, 2000

Soumyanath, VLSI98, and Sato, ISSCC00
Peak Noise Measurement Structure

Variable impedance driver

Vin
Ain
Vref
mode

Aen 3
Ven 3

aggressor-1
dummy line

victim
aggressor-2
dummy line

Vref

out2
Accurate Peak Detector

Cascaded low-gain differential amplifiers

Level Shifter

Latch

Reset

Vdd

OUT

Reset

vf

Vref

OUT
APD Sensitivity

\[ \Delta V = V_{\text{ref}} - V_{\text{peak}} \text{(mV)} \]

-50
-100
-150
0
50
100
150

Pulse Base Width Tw (ns)

Conventional (WRC)

Proposed (APD)

Noise

\( T_c = 100 \text{ ps} \)

\( 800 \text{ ps} \)

\( 700 \text{ ps} \)

\( T_c = 100 \text{ ps} \)

Successful peak detection
RC vs. RLC noise

- The APD concept is still valid
- Will miss the *shape* of the waveform though (ringing)
The long-range coupling of L and locality of C help isolate their effects

Test circuits exploiting this can attempt to directly measure L noise (using APD or similar concept)
Impact of Coupling Noise on Delay

Ain

Vin

Timing difference

Victim line waveform at far-end

Noise

Delay change
Delay-Change Measurement Structure

Variable impedance driver

Vin

Ain

Vref

mode

out0

out1

out2

Ven

Aen

3

3

dummy line

aggressor-1

victim

aggressor-2

dummy line

buffer

Vref

WRC

WRC

WRC

APD
Noise Peak Measurement

![Graph showing Vpeak/Vdd (%) vs Victim driver NMOS Transistor size Wn (µm)]

- Measurement
- Simulation

- W=1.2µm, S=1.2µm
- W=2.4µm, S=1.2µm

- Metal 4, 6mm

Victim driver NMOS Transistor size Wn (µm)
Delay-Change Curve (DCC)

- Aggressor/Victim Timing Difference (ns)
  - Out-phase: Victim $W_n=20\ \mu m$
  - In-phase: Aggressor $W_n=70\ \mu m$

- Measurement
- Simulation

- Delay change due to aggressor (ps)
  - Noise base width

- Aggressor/Victim Timing Difference (ns)
  - $M3, 3mm, W=S=0.6\ \mu m$
  - $10\ \mu m$
Measured Data: Bus-line Example

![Graph showing measured data]

- Peak-to-peak delay change (ps)
- Noise base width (ns)
- Aggressor and victim driver NMOS transistor size $W_n$ ($\mu$m)
- Metal 4, 3mm
- $W=S=1.2 \mu$m
Applying DCC

- Fundamental relationship between DCC and crosstalk noise waveform
Objective

Bi-directional transformation between

Delay Change Curve (DCC)  Victim waveform

[Diagram showing the relationship between Agg./Vic. timing difference and delay change, as well as the transformation to the victim waveform over time.]
Victim Waveform Model

\[ g(t) = \begin{cases} 
0 & (t<0) \\
V_{dd} \{1-\exp(-t/\tau_r)\} & (0\leq t) 
\end{cases} \]  \hspace{1cm} (1a)

\[ f(t,k) = \begin{cases} 
0 & (t-k<0) \\
V_p/t_a(t-k) & (0\leq t-k<t_a) \\
V_p \exp\{-(t-k-t_a)/\tau_d\} & (t_a\leq t-k) 
\end{cases} \]  \hspace{1cm} (2a)

(1b) and (2b) and (2c)
Derivation of the DCC Equation

Assumption: the victim waveform with noise is sum of without-noise victim waveform and the coupling noise

F. Dartu et. al. Design Automation Conference, 1997

DCC can be analytically calculated by expanding exponential function around t=t0

\[ \Delta t = \begin{cases} 0 & \text{(3a)} \\ \frac{1}{\tau_d} (k-t_0) & \text{(3b)} \\ \frac{\tau_d}{2 - \frac{\tau_d}{\tau_r} \exp\left\{\frac{(k-t_0+\tau_d)}{\tau_d}\right\}} & \text{(3c)} \end{cases} \]

\[ \Delta t_{peak} = \tau_r \ln(2\rho+1) \quad \text{(4)} \]
Waveform Reconstruction Procedure

x-intercept of Eq.(3b)

curve fit on Eq.(3c)

max. delay change

slope of Eq.(3b)

\[ t_a = \frac{2 \tau_r}{\rho} \frac{1 - \Delta t'}{\Delta t'} \]

\[ V_p = \frac{V_{dd}}{2} \{ \exp(\frac{-\Delta t_{peak}}{\tau_r}) - 1 \} \]

\[ \tau_r = t_0 / \ln 2 \]

Delay Change Curve (DCC)  Victim waveform
Example Applications

1. Victim waveform reconstruction using DCC
   - Indirect measurement technique of the coupling noise

   ![DCC Diagram](image1)

   ![Victim Waveform Diagram](image2)

   Eq.(3)

   Eq.(4)

   Eq.(1) and (2)

2. Accurate & fast estimation of DCC
   - Noise aware Static Timing Analysis (STA)
Waveform Reconstruction (1)

- Simulated
- Extracted from DCC

Delay change (ns)

Timing difference (ns)

Vic. Wn = 14.4 (um)
Agg. Wn = 33.6 (um)

W = S = 0.32 (um)
Length = 2 (mm)

V_p = 0.86 (V)

\( t_a = 130 \) (ps)

\( \tau_d = 250 \) (ps)

\( \tau_r = 220 \) (ps)

Victim voltage (V)

1 ns

Coupling noise

Without-noise victim waveform
Waveform reconstruction (2)

- Simulated (x2 Rsheet)
- Extracted from DCC

Vic. Wn = 4.8 (um)
Agg. Wn = 33.6 (um)

- Vp = 0.78 (V)
- t_a = 310 (ps)
- \( \tau_d = 600 \) (ps)
- \( \tau_r = 580 \) (ps)

Delay change (ns)
Timing difference (ns)
Voltage (V)
DCC Estimation

Metal-1, 2mm

38 SPICE simulations

1 SPICE simulation, parameter fitting, and Eq. (3).

Aggressor/Victim timing difference $k$ (ns)

Victim delay change (ns)

Eq. (4)
### Speed and Accuracy

<table>
<thead>
<tr>
<th></th>
<th>Conventional</th>
<th>Proposed</th>
</tr>
</thead>
<tbody>
<tr>
<td># simulations</td>
<td>38</td>
<td>1</td>
</tr>
<tr>
<td>Total CPU time</td>
<td>1142 (sec)</td>
<td>38 (sec)</td>
</tr>
<tr>
<td>Speedup</td>
<td>1</td>
<td>30.9 x</td>
</tr>
<tr>
<td>DCC peak w/Eq.(3)</td>
<td>-</td>
<td>16 %</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>w/Eq.(3),(4)</td>
<td>-</td>
<td>0.4 %</td>
</tr>
<tr>
<td>DCC half max. width</td>
<td>-</td>
<td>6 %</td>
</tr>
</tbody>
</table>

The use of analytical equation may further reduce the calculation time.
Bibliography

- B. Kleveland et al., “Line inductance extraction and modeling in a real chip with power grid ,” IEDM, 1999.
Bibliography

Bibliography
