Literature Review of Statistical Timing Analysis

Jinjun Xiong

1. REVIEW

1.1 Path-based Approach

In [1], a path-based SSTA approach is proposed which consists of two phases (1) identifying logically sensitizable critical path based upon SSTA; and (2) identifying timingly true critical path based on the statistical timing information.

The first phase has three steps. Firstly, a worst-case deterministic timing analysis is performed so that the critical nodes and timing slacks for all nodes are obtained. Based upon criticality of nodes and slackness of the nodes, paths that consists of critical nodes and nodes with certain probability being critical are selected. Secondly, the PDFs of node delays are obtained via Monte Carlo simulation. The correlation between nodes are only considered during the Monte Carlo simulation stage, where nodes are divided into groups and each group share one same correlation factor. After obtained individual PDFs, no correlation information between nodes is kept. Thirdly, paths that are logically sensitizable are selected based upon sensitization criterion.

The second phase is to use Monte Carlo simulation to find the timing distribution for each selected logically sensitizable critical path. Paths are ordered according to their criticality probabilities.

———-

In [2], a path-based statistical delay computation is proposed to consider both inter-die variation and intra-die variation. The intra-die spatial variation is modeled by a hierarchical quad-tree like linear combinations of independent variations. Gate delay is a function of the underlying channel length variation.

As there is one common inter-die variation among all devices at the same die, a Monte Carlo simulation is performed to obtain the distribution of path delay variation due to intra-die variation.

For the intra-die variation, the gate delay and slope are first linearized as a combination of the underlying channel length variation, next stage gate input capacitance variation and previous stage gate slope. The later two are further transformed to linear combinations of the underlying channel length variations. Therefore, at the end, the intra-die path delay is still a linear combinations of the underlying channel length variations, even after considering spatial correlations. By approximating the channel length variation as a normal distribution, the intra-die induced path delay variation is also a normal distribution.

The final path-delay variation is the sum of intra-die induced path variation and inter-die induced path variation, which can be obtained via numerical convolution.

1.2 Block-based Approach

[3] is a block-based SSTA that does not assume normal distribution. In contrast, a piece-wise uniform distribution (PDF) or a piece-wise linear distribution (CDF) is assumed for gate delays. Moreover, the gate delays are assumed to be independent.

Because of the independence assumption about gate delays, the sum operation and max operation can be easily computed, i.e., sum's PDF is obtained via convolution of two PDFs, the max's CDF is obtained via product of two CDFs. For piece-wise linear functions, the above operations can be computed efficiently.

To handle the path re-convergence problem, a *dependent list* is maintained at each node that lists the all previous stage nodes on which current node's arrival time depends. It has been shown in [3] that if two nodes have only one common predecessor node, closed formulae can be used to compute the node's CDF efficiently, provided that the CDF can be characterized by two moments (like normal distribution). In case of more than two common nodes, a heuristic algorithm is proposed to reduce the multiple dependent nodes to one node so that the developed closed formulae can be used. To further speed up the algorithm, the size of the dependent list is left as a tuning parameter, and moreover, after maximum operation, the dependent list is flushed.

———-

[4] considers both inter-die variation and intra-die variation (spatial variation). A first order variation model is proposed for device delay, which is a linear function of underlying channel length variation. The $Leff$ variation is a sum of inter-die variation and intra-die variations. Inter-die variation is modeled by one common random variable for all devices on the die. The intra-die variation is modeled by using a hierarchical quad-tree structure. All component-wise random variables are assumed to be independent.

For SSTA, [4] assumed that the correlation due to path-reconvergence can be ignored as that only results in pessimistic results according to [5]. Therefore, only spatial correlation is considered in [4].

Because the gate delay is represented as a linear combination of independent random variables, the sum operation is computed easily and similar linear combination forms can be preserved after sum operation.

