Physical Synthesis 2.0

Andrew B. Kahng
UCSD CSE and ECE Departments

abk@ucsd.edu
http://vlsicad.ucsd.edu
Concept: “Design Principles”

- Partition the problem \(\rightarrow\) divide and conquer, hierarchy
  - Different abstraction levels: RT-level, gate-level, switch-level, transistor-level

- Orthogonalize concerns
  - Function vs. implementation
  - Logic vs. timing vs. embedding

- Solve chicken-egg conundrums

- Constrain the design space to simplify the design process
  - Balance between design complexity and performance
  - E.g., standard-cell methodology
  - \(\rightarrow\) “freedom from choice”
Flow expands in two directions
  - System-Level Design
  - Design for Manufacturability (DFM)

More design care-abouts
  - Area, Timing, Power, Signal Integrity, Reliability, Cost

Key challenges: loops, chicken-egg
  - “Design closure” through tight integrations
  - RTL, GDSII “signoffs” = business structure of semiconductor creation

“One-pass flow”: required for Productivity, requires Predictability
  - By Guardbands?
  - By “Unifications”?
  - By Statistics?
  - By Methodology (to avoid issues)?
Outline

• Why Physical Synthesis
• Physical Synthesis 1.0
• Example Challenges / Stressors
  • FinFET
  • Noise and Chaos
  • Clock Skew
  • Complexity and Hyperlocality
  • Better (and, more complex) Signoff
  • New Mixed-Height Sweet Spot?
• Physical Synthesis 2.0?
Logic Design Needs Spatial Information

- High aspect ratio floorplan: shift one macro block from left to right, and vary its shape (with constant area)
- 10% power range (post-route): center location, taller blockage = more power, more contribution of wire (delays)
- Separation of logical, temporal, spatial must crumble

Shift the location of blockage

Macro size
- 260µm x 65µm
- 184µm x 92µm
How Do We Predict Spatial Information?

- Predict by modeling
  - Machine learning, regression, etc.
  - *(Don’t dismiss this!)*
    - [DAC00] http://vlsicad.ucsd.edu/Publications/Conferences/112/c112.pdf

- Predict by assuming and enforcing
  - Make a prediction, then make the prediction come true
  - (Constant-delay methodology)

- Predict by doing
  - Constructive prediction
  - (Run under the hood – quick and dirty, else no leverage)
Outline

• Why Physical Synthesis

• Physical Synthesis 1.0

• Example Challenges
  • FinFET
  • Noise and Chaos
  • Clock Skew
  • Complexity and Hyperlocality
  • Better (and, more complex) Signoff
  • New Mixed-Height Sweet Spot?

• Physical Synthesis 2.0?
Synthesis vs. Physical Synthesis

- Synthesis (DC, RC)
  - Elaboration, mapping to generic gates
  - Clock gating
  - Apply timing constraints, remap / optimize
  - Multibit FF optimization
  - MBIST insertion
  - Scan chain stitching
  - Further optimization, area recovery

- Physical Synthesis (DCT/DCG, RCP)
  - LEF list
  - Tech file, map file
  - tluplus_{max,min}
  - floorplan DEF
  - {min,max}_routing_layer
Physical Synthesis

• In
  • RTL + SDC + Library models + Floorplan DEF

• Out
  • Better netlist (usually), at one (worst) corner
  • Better netlist (usually) + placed DEF (not legalized)
  • **N.B.: very fast TAT required by customers**

• **Netlist (+ placed DEF) is passed to P&R + signoff**
  • Place, placeOpt, CTS, CTSOpt, route, routeOpt, leakage recovery, timing closure
  • **Different companies and tools in a long tool chain**
**Example**

**Physical Synthesis**

- Floorplan Specified by designers
- Netlist + initial placement
- Floorplan in DEF or physical guidance
- P&R flow
- Routed Results

Floorplan information:

- Placement region for standard cells
- Blockage
- RC tech file (tluplus, captable)
- Libraries, LEF, tech files
- e.g., DCT (Physical Synthesis)
Note: “P&R + Signoff” is Complicated!


