Body Bias Clustering for Low Test-Cost Post-Silicon Tuning

Shuta Kimura, Masanori Hashimoto, Takao Onoye
Dept. Information Systems Engineering, Osaka University
hasimoto@ist.osaka-u.ac.jp
Outline

• Background & objective
• Proposed body bias clustering
• Experiments
• Conclusion
Backgrounds

- Larger variation in delay and leakage
- Design-time efforts insufficient

Post-fabrication performance compensation is demanded.

Body biasing

Delay and leakage can be adjusted by altering body voltages.

For body biasing, large well separation is necessary.
Granularity of body biasing

Chip

Block/IP

Clustered gates\footnote{1}

Gate

Overhead for well separation

Granularity

Negligible  <  A few %  <  Acceptable  <<  Impractical

Compensable variability

Die-to-die

Within-chip spatial

Random

Gates insensitive to circuit delay are also biased.

⇒ Wasteful increase in leakage

Only timing-critical gates are highly biased

⇒ Min. increase in leakage

Low test cost post-silicon tuning

- [Conventional] each cluster can freely take a bias voltage.

  For all configurations, delay test and leakage measurement are necessary.

  ![Cluster Diagram]

  - Low voltage (slow, small leak)
  - High voltage (fast, large leak)

  Large tuning cost

- [Low cost\(^1\)] only configurations w/ monotonic increase in high voltage clusters

  Test starts lowest level and continues until timing satisfied.
  No leakage measurement necessary

  Large reduction in tuning cost

Limitations of previous work

• Previous work\(^1\) focused on subthreshold circuits
  – No die-to-die variations are considered assuming random Vth variation is dominant.
  – Thus, no correlation between delay and leakage is considered.
  – Only two bias voltages usable.

\(^1\) K. Hamamoto, M. Hashimoto, Y. Mitsuyama, and T. Onoye, “Tuning-friendly body bias clustering for compensating random variability in subthreshold circuits,” ISLPED’09
Objective

➢ To develop a body bias clustering method for low test-cost post-silicon

  – Extension to more than 3 body voltages
  – Consideration of D2D, spatially-correlated and random variations
  – Consideration of correlation between delay and leakage
  – Design implications on how to synthesize a circuit to which post-silicon tuning will be applied.
Outline

• Background & objective
• Proposed body bias clustering
• Experiments
• Conclusion
Extensions to more than three bias voltages

Cluster-first alternation

Voltage-first alternation

Monotonicity in delay and leakage preserved.
Maximum # of testable levels limited

Increase in # of body voltages may result in
(# of possible levels) > (max. # of testable levels)

Need to pick up during optimization
Body voltage selection

Body voltage generator can often produce several voltages, but their on-chip distributions are expensive.

Need to select small # of voltages

Select voltages used for body biasing during optimization
Problem formulation

- **Inputs**
  - a placed circuit
  - constraints (delay, yield, #clusters, max. # of tests, # of usable body voltages)

- **Outputs**
  - cluster organization
  - a set of body voltages
  - a set of configurations for test

- **Objective function**

\[
\text{(Avg. leakage after post-silicon tuning)} \times \text{(penalty)}
\]

Becomes large when yield const. not satisfied.
Clustering procedure

1. Placed circuit
2. Divide circuit into islands
3. Generate an initial solution
4. Evaluate obj. function
5. Generate a neighbor solution
6. Evaluate obj. function
7. Better solution?
   - Yes: Record current solution
   - No: Divide circuit into islands
8. SA finish?
   - Yes: A solution
   - No: Ordinary simulated annealing
Leakage current after post-silicon tuning

\[
\text{leakage} = \sum_{i=1}^{\text{level}} \int_{\text{cond.}} \mu_i(x) \, dx
\]

**Leakage at level** \(i\)

Variation parameters

Integration is carried out in subspace of \(X\) where level \(i\) is selected.

\[(d^{(i)}(X) < d < d^{(i-1)}(X)) \quad (i > 1)\]

**Delay at level** \(i\)
Leakage computation w/ Monte Carlo integration

Virtually fabricate many chips and simulate tuning

(1) Derive canonical forms of circuit delay and leakage w/ same random variables

\[ D^{(i)}(X) = \ldots \] [1]
\[ L^{(i)}(X) = \ldots \] [2]

(2) Virtually fabricate many chips (generate many \(X\)s randomly)

(3) Simulate post-silicon tuning

(4) Compute avg. leakage

Outline

• Background & objective
• Proposed body bias clustering
• Experiments
• Conclusion
Performance improvement w/ proposed method

Effective for pushing performance w/ smaller leakage increase

64-bit multiplier synthesized for minimizing delay
65nm process
\( \sigma_{\text{global}} = 25\text{mV}, \sigma_{\text{random}} = 15\text{mV} \)

#islands = 16
#clusters = 4
max. # of levels = 5
Yield > 98%

# of bias voltages = 2
from ZBB
RBB 300, 200, 100mV
FBB 300, 200, 100mV
At same # of testable levels, difference is small.
For each # of testable levels, min # of body voltages is sufficient.
Implications on synthesizing a circuit for post-silicon tuning

When not aiming at fastest speed, tuning should not be used beyond the speed w/o compensation.
Outline

• Background & objective
• Proposed body bias clustering
• Experiments
• Conclusion
Conclusions

- Developed a low test-cost body bias clustering that explicitly minimize leakage after compensation considering correlation between delay and leakage
  - Extended to handle more than three body voltages
    * When # of testable levels is smaller than # of clusters, only two voltages are sufficient.
  - Clustering attained 8.3% maximum delay reduction w/ 25.3% smaller leakage compared to non-clustering.
  - When not pursuing the maximum speed, tuning should not be used to increase the speed.