The main challenge is due to the maximum operation. The authors walk around such a problem by computing an upper bound CDF of the maximum operation, i.e., taking the component wise maximum operations instead of the maximum of the whole sum. However, this approximation may result in a very loose upper bound. Therefore, the authors further refine such a problem by using the following heuristic approach. First, at each node, the pair-wise arrival times are merged via the upper bound maximum operation. The least mean one is kept to replace the original pair of arrival times. This procedure is repeated until a finite number of arrival times are left at the node. Then all of these remaining arrival times are propagated to the next stage instead of taking the maximum operation at the current stage. At the final end of the primary output, another upper bound maximum operation is taken to compute the delay distribution.

Experiment are done on ISCAS85 benchmark and results are compared with Monte Carlo simulation. Figure-of-merits are the 99on the distribution.

————-

Under the normal distribution assumption, [6] modeled the gate delay as a normal random variable, and the arrival time at all nodes as normal distributions as well. Therefore, under block-based SSTA, sum and max operations are two of the atomic operations that need to be solved statistically.

Because [6] considered both the inter-die variations and the intra-die variations, though the model is not clearly specified in the paper, the above two operations need to consider correlation between all random variables. However, as normal distribution is assumed, only the covariance between any two random variables need to be computed.

The closed formula for the sum operation is easy to derive. The pain is due to max operation. To solve this problem, a table-look-up based approach is taken that converts the maximum between two random variables into a one-dimensional table look-up in order to compute the mean, the variance and covariance. Experiment results based upon ISCAS85 benchmark are reported. The comparison is between SSTA and Monte Carlo simulation at the 6σ point.

An extension of this work is to consider timing window statistically.

————-

[7] approximates the delay in the timing graph as a linear combination of independent global variations, which is called *canonical* form. Furthermore, by assuming the variations are normal variables, they also obtain that the delay is also a normal distribution.

To solve the max operation, [7] proposed the idea of *tightness*, which defines the probability of one variable dominates the other. Moreover, they assume that the max operation preserves the canonical form, which is just a linear combination of the individual terms with the weights of each term being the tightness probability.

————-

[8] considers the spatial correlations by assuming that the correlation of underlying physical parameters in different regions are known in terms of a covariance matrix. Because the delay of devices or wires can be modeled as a linear function of the underlying random variations, which are correlated, a PCA (principle component analysis) method is employed to decoupling the correlated terms. After PCA, the delay of devices or wires can be written as a linear combination of independent random variables, which are independent component obtained from PCA.

To solve the max operation, [8] also enforced that the max operation preserves the form as a linear combination of independent PCA components. By employing an approximation formula from [8], the authors derived closed formulae to compute the new coefficients after the max operation.

2. REFERENCES

[1] J.-J. Liou, A. Krstic, L.-C. Wang, and K.-T. Cheng, "False-path-aware statistical timing analysis and efficient path selection for delay testing and timing validation," June 2002.

[2] A. Agarwal, D. Blaauw, V. Zolotov, S. Sundareswaran, M. Zhao, K. Gala, and R. Panda, "Statistical delay computation considering spatial correlations," Jan 2003.

[3] A. Devgan and C. Kashyap, "Block-based static timing analysis with uncertainty," in *ICCAD 03*, Nov 2003.

[4] A. Agarwal, D. Blaauw, and V. Zolotov, "Statistical timing analysis for intra-die process variations with spatial correlations," Nov. 2003.

[5] A. Agarwal, V. Zolotov, and D. Blaauw, "Statistical timing analysis using bounds and selective enumeration," Sept. 2003.

[6] J. Le, X. Li, and L. Pileggi, "Stac: Statisical timing analysis with correlation," in *DAC 04*, Jun 2004.

[7] C. Visweswariah, K. Ravindran, K. Kalafala, S. G. Walker, and S. Narayan, "First-order incremental block-based statistical timing analysis," in *Proc. Design Automation Conf*, Jun 2004.

[8] H. Chang and S. Sapatnekar, "Statistical timing analysis considering spatial correlations using a single PERT-like traversal," Nov. 2003.