TOP-LEVEL NETLIST / SPEF
BLOCK-LEVEL NETLIST / SPEF

Static Timing Analysis for all Modes / Corners

Breakdown of Timing Violations on per Block Basis

Manual Repair of Timing Failures

Timing Closed

Operations Permitted at Each Iteration (in order of preference)
1. Vt Swap, Resizing, Buffer Insertion, NDR Changes, Useful Skew
2. Vt Swap, Resizing, Buffer Insertion, NDR Changes
3. Vt Swap, Resizing, Buffer Insertion
4. Vt Swap, Resizing
5. Vt Swap

Violation Classes Addressed for Each Iteration (in order of priority)
1. Electrical Rule Violations
2. Noise Violations
3. Setup Violations
4. Hold Violations
Since That Article Was Written:

- 90nm
  - Temp inversion
  - Noise
  - MCMM

- 65nm
  - Maxtrans
  - PBA
  - EM

- 45/40nm
  - AOCV / POCV

- 28nm
  - Dynamic IR
  - Fixed-margin spec
  - Phys-aware timing ECO

- 20nm
  - Multi-patterning
  - MOL
  - BEOL R

- 16/14nm
  - Cell-POCV
  - Min implant
  - MOL, BEOL R

- 10nm ≤ 7nm
  - LVF
  - BEOL, MOL variations
  - Signoff criteria with AVS
  - SOC complexity
  - Fill effects
  - Layout rules

[DAC15]

A. B. Kahng, Physical Synthesis 2.0, IWLS-2015 Keynote
How Can Physical Synthesis Possibly Work?

• “If it sounds too good to be true, it usually is …”

• What do we do with constraints at (physical) synthesis stage?
  • Overconstrain the clock period in synthesis (was by 20%, now by ~10%)
  • Utilization: 60% target in synthesis (sometimes 50%, 55%) \( \rightarrow \) 85+% post-placement

• Which detailed placer, CTS tool, router, optimizer?
  • Complex tool “sensitivities” (noisy, chaotic behavior)
  • Information that is ignored (advanced manufacturing)
  • Information that is never available (CTS, SI)

• What explains “success”? Guardbands, low expectations…?
  • Designers’ preoccupation with area and schedule helps…
Challenges

• FinFET, BEOL scaling effects
  • Drive
  • Resistivity
  • Gate-wire balance

• Clock effects
  • Skew across corners
  • Top-level clock distribution (CGCs, muxes, dividers, …)
  • Useful skews = area vs. delay tradeoffs

• “Extreme localization” effects
  • Advanced (multi-)patterning
  • Pin access, congestion, coupling
  • Breakdown of placement-optimization separation
Questions

• If Logic Synthesis can’t know outcomes at end of Physical Design, can it be doing the right thing? *(Simple information arguments) (What margin is left on the table? Are we seeing placebo effects (association vs. causation etc.)?)*

• Can Logic Synthesis be made better aware of future Physical Design outcomes?

• Is Logic Synthesis at risk of being eclipsed by Physical Design? *(Venus-Mars → Sun-Moon, etc.)*
Outline

• Why Physical Synthesis
• Physical Synthesis 1.0
• **Example Challenges**
  • FinFET
  • Noise and Chaos
  • Clock Skew
  • Complexity and Hyperlocality
  • Better (and, more complex) Signoff
  • A Mixed-Height Sweet Spot?
• Physical Synthesis 2.0?
FinFET: Current Density + Discreteness

- Better electrostatic control + continued gate length scaling
  - Drive current $\uparrow \Rightarrow$ cell height $\downarrow$ (e.g., 8.25T), better area density (w/ fin height $\uparrow$)
  - Effective width $\approx 1.6x$ equivalent area with planar devices

- Current density $\uparrow$, plus fin discreteness challenges
**FinFET: Aggressive Voltage Scaling**

- FinFET enables voltage scaling for reduced dynamic power
  - Better electrostatic control $\rightarrow$ better performance at low supply voltage
- **High-performance mode**: wire-dominated
- **Low-performance mode**: gate-dominated

![Graph showing delay vs. V_{DD}](image)

