Reversible Statistical \textit{max/min} Operation: Concept and Applications to Timing

\textit{Debjit Sinha, Chandu Visweswariah, Jinjun Xiong\*, Vladimir Zolotov\*, and Natesan Venkateswaran}

IBM Electronic Design Automation, USA
*IBM T. J. Watson Research Center, USA

June 3-7, 2012
DAC 2012, San Francisco, California
Background: Statistical timing

- Increasing significance of variability
- Statistical static timing analysis (SSTA) and optimization
  - Timing quantities (e.g. delay, arrival time) represented by random variables with known distributions (e.g. Gaussian)

- Fundamental operations in timing analysis: Addition (*add*), Subtraction (*sub*), Maximum (*max*) and Minimum (*min*)
Background: Linear Gaussian statistical timing model

- Captures first-order sensitivity of any timing quantity to sources of variability [Chang et al. ICCAD’03; Visweswariah et al., Le et al. DAC’04]
  - Reasonably accurate for most variation sources
  - Simple – Less memory and run-time compared to complex (non-linear and non-Gaussian models)
  - Wider adoption in industrial SSTA tools

- Linear canonical form

\[ a_0 + a_1 \Delta X_1 + a_2 \Delta X_2 + \cdots + a_n \Delta X_n + a_{n+1} \Delta R_a \]

- Constant (nominal) value in the absence of variations
- Sensitivities
- Global random variables; these are probability distributions
- Independently random uncertainty
Background: *max/min* operation

- **Deterministic** *max/min* is straightforward
  - Guarantees the result is exactly same as one operand
  - Example: *deterministic max*{4, 5} = 5

- **Statistical** *max/min* of Gaussians is non-trivial
  - Result typically combination of operands (unlike deterministic)
    - Preserves *signatures* of operands
  - Approximated to a Gaussian using Clark’s approach to computing moments of the *max/min* of two Gaussians
    - [Chang *et al.* ICCAD’03; Visweswariah *et al.* DAC’04]
  - Performed pair-wise for *n* Gaussians

\[
\text{max}\{X_1, X_2, X_3, X_4\} \approx \text{max}\{X_1, \text{max}\{X_2, \text{max}\{X_3, X_4\}\}\}
\]
Statistical **max/min** during incremental timing

- Arrival time at Z \( (AT_Z) = \max\{AT_A, AT_B, AT_C\} \)

Consider gate re-sizing during incremental timing optimization

- Updated arrival time at Z \( (AT_Z^*) = \max\{AT_A^*, AT_B, AT_C\} \)
- Traditional statistical **max/min** non-incremental
  - Performed over all operands irrespective of how many changed
  - Inefficient for cases with large number of operands
  - How to perform “Incremental statistical **max/min**”?
Reversible-statistical max/min

- Given operand $X_1$ and result ($\text{max}^*$) $X_m$, compute operand $X_2$:
  \[
  \text{max}\{X_1, X_2\} = X_m
  \]

- Definition: Reversible-max\{X_1, X_m\} = X_2

- Deterministic reversible-max trivial, result not guaranteed
  - Reversible-max\{4, 5\} = 5
  - Reversible-max\{4, 4\} = ? (Non-unique! Could be any value <= 4)

- Statistical reversible-max non-trivial
  - Result exists – Recall statistical max is a combination of operands
  - Barring pathological cases
  - Reversible-statistical max\{X_1, X_m\} = X_2 can be computed

Statistical max of Gaussians approximated to a Gaussian  Result: Gaussian
Computing “unique” reversible-statistical \textit{max}

- Definitions and algorithm

\[
\phi(x) \triangleq \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)
\]

\[
\Phi(y) \triangleq \int_{-\infty}^{y} \phi(x)dx
\]

\[
F(\lambda) \triangleq \frac{(\lambda^2 + 1)\Phi(\lambda) + \lambda\phi(\lambda)}{[\lambda\Phi(\lambda) + \phi(\lambda)]^2}
\]

**Algorithm:** Reversible statistical \textit{max}

- **Input:** Gaussians \(X_1 \sim N(\mu_1, \sigma_1^2)\), \(X_m \sim N(\mu_m, \sigma_m^2)\)
- **Output:** Gaussian \(X_2 : \max(X_1, X_2) = X_m\)

