Interval-Based Robust Statistical Techniques for Non-Negative Convex Functions, with Application to Timing Analysis of Computer Chips

Michael Orshansky
Wei-Shen Wang
Department of Electrical and Computer Engineering
University of Texas at Austin
Austin, TX 78712,USA
orshansky@mail.utexas.edu

Martine Ceberio
Gang Xiang
Department of Computer Science
University of Texas at El Paso
El Paso, TX 79968, USA
mceberio@cs.utep.edu,
gxiang@utep.edu

ABSTRACT

In chip design, one of the main objectives is to decrease its clock cycle. On the design stage, this time is usually estimated by using worst-case (interval) techniques, in which we only use the bounds on the parameters that lead to delays. This analysis does not take into account that the probability of the worst-case values is usually very small; thus, the resulting estimates are over-conservative, leading to unnecessary over-design and under-performance of circuits. If we knew the *exact* probability distributions of the corresponding parameters, then we could use Monte-Carlo simulations (or the corresponding analytical techniques) to get the desired estimates. In practice, however, we only have *partial* information about the corresponding distributions, and we want to produce estimates that are valid for all distributions which are consistent with this information.

In this paper, we develop a general technique that allows us, in particular, to provide such estimates for the clock time.

1. CASE STUDY

Decreasing clock cycle: a practical problem. In chip design, one of the main objectives is to decrease the chip's clock cycle. It is therefore important to estimate the clock cycle on the design stage.

The clock cycle of a chip is constrained by the maximum path delay over all the circuit paths $D \overset{\text{def}}{=} \max(D_1, \ldots, D_N)$, where D_i denotes the delay along the i-th path. Each path delay D_i is the sum of the delays corresponding to the gates and wires along this path. Each of these delays, in turn, depends on several factors such as the variation caused by the current design practices, environmental design characteristics (e.g., variations in temperature and in supply voltage), etc.

Traditional (interval) approach to estimating the clock cycle. Traditionally, the delay D is estimated by using the worst-case analysis, in which we assume that each of the corresponding factors takes the worst possible value (i.e., the value leading to the largest possible delays). As a result, we get the time delay that corresponds to the case when all the factors are at their worst.

It is necessary to take probabilities into account. The worst-case analysis does not take into account that different factors come from independent random processes. As a result, the probability that all these factors are at their worst is extremely small. For example, there may be slight variations of delay time from gate to gate, and this can indeed lead to gate delays. The worst-case analysis considers the case when all these random variations lead to the worst case; since these variations are independent, this combination of worst cases is highly unprobable.

As a result, the current estimates of the chip clock time are over-conservative, over up to 30% above the observed clock time. Because of this over-estimation, the clock time is set too high – i.e., the chips are usually over-designed and under-performing; see, e.g., [6, 7, 8, 22, 21, 23, 24]. To improve the performance, it is therefore desirable to take into account the probabilistic character of the factor variations.

Robust statistical methods are needed. If we knew the *exact* probability distributions of the corresponding parameters, then we could use Monte-Carlo simulations (or the corresponding analytical techniques) to get the desired estimates. In practice, however, we only have *partial* information about the corresponding distributions. For a few parameters, we know the exact distribution, but for most parameters, we only know the mean and some characteristic of the deviation from the mean – e.g., the interval that is guaranteed to contain possible values of this parameter.

In principle, we could pick up some distributions which are consistent with this partial information – e.g., truncated normal distributions, compute the maximum delays D corresponding to all these distributions, and then take the largest $D_{\max}$ of these computed maximum delays D as the clock time. This procedure will guarantee that the path delay D does not exceed the clock time if the actual distribution is one of the picked ones. However, it is quite possible that some other possible distributions (different from the ones

SAC'06 April 23-27, 2006, Dijon, France

we picked), the corresponding path delay D is larger than $D_{\max}$. As a result, we may be underestimating the clock time. If we set the clock time too low, we may have operations that did not have time to finish before the next cycle starts – and this is even worse than overestimating.