Gate-Wire Balancing

- Unbalanced gate-wire delay causes **severe** delay variation on data and clock paths across modes
- Delay variation in clock paths == skew variation
  => Increased difficulty for timing closure ("ping-pong effect")
- Minimization of skew variation is important for timing closure
  (Our work at DAC15 uses global-local optimization achieves 22% skew variation reduction)

Latency skew

<table>
<thead>
<tr>
<th>Corner</th>
<th>Clock latency</th>
<th>Skew</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Launch</td>
<td>Capture</td>
</tr>
<tr>
<td>SS, 0.7V, -25°C</td>
<td>1.0</td>
<td>1.1</td>
</tr>
<tr>
<td>FF, 1.1V, -25°C</td>
<td>0.9</td>
<td>0.7</td>
</tr>
</tbody>
</table>

Low voltage: gate delay dominates
High voltage: wire delay dominates
=> Skew reversal
=> Power/area overheads
FinFET: Less Body Effect, Richer Libraries?

- FinFET 4-input NAND ~ planar bulk 3-input NAND
- More complex cells / higher fan-in cells could be made available to synthesis

Number of fan-in limited by body effect
Pin Accessibility Below 20nm

- Routing challenged by complex rules for multi-patterning
- Limited pin access with small track cells
  - Wider power rail for reliable connection
  - Fewer pin access points
- Complex design rules

Pin accessibility problem

$\Rightarrow$ conflict between area reduction and routability

[21 A. B. Kahng, Physical Synthesis 2.0, IWLS-2015 Keynote]
Outline

• Why Physical Synthesis
• Physical Synthesis 1.0
• Example Challenges
  • FinFET
  • Noise and Chaos
  • Clock Skew
  • Complexity and Hyperlocality
  • Better (and, more complex) Signoff
  • New Mixed-Height Sweet Spot?
• Physical Synthesis 2.0 ?

Slack vs. Layout Context

- Layout knobs: SRAM pitches and buffer keepout distances
- Post-P&R slacks of five embedded memories is “chaotic”
- Physical synthesis challenge: Logic optimization given “chaos”

Testcase: Logic from OpenCores GPU THEIA + SRAMs
Slack vs. Clock Period

- $\Delta$ path slack is 81ps at signoff clock period of 1.0ns
- Changing clock period to 0.82ns changes $\Delta$ path slack to 143ps!
Non-SI vs. SI

- Top-1000 critical paths from Viterbi design (clock period = 1.0ns)
- Slack diverges by 81ps !!! ~4 stages of logic at 28nm FDSOI
- Unfortunately, we don’t know coupling before routing !!!

![Graph showing Path Slack in Non-SI Mode vs. Path Slack in SI Mode](image)

81ps

Ideal correlation
WLM, RC (Interconnect proxy) Effects

- Example: SOCE-based “Shrunk2D” (S2D) flow [1]
- Perform synthesis with different WLM caps, P&R with S2D flow
- Shown: total power (#buffers, #instances, instance area, WL, … similar)

Outline

• Why Physical Synthesis
• Physical Synthesis 1.0
• **Example Challenges**
  • FinFET
  • Noise and Chaos
  • **Clock Skew**
  • Complexity and Hyperlocality
  • Better (and, more complex) Signoff
  • A Mixed-Height Sweet Spot?
• Physical Synthesis 2.0 ?
Sensitivity of CTS Outcomes to Layout Contexts

- Delay varies by up to **43%** with clock entry point locations
- Delay varies by up to **45%** with core aspect ratio
- NDRs, fill, buffer sizes, max fanout / max trans rules, …
  \[\Rightarrow\] **100ps impacts on insertion delays, skew, slacks**

A. B. Kahng, Physical Synthesis 2.0, IWLS-2015 Keynote
Useful Skew Improves Timing

- Useful skew optimization adjusts clock sink latencies to improve timing
- Our predictive useful skew flow resolves the “chicken-and-egg loop” → further improved timing

Zero skew

Useful skew

Delay/Slack Clock latency

Useful skew improves timing

6 testcases {3 RTLs x 2 clock periods}
Conventional Useful Skew Optimization

- Standard useful skew flow has **chicken-egg problem**

  - Netlist and placement assume zero skew
  - Useful skew optimization relies on placement

- One solution: Back-annotation flows (large runtime)

  - RTL netlist
  - Synthesis
  - Placement / Place Opt.
  - CTS
  - CTS Opt.
  - Routing / Route Opt.