1. If \(X_1 = X_m\), return \textit{failure}
2. Compute \(X_0 \sim N(\mu_0, \sigma_0^2) = X_m - X_1\)
3. Evaluate \(\frac{\mu_0^2 + \sigma_0^2}{\mu_0^2}\)
4. Find \(\lambda\) such that \(F(\lambda) = \frac{\mu_0^2 + \sigma_0^2}{\mu_0^2}\)
5. Re-construct \(X \sim N(\mu, \sigma^2)\) using \(\sigma = \frac{\mu_0}{\lambda\Phi(\lambda) + \phi(\lambda)}\) and \(\mu = \lambda\sigma\)
6. Re-construct \(X_2 = X + X_1\)

Main step for reversible \textit{max} computation

Used to obtain a canonical form for \(X_2\)
Reversible-\textbf{max} for incremental statistical \textbf{max}

- Given N Gaussians and their \textit{max}:
  \[ X_m = \text{max}\{X_1, X_2, \ldots, X_k, \ldots, X_N\} \]

- Consider incremental update of \underline{one} operand
  \[ X_k \rightarrow X_k^* \]

- Problem: Compute \underline{incremental-max} \underline{X}_m^*:
  \[ X_m^* = \text{max}\{X_1, X_2, \ldots, X_k^*, \ldots, X_N\} \]

- Proposed solution steps:
  - Compute \underline{reversible-max}\{X_k, X_m\} = X_m^{-k}
    \[ \text{One operation \ Assume cost = } (1 + \delta) \]
  - Conceptually \[ X_m^{-k} = \text{max}\{X_1, X_2, \ldots, X_{k-1}, X_{k+1}, \ldots, X_N\} \]
  - Compute \[ X_m^* = \text{max}\{X_{m^{-k}}, X_k^*\} \] \[ \text{Operation cost = 1} \]

- Run-time gain ratio: \[ \frac{N - 1}{2 + \delta} \]
Generalization for multiple operand updates

- Consider M of N Gaussians being incrementally updated
  \[ \{X_k, X_{k+1}, \ldots, X_{k+M-1}\} \to \{X_k^*, X_{k+1}^*, \ldots, X_{k+M-1}^*\} \]

- Computing **incremental-max** \( X_m^* \):
  \[ X_m^* = \max\{X_1, \ldots, X_k^*, X_{k+1}^*, \ldots, X_{k+M-1}^*, \ldots, X_N\} \]

**Algorithm:** Incremental max for M of N Gaussians

**Input:** Gaussians \( \{X_1, \ldots, X_N\} \), \( X_m \),
\[ \{X_k^*, \ldots, X_{k+M-1}^*\} \]

**Output:** Gaussian \( X_m^* \)

1. Compute \( X_a = \max(X_k, X_{k+1}, \ldots, X_{k+M-1}) \)
2. Re-construct \( X_m^{-k,M} = \text{reversible max} (X_a, X_m) \)
3. If failure in re-constructing \( X_m^{-k,M} \), revert to traditional max for \( X_m^* \); exit
4. Compute
   \[ X_m^* = \max(X_m^{-k,M}, X_k^*, X_{k+1}^*, \ldots, X_{k+M-1}^*) \]

- **Run-time gain ratio:** \[ \frac{N - 1}{2M + \delta} \]
  Indicates potential of significant run-time gain for cases with **large** N and **small** M
Graphical illustration of run-time gain

Gain: \( \frac{N - 1}{2M + \delta} \)

Plots of run-time gain as functions of \( N \) and \( M \)
Applications in statistical timing

- Incremental chip-slack (yield) computation
  - Update

- Incremental chip timing yield prediction
  - Without any timing propagation!

- Incremental chip timing yield gradient computation

- Incremental criticalities [Xiong et al. DATE’08]

Chip slack = \( \min \{S_1, S_2, S_3^{\text{new}}, \ldots, S_n\} \)

- Incremental
- Efficient
- Accurate

Please see poster for details on these applications
Conclusions

- Concept of **reversible-statistical max/min** introduced
  - Formal proof of uniqueness of reversible-max/min presented
  - Enables incremental statistical max/min
  - Run-time independent of total number of operands, depends only on number of updated operands
  - Potential of significant run-time gain \( \left( \frac{N - 1}{2M + \delta} \right) \) – Validated by experimental results

- Several applications of concept to timing highlighted
  - Incremental chip slack/yield computation
  - Incremental chip yield prediction
  - Incremental yield gradient estimation
  - Incremental criticalities

Thank you