It is therefore desirable to provide bounds that work for all the distributions which are consistent with the given information. In statistics, estimates which are guaranteed for all distributions from some non-parametric class are called *robust* (see, e.g., [13]). In these terms, our objective is to provide robust statistical estimates for the clock time.

What we do in this paper. In this paper, we develop general techniques that allow us, in particular, to provide robust estimates for the clock time.

In deriving these estimates, we will use the extensions of interval methods to cases with partial information about probabilities described, e.g., in [11, 17, 18, 19]; see also [1, 2, 3, 4, 5, 20].

2. TOWARDS A MATHEMATICAL FORMULATION OF THE PROBLEM

Case study: how the desired delay D depends on the parameters. The variations in the each gate delay d are caused by the difference between the actual and the nominal values of the corresponding parameters. It is therefore desirable to describe the resulting delay d as a function of these differences $x_1, \ldots, x_n$. Since these differences are usually small, we can safely ignore quadratic (and higher order) terms in the Taylor expansion of the dependence of d on x_j and assume that the dependence of each delay d on these differences can be described by a linear function.

As a result, each path delay D_i – which, as we have mentioned, is the sum of delays at different gates and wires – can also be described as a linear function of these differences, i.e., as $D_i = a_i + \sum\limits_{j=1}^{n} a_{ij} \cdot x_j$ for some coefficients a_i and a_{ij}.

Thus, the desired maximum delay $D = \max\limits_i D_i$ has the form

$$D = \max_i \left(a_i + \sum_{j=1}^{n} a_{ij} \cdot x_j \right). \qquad (1)$$

How we can describe such functions in general terms. In this paper, we will use two properties of the time delay. First, we will use the fact that the time delay is always non-negative; second, we will use the fact that the dependence (1) is convex.

Let us recall that a function $f : R^m \to R$ is called *convex* if

$$f(\alpha \cdot x + (1-\alpha) \cdot y) \le \alpha \cdot f(x) + (1-\alpha) \cdot f(y)$$

for every $x, y \in R^m$ and for every $\alpha \in (0,1)$. It is known that the maximum of several linear functions is convex, so the function (1) is convex. Vice versa, every convex function can be approximated, with an arbitrary accuracy, by maxima of linear functions – i.e., by expressions of type (1).

So, in general terms, we can say that we are interested in the robust statistical properties of the value $y = F(x_1, \ldots, x_n)$, where F is a non-negative convex function of the variables x_j.

Our objective. We want to find the smallest possible value y_0 such that for all possible distributions consistent with the known information, we have $y \le y_0$ with the probability $\ge 1 - \varepsilon$ (where $\varepsilon > 0$ is a given small probability).

What information we can use. What information can we use for these estimations? We can safely assume that different factors x_j are statistically independent. About some of the variables x_j, we know their exact statistical characteristics; about some other variables x_j, we only know their interval ranges $[\underline{x}_j, \overline{x}_j]$ and their means E_j.

Additional property: the dependency is non-degenerate. We only have partial information about the probability distribution of the variables x_j. For each possible probability distribution p, we can find the largest value y_p for which, for this distribution, $y \le y_p$ with probability $\ge 1 - \varepsilon$. The desired value y_0 is the largest of the values y_p corresponding to different probability distributions p: $y_0 = \sup\limits_{p \in \mathcal{P}} y_p$, where $\mathcal{P}$ denotes the class of probability distributions p which are consistent with the known information.

If we learn some additional information about the distribution of x_j – e.g., if we learn that x_j actually belongs to a proper subinterval of the original interval $[\underline{x}_j, \overline{x}_j]$ – we thus decrease the class $\mathcal{P}$ of distributions p which are consistent with this information, to a new class $\mathcal{P}' \subset \mathcal{P}$. Since the class has decreased, the new value $y_0' = \sup\limits_{p \in \mathcal{P}'} y_p$ is the maximum over a smaller set and thus, cannot be larger than the original value y_0: $y_0' \le y_0$.