Wang et al. in DAC06 propose to back-annotate useful skew from post-placement to before-synthesis

A. B. Kahng, Physical Synthesis 2.0, IWLS-2015 Keynote
NOLO: No-Loop Useful Skew Optimization

- Our work: Cure the chicken-egg problem with delay prediction

- Use setup slacks from LVT-only synthesis
  ⇒ estimation of achievable slacks

- Use hold slacks from multi-VT synthesis
  ⇒ reduce pessimism

- Advantage: One-pass approach, not constrained by placement
Experimental Results

- Predictive flow achieve similar or better timing and much smaller runtime
Outline

• Why Physical Synthesis
• Physical Synthesis 1.0

• Example Challenges
  • FinFET
  • Noise and Chaos
  • Clock Skew
  • Complexity and Hyperlocality
  • Better (and, more complex) Signoff
  • A Mixed-Height Sweet Spot?

• Physical Synthesis 2.0?
BEOL Multi-Patterning Impacts

Wire1\textwidth = M\textwidth

Wire2\textwidth = M\textwidth - 2\times S\textwidth

Line-end extensions

Floating fill wires

Line-end cuts

\sigma_{\text{M}}

\sigma_{\text{S}}

0.5\sigma_{\text{M}}

block
Placement-Sizing Interference

- New “interferences” between post-layout optimization and P&R
- Rules for device layers (FEOL) become considerably more complex and restrictive
  - Minimum implant width rules for implant region
  - Minimum notch and jog width rule for oxide diffusion (OD)
Placement-Sizing Interference (cont.)

- Drain-to-drain abutment (DDA)

- Example solution

Intertwine the historically separate tasks of P&R and post-route optimization
Outline

• Why Physical Synthesis
• Physical Synthesis 1.0
• Example Challenges
  • FinFET
  • Noise and Chaos
  • Clock Skew
  • Complexity and Hyperlocality
  • Better (and, more complex) Signoff
  • A Mixed-Height Sweet Spot?
• Physical Synthesis 2.0?
I. Flexible Timing Models

• Setup time, hold time and clock-to-q (c2q) delay of FF ⇒ values interdependent, but NOT fixed

• Flexible FF timing model can exploit operating (function/test) modes ⇒ “Free” pessimism reduction in STA

• **Goal**: Find best \{setup, hold, c2q\} for each FF instance

• Sequential LP:
  • setup-c2q opt
  • hold-c2q opt

[ISQED14]
Flexible Timing Model → Recover Margin

- Independent datapaths in PBA: using fixed FF timing model loses performance optimization opportunity
Improved Timing Signoff Flow

Takeaways
- Fix timing violations “for free”
- 48ps average improvement of slack over 5 designs in a foundry 65nm technology

Next
- Better exploitation of disjoint cycles/modes
- More accurate modeling of setup-hold-c2q tradeoff
- Circuit optimization should natively exploit FF timing model flexibility

A. B. Kahng, Physical Synthesis 2.0, IWLS-2015 Keynote
II. Signoff Definition (e.g., with AVS, Aging)

- $V_{\text{BTI}}$: Voltage for BTI-aging estimation
- $V_{\text{lib}}$: Supply voltage for timing library characterization
- $V_{\text{final}}$: $V_{\text{dd}}$ of a circuit with AVS at end-of-lifetime

**Chicken & Egg Loop**

- $V_{\text{BTI}}$ and $V_{\text{lib}}$ depend on aging during AVS ($V_{\text{final}}$)
- $V_{\text{final}}$ depends on circuit
- Circuit implementation depends on $V_{\text{BTI}}$ and $V_{\text{lib}}$
Observations and Heuristics

Observation #1: $V_{\text{final}}$ is not sensitive to cells along the timing-critical path.

Observation #2: $\Delta V_t$ with a constant $V_{\text{final}}$ throughout lifetime $\approx$ adaptive $V_{\text{dd}}$.