From the purely mathematical viewpoint, it is, in principle, possible that the desired value y does not actually depend on some of the variables x_j. In this case, if we narrow down the interval of possible values of the corresponding variable x_j, this will not change the resulting value y_0.

For the chip design problem, it is reasonable to assume that such variables have already been weeded out, and that the resulting function $F(x_1, \ldots, x_n)$ is *non-degenerate* in the sense that every time we narrow down one of the intervals $[\underline{x}_j, \overline{x}_j]$, the resulting value y_0 actually decreases: $y_0' < y_0$.

As a result, we arrive at the following problem.

3. FORMULATION OF THE PROBLEM AND THE MAIN RESULT

GIVEN:

- natural numbers n, and $k \le n$;
- a real number $\varepsilon > 0$;
- a function $y = F(x_1, \ldots, x_n)$ (algorithmically defined) such that for every combination of values $x_{k+1}, \ldots, x_n$, the dependence of y on $x_1, \ldots, x_k$ is convex;
- $n - k$ probability distributions $x_{k+1}, \ldots, x_n$ – e.g., given in the form of cumulative distribution function (cdf) $F_j(x)$, $k+1 \le j \le n$;
- k intervals $\mathbf{x}_1, \ldots, \mathbf{x}_k$, and
- k values $E_1, \ldots, E_k$,

such that for every $x_1 \in [\underline{x}_1, \overline{x}_1], \ldots, x_k \in [\underline{x}_k, \overline{x}_k]$, we have $F(x_1, \ldots, x_n) \ge 0$ with probability 1.

TAKE: all possible joint probability distributions on R^n for which:

- all n random variables are independent;

- for each j from 1 to k, $x_j \in \mathbf{x}_j$ with probability 1 and the mean value of x_j is equal to E_j;
- for $j > k$, the variable x_j has a given distribution $F_j(x)$.

FIND: find the smallest possible value y_0 such that for all possible distributions consistent with the known information, we have $y \stackrel{\text{def}}{=} F(x_1, \ldots, x_n) \le y_0$ with probability $\ge 1 - \varepsilon$.

PROVIDED: that the problem is *non-degenerate* in the sense that if we narrow down one of the intervals $\mathbf{x}_j$, the value y_0 decreases.

The following result explains how we can compute this value y_0.

Proposition 1. *The desired value y_0 is attained when for each j from 1 to k, we use a 2-point distribution for x_j, in which:*

- $x_j = \underline{x}_j$ *with probability* $\underline{p}_j \stackrel{\text{def}}{=} \dfrac{\overline{x}_j - E_j}{\overline{x}_j - \underline{x}_j}$.
- $x_j = \overline{x}_j$ *with probability* $\overline{p}_j \stackrel{\text{def}}{=} \dfrac{E_j - \underline{x}_j}{\overline{x}_j - \underline{x}_j}$.

Comment. The proof of Proposition 1 is given in the special (last) section of this paper.

4. RESULTING ALGORITHM FOR COMPUTING Y_0

Because of Proposition 1, we can compute the desired value y_0 by using the following Monte-Carlo simulation:

- We set each value x_j, $1 \le j \le k$, to be equal to $\overline{x}_j$ with probability $\overline{p}_j$ and to the value $\underline{x}_j$ with the probability $\underline{p}_j$.
- We simulate the values x_j, $k < j \le n$, as random variables distributed according to the distributions $F_j(x)$.
- For each simulation s, $1 \le s \le N_i$, we get the simulated values $x_j^{(s)}$, and then, a value $y^{(s)} = F(x_1^{(s)}, \ldots, x_n^{(s)})$. We then sort the resulting N_i values $y^{(s)}$ into an increasing sequence
$$y_{(1)} \le y_{(2)} \le \ldots \le y_{(N_i)},$$
and take, as y_0, the $N_i \cdot (1-\varepsilon)$-th term $y_{(N_i \cdot (1-\varepsilon))}$ in this sorted sequence.

Comment about Monte-Carlo techniques. Before presenting the algorithm for computing the upper bound on y_0, let us remark that some readers may feel uncomfortable with the use of Monte-Carlo techniques. This discomfort comes from the fact that in the *traditional* statistical approach, when we know the exact probability distributions of all the variables, Monte-Carlo methods – that simply simulate the corresponding distributions – are inferior to analytical methods. This inferiority is due to two reasons:

- First, by design, Monte-Carlo methods are approximate, while analytical methods are usually exact.
- Second, the accuracy provided by a Monte-Carlo method is, in general, proportional to $\sim 1/\sqrt{N_i}$, where N_i is the total number of simulations. Thus, to achieve reasonable quality, we often need to make a lot of simulations – as a result, the computation time required for a Monte-Carlo method becomes much longer than for an analytical method.

In *robust* statistic, there is often an additional reason to be uncomfortable about using Monte-Carlo methods:

- Practitioners use these methods by selecting a finite set of distributions from the infinite class of all possible distributions, and running simulations for the selected distributions.
- Since we do not test all the distributions, this practical heuristic approach sometimes misses the distributions on which the minimum or maximum of the corresponding distribution is actually attained.

In our case, we also select a finite collection of distributions from the infinite set. However, in contrast to the heuristic (un-justified) selection – which is prone to the above criticism, our selection is *justified.* Proposition 1 *guarantees* that the values corresponding to the selected distributions indeed provide the desired value y_0 – the largest over all possible distributions $p \in \mathcal{P}$.

In such situations, where a justified selection of Monte-Carlo methods is used to solve a problem of robust statistics, such Monte-Carlo methods often lead to *faster* computations than known analytical techniques. The speed-up caused by using such Monte-Carlo techniques is one of the main reasons why they were invented in the first place – to provide fast estimates of the values of multi-dimensional integrals. Many examples of efficiency of these techniques are given, e.g., in [25]; in particular, examples related to estimating how the uncertainty of inputs leads to uncertainty of the results of data processing are given in [26].

Comment about non-linear terms. In the formula (1), we ignored quadratic and higher order terms in the dependence of each path time D_i on the parameters x_j. It is known that the maximum $D = \max_i D_i$ of convex functions D_i is always convex. So, according to Proposition 1, the above algorithm will work if we take quadratic terms into consideration – provided that each dependence $D_i(x_1, \ldots, x_k, \ldots)$ is still convex.

5. PROOF OF PROPOSITION 1

By definition, y_0 is the largest value of y_p over all possible distributions $p \in \mathcal{P}$. This means that for the given y_0, for all possible distributions $p \in \mathcal{P}$, we have $\text{Prob}(D \le y_0) \ge 1 - \varepsilon$. Let $p \in \mathcal{P}$ be the "worst-case" distribution, i.e., the distribution for which the probability $\text{Prob}(D \le y_0)$ is the smallest. Let us show that this "worst case" occurs when all k variables $x_1, \ldots, x_k$ have the 2-point distributions described in Proposition 1.

Let us fix the value $j \le k$ and show that in the "worst case", x_j indeed has the desired 2-point distribution. Without losing generality, we can take $j = 1$. Let us fix the distributions for $x_2, \ldots, x_k$ as in the worst case. Then, the fact that the probability $\text{Prob}(D \le y_0)$ is the smallest means that if we replace the worst-case distribution for x_1 with some other distribution, we can only increase this probability. In other words, when we correspondingly fix the distributions for $x_2, \ldots, x_k$, the probability $\text{Prob}(D \le y_0)$ attains the smallest possible value at the desired distribution for x_1.