Heuristic #1: Use average of critical path replicas to estimate $V_{\text{final}}$ ($V_{\text{heur}}$).

Heuristic #2: approximate $V_{\text{dd}}$ in AVS by constant $V_{\text{heur}}$.

Solve “Chicken & Egg Loop” by having $V_{\text{BTI}} = V_{\text{lib}} = V_{\text{heur}} \approx V_{\text{final}}$.

A. B. Kahng, Physical Synthesis 2.0, IWLS-2015 Keynote
“Knee” Point for Signoff Definition

Optimistic aging library → large power penalty

Overly pessimistic aging library → large area penalty

Our method finds “Knee” point for balanced area and power tradeoff

Experiment setup:
DC/AC BTI @ 125°C
32nm PTM technology
4 benchmark circuit implementations

A. B. Kahng, Physical Synthesis 2.0, IWLS-2015 Keynote
Outline

• Why Physical Synthesis
• Physical Synthesis 1.0
• Example Challenges
  • FinFET
  • Noise and Chaos
  • Clock Skew
  • Complexity and Hyperlocality
  • Better (and, more complex) Signoff
  • A Mixed-Height Sweet Spot?
• Physical Synthesis 2.0?
Mixed Cell Height Implementation (!)

- Large cell height $\rightarrow$ better timing, but large area and power
- Small cell height $\rightarrow$ smaller area/power per gate, but large delay and more #buffers
- Mixing cell height enables tradeoffs between performance and area/power (recall FinFET introduction!) $\rightarrow$ better design QoR
  - E.g., use large-height high-fanin cells to improve pin accessibility
  - Already have flop trays, etc. as problematic multi-height instances

Technology: 28nm LP
In red are 12T cells = larger area, smaller delay
In blue are 8T cells = smaller area, larger delay
Cost of Mixing Cell Heights

- “Breaker cells” are required to align regions with different cell heights
  → Optimization must comprehend corresponding area cost

Assume: M2 pitch = 64nm

Y directional shift
one M2 pitch

64nm

48nm

64nm

X directional shift
four sites

No routing blockage
Routing blockage on M1/M2
Optimization Flow

- Initial placement uses modified LEF → enable optimization with a conventional flow
- Slicing-based partition with DP to divide die area into regions with different cell heights
- Internal-timer guided placement legalization
- Floorplan update with “breaker cell” penalty
- Row-based cell mapping places cells onto rows with corresponding heights
Example of Optimization Flow

Initial placement
(8T/12T cells are “freely” placed)

Partitioning
(Yellow blocks = regions)

Legalization

Technology: 28nm LP
Design: AES
8T cells are in blue
12T cells are in red

Mixed-height placement

New floorplan

A. B. Kahng, Physical Synthesis 2.0, IWLS-2015 Keynote
Benefits from Mixing Cell Heights

- Technology: 28nm LP (12T/8T)  Design: AES
- 25% area reduction as compared to 12T-only design
- 20% performance improvement compared to 8T-only design
Outline

• Why Physical Synthesis
• Physical Synthesis 1.0
• Example Challenges
  • FinFET
  • Noise and Chaos
  • Clock Skew
  • Complexity and Hyperlocality
  • Better (and, more complex) Signoff
  • A Mixed-Height Sweet Spot?
• Physical Synthesis 2.0 ?
Physical Synthesis 2.0

- It’s the predictability! (and, prediction is challenged…)
  - New devices and patterning technologies
  - Complex PD tool chain; chaotic behavior of tools and flows
  - Oblivious to clocks, corners, coupling → how can Physical Synthesis be doing the right thing? (= target for margin recovery!)

- What will Physical Synthesis 2.0 look like?
  - (1) Higher-level value: what Physical Design cannot do
    - Datapath architecture selection
    - Resource sharing
    - Mux mapping
  - (2) Other types of prediction (machine learning, big data, etc.)!
  - (3) Constructive prediction deeper into implementation flow
    - (More integration… 😊) Clock and MCMM awareness
    - Hyperlocality awareness: coloring, congestion, coupling, interactions …
THANK YOU!