In reality, the distribution for x_1 is located on an interval $\mathbf{x}_1 = [\underline{x}_1, \overline{x}_1]$, i.e., on a set with infinitely many points. However, with an arbitrary large value N (and thus, for an arbitrarily small discretization error $\delta = (\overline{x}_1 - \underline{x}_1)/N$), we can assume that all the distributions are located on a finite grid of values

$$v_0 \stackrel{\text{def}}{=} \underline{x}_1, \quad v_1 \stackrel{\text{def}}{=} \underline{x}_1 + \delta, \quad v_2 \stackrel{\text{def}}{=} \underline{x}_1 + 2\delta, \ldots, v_N = \overline{x}_1.$$

The smaller δ, the better this approximation. Thus, without losing generality, we can assume that the distribution of x_1 is located on finitely many points v_i.

In this approximation, the probability distribution for x_1 can be described by the probabilities $q_i \stackrel{\text{def}}{=} p_1(v_i)$ of different values v_i.

The minimized probability $\text{Prob}(D \le y_0)$ can be described as the sum of the probabilities of different combinations $(x_1, \ldots, x_n)$ over all the combinations for which $D(x_1, \ldots, x_n) \le y_0$. We assumed that all the variables x_j are independent. Thus, the probability of each combination $(x_1, \ldots, x_n)$ is equal to the product of the corresponding probabilities $p_1(x_1) \cdot p_2(x_2) \cdot \ldots$ Since the probability distributions for $x_2, \ldots$ are fixed, the minimized probability is thus a linear combination of probabilities $p_1(v_i)$, i.e., of the probabilities q_i. In other words, the minimized probability has the form $\sum\limits_{i=0}^{N} c_i \cdot q_i$ for some coefficients c_i.

By describing the probability distribution on x_1 via the probabilities $q_i = p_1(v_i)$ of different values $v_i \in [\underline{x}_1, \overline{x}_1]$, we automatically restrict ourselves to distributions which are located on this interval. The only restrictions that we have on the probability distribution of x_1 is that it is a probability distribution, i.e., that $q_i \ge 0$ for all i and $\sum\limits_{i=0}^{N} q_i = 1$, and that the mean value of this distribution is equal to E_1, i.e., that $\sum\limits_{i=0}^{N} q_i \cdot v_i = E_1$. Thus, the worst-case distribution for x_1 is a solution to the following linear programming problem:

Minimize

$$\sum_{i=0}^{N} c_i \cdot q_i$$

under the constraints

$$\sum_{i=0}^{N} q_i = 1,$$

$$\sum_{i=0}^{N} q_i \cdot v_i = E_1,$$

$$q_i \ge 0, \quad i = 0, 1, 2, \ldots, N.$$

It is known that the solution to the linear programming problem is always attained at a vertex of the corresponding constraint set. In other words, in the solution to the linear programming problem with $N+1$ unknowns $q_0, q_1, \ldots, q_N$, at least $N+1$ constraints are equalities. Since we already have 2 equality constraints, this means that out of the remaining constraints $q_i \ge 0$, at least $N-1$ are equalities. In other words, this means that in the optimal distribution, all but two values of $q_i = p_1(v_i)$ are equal to 0.

Thus, the "worst-case" distribution for x_1 is located on 2 points v and v' within the interval $[\underline{x}_1, \overline{x}_1]$. Let us prove, by reduction to a contradiction, that these two points cannot be different from the endpoints of this interval. Indeed, let us assume that they are different. Without losing generality, we can assume that $v \le v'$. Then, this "worst-case" distribution is actually located on the proper subinterval $[v, v'] \subset [\underline{x}_1, \overline{x}_1]$ of the original interval $\mathbf{x}_1$. Since the maximum y_0 of y_p is attained on this distribution, replacing the original interval $\mathbf{x}_1$ with its proper subinterval $[v, v']$ would not change the value y_0 – while our assumption of non-degeneracy states that such a replacement would always lead to a smaller value y_0. This contradiction shows that the values v and v' – on which the worst-case distribution is located – have to be endpoints of the interval $[\underline{x}_1, \overline{x}_1]$.

In other words, we conclude that the worst-case distribution is located at 2 points: $\underline{x}_1$ and $\overline{x}_1$. Such a distribution is uniquely determined by the probabilities $\underline{p}_1$ and $\overline{p}_1$ of these two points. Since the sum of these probabilities is equal to 1, it is sufficient to describe one of these probabilities, e.g., $\overline{p}_1$; then, $\underline{p}_1 = 1 - \overline{p}_1$. The condition that the mean of x_1 is E_1, i.e., that

$$\underline{p}_1 \cdot \underline{x}_1 + \overline{p}_1 \cdot \overline{x}_1 = (1 - \overline{p}_1) \cdot \underline{x}_1 + \overline{p}_1 \cdot \overline{x}_1 = E_1,$$

uniquely determines $\overline{p}_1$ (and hence $\underline{p}_1$) – exactly by the expression from Proposition 1. The statement is proven.

6. CONCLUSIONS

In chip design, one of the main objectives is to decrease its clock cycle.

On the design stage, this time is usually estimated by using worst-case (interval) techniques, in which we only use the bounds on the parameters that lead to delays. This analysis does not take into account that the probability of the worst-case values is usually very small; thus, the resulting estimates are over-conservative, leading to unnecessary over-design and under-performance of circuits. Instead of the largest possible value of the delay, it is reasonable to determine the clock time as the time y_0 for which the probability that the actual delay y exceeds y_0 does not exceed a given small value ε.

If we knew the *exact* probability distributions of the corresponding parameters, then we could use Monte-Carlo simulations (or the corresponding analytical techniques) to get the desired value y_0. In practice, however, we only have *partial* information about the corresponding distributions, and we want to produce the value y_0 which is valid for all distributions which are consistent with this information.

In this paper, we describe a general technique that allows us, in particular, to compute this value y_0. This technique uses Monte-Carlo simulations with specially selected "worst-case" distributions, distributions for which the delay is provably largest among all distributions from the given class. Thus, to guarantee that $\text{Prob}(y \le y_0) \ge 1 - \varepsilon$ *for all* distributions from the given class, it is sufficient to check this inequality for the selected "worst-case" distributions.

7. ACKNOWLEDGMENTS

This work was supported in part by the Army Research Lab grant DATM-05-02-C-0046. The authors are thankful to Vladik Kreinovich for his help and to the anonymous referees for their valuable suggestions.

8. REFERENCES

[1] D. Berleant, M.-P. Cheong, C. Chu, Y. Guan, A. Kamal, G. Sheblé, S. Ferson, and J. F. Peters, Dependable handling of uncertainty, *Reliable Computing* 9(6) (2003), pp. 407–418.

[2] D. Berleant, L. Xie, and J. Zhang, Statool: a tool for Distribution Envelope Determination (DEnv), an interval-based algorithm for arithmetic on random variables, *Reliable Computing* 9(2) (2003), pp. 91–108.

[3] D. Berleant and J. Zhang, Using Pearson correlation to improve envelopes around the distributions of functions, *Reliable Computing*, 10(2) (2004), pp. 139–161.

[4] D. Berleant and J. Zhang, Representation and Problem Solving with the Distribution Envelope Determination (DEnv) Method, *Reliability Engineering and System Safety*, 85 (1–3) (July-Sept. 2004).

[5] D. Berleant and J. Zhang, Using Pearson correlation to improve envelopes around the distributions of functions, *Reliable Computing*, 10(2) (2004), pp. 139–161.

[6] D. Boning and S. Nassif, "Models of process variations in device and interconnect", in: A. Chandrakasan (ed.), *Design of High-Performance Microcomputer Ciucruits*, 2000.

[7] D. Chinnery and K. Keutzer, "Closing the gap between ASICs and custom", *Proc. of the Design Automation Conference DAC'2000.*

[8] D. Chinnery and K. Keutzer (eds.), *Closing the Gap Between ASICs and Custom*, Kluwer, 2002.

[9] R. E. Edwards, *Functional analysis: theory and applications*, Dover Publ., N.Y., 1995.

[10] S. Ferson, *RAMAS Risk Calc 4.0: Risk Assessment with Uncertain Numbers*, CRC Press, Boca Raton, Florida, 2002.

[11] S. Ferson, L. Ginzburg, V. Kreinovich, and J. Lopez, "Absolute Bounds on the Mean of Sum, Product, etc.: A Probabilistic Extension of Interval Arithmetic", *Extended Abstracts of the 2002 SIAM Workshop on Validated Computing*, Toronto, Canada, May 23–25, 2002, pp. 70–72.

[12] J. Galambos, *The Asymptotic Theory of Extreme Order Statistics*, Wiley, New York, 1987.

[13] P. J. Huber, *Robust statistics*, Wiley, New York, 2004.

[14] L. Jaulin, M. Kieffer, O. Didrit, and E. Walter, *Applied interval analysis: with examples in parameter and state estimation, robust control and robotics*, Springer-Verlag, London, 2001.

[15] S. Kotz and S. Nadarajah, *Extreme Value Distributions: Theory and Applications*, Imperial College Press, London, UK, 2000.

[16] V. Kreinovich and S. Ferson, "Computing best-possible bounds for the distribution of a sum of several variables is NP-hard", *International Journal of Approximate Reasoning*, to appear.

[17] V. Kreinovich, S. Ferson, and L. Ginzburg, "Exact Upper Bound on the Mean of the Product of Many Random Variables With Known Expectations", *Reliable Computing*, 9(6) (2003), pp. 441–463.

[18] V. Kreinovich and L. Longpré, "Computational complexity and feasibility of data processing and interval computations, with extension to cases when we have partial information about probabilities", In: V. Brattka, M. Schroeder, K. Weihrauch, and N. Zhong, *Proceedings of the Conference on Computability and Complexity in Analysis CCA'2003*, Cincinnati, Ohio, USA, August 28–30, 2003, pp. 19–54.

[19] V. Kreinovich, G. N. Solopchenko, S. Ferson, L. Ginzburg, and R. Aló, "Probabilities, intervals, what next? Extension of interval computations to situations with partial information about probabilities", *Proceedings of the 10th IMEKO TC7 International Symposium on Advances of Measurement Science*, St. Petersburg, Russia, June 30–July 2, 2004, Vol. 1, pp. 137–142.

[20] A. T. Langewisch and F. F. Choobineh, "Mean and variance bounds and propagation for ill-specified random variables", *IEEE Transactions on Systems, Man, and Cybernetics, Part A*, 34(4) (2004), pp. 494–506.

[21] M. Orshansky, "Increasing circuit performance through statistical design techniques", in: D. Chinnery and K. Keutzer (eds.), *Closing the Gap Between ASICs and Custom*, Kluwer, 2002.

[22] M. Orshansky and A. Bandyopadhyay, "Fast statistical timing analysis handling arbitrary delay correlations", *Proc. of the Design Automation Conference DAC'2004*, San Diego, California, June 7–11, 2004, pp. 337–342.

[23] M. Orshansky and K. Keutzer, "A general probabilistic framework for worst case timing analysis," *Proceedings of the Design Automation Conference DAC'2002*, June 2002, pp. 556–561.

[24] M. Orshansky and K. Keutzer, "From Blind Certainty to Informed Uncertainty," *Proceedings of the TAU Workshop*, December 2002.

[25] S. Rajasekaran, P. Pardalos, J. Reif, and J. Rolim (eds.), *Handbook on Randomized Computing*, Kluwer, 2001.

[26] R. Trejo and V. Kreinovich, "Error Estimations for Indirect Measurements: Randomized vs. Deterministic Algorithms For 'Black-Box' Programs", In: S. Rajasekaran, P. Pardalos, J. Reif, and J. Rolim (eds.), *Handbook on Randomized Computing*, Kluwer, 2001, pp. 673-729.

[27] H. M. Wadsworth Jr., *Handbook of statistical methods for engineers and scientists*, McGraw-Hill, N.Y., 1990.