Probability Density Function Based Reliability Evaluation of Large-Scale ICs

Nicoleta Cucu Laurenciu, and Sorin D. Cotofana
Computer Engineering Laboratory,
Delft University of Technology, The Netherlands.
{N.CucuLaurenciu, S.D.Cotofana}@tudelft.nl

***Abstract*—For the current advanced technology nodes, an accurate, yet fast reliability analysis is needed at design time, to enable the comparison between different circuit architectures, and thus a reliability-aware design and synthesis process. To this end we propose a reliability assessment framework that is able to estimate more accurately the circuit reliability and which can be applied to large-scale circuit settings, by: (i) taking into account the circuit topology (and implicitly its reconvergent fanouts), the input vectors, the environmental conditions and fault scenarios, (ii) employing a range of probabilities, i.e., a Probability Density Function (PDF), instead of hitherto single probability value, in order to quantify the circuit reliability, (iii) employing variational inference, to derive the circuit primary output PDFs, given its primary inputs PDFs, and (iv) adapting the traditional variational inference approach to exploit the peculiarities of the probabilistic model afferent to logic circuits, for convergence speed improvements and thus applicability in large scale circuits settings.**

***Index Terms*—variational inference, probability density function, IC reliability.**

I. Introduction

Traditionally, reliability begins to be evaluated late in the Integrated Circuit (IC) life cycle [1] (i.e., shortly before manufacturing release). However, in the realm of variability and higher failure probabilities expected for emerging nanodevices and their afferent interconnects, reliability should be also addressed upstream, from the early design inception phase to the in-field operation phase. Building architectures from prohibitively unreliable emerging nano-devices, expected to exhibit increased susceptibility to variations (e.g., manufacturing, permanent, and transient failures), requires the inclusion of reliability as an optimization goal (besides power, area, and time) in the forthcoming EDA tools. Considering this context, an accurate yet fast reliability analysis is needed at design time, to allow for instance, a gate-level comparison of different realizations of the same logic function and enable a reliability driven synthesis process.

To this end, various probabilistic analytical approaches for circuit reliability evaluation have been proposed, such as the Probability Element (ProxEl) method [2], the Probabilistic Transfer Matrices (PTM) formalism [3], the Probabilistic Gate Model (PGM) [4], [5], [6], the single-pass reliability analysis method [7], [8], the Signal Probability Reliability Analysis (SPRA) method [9], [10]. Only recently, probabilistic graphical models, have been applied in the context of IC reliability evaluation. Most prominently, Bayesian Networks (BNs) allows one to capture both temporal and spatial circuit decencies in a comprehensive manner, providing an exact and minimal probabilistic model for reasoning and inference in causal logic networks. The BN formalism has been exploited in [11], [12], [13] for reliability evaluation of small circuits. The preponderant corpus of past approaches exhibit high computational complexity for accurate reliability estimates, which cautions on using them for reliability estimation of medium and large circuits. Furthermore, hitherto reliability evaluation approaches customary posit a single value for a gate probability of failure. While benefiting from a relative simplicity of implementation, the single-probability approach may not suffice for accurate reliability assessment. Another aspect is related to the fact that the majority of previous approaches evaluate the reliability of a circuit starting from the gate level and furthermore, most of them rely on the assumption that all gates have the same reliability. However, accurate computation of a circuit reliability at gate-level is of foremost importance, since very small reliability estimation errors at the gate-level can severely impact the reliability evaluation of circuits comprising large numbers of gates [14], [15].

In light of the above we aim to provide a probabilistic inference framework for reliability assessment that is accurate enough and suitable for fast, large-scale circuit settings. Instead of using a single probability value to reflect a circuit reliability, in our view, a promising avenue toward a more appropriate approach to model the faulty circuits stochastic behavior, would be to consider a range of failure probabilities, i.e., a Probability Density Function (PDF). In this paper, we propose a PDF-based circuit reliability assessment framework, that yields a more accurate circuit reliability estimate by: (i) capturing the circuit topology, and its temporal and spatial correlations (e.g., reconvergent fan-outs), (ii) taking into account the reliability of individual gates, the workload (input vectors), environmental conditions variations, and various fault scenarios, (iii) employing variational inference, to derive the PDFs of a circuit primary outputs, given its primary inputs PDFs, and (iv) adapting the traditional variational inference approach to exploit the peculiarities of the probabilistic model afferent to logic circuits, for convergence speed improvements and thus applicability in large scale circuits settings. Our framework provides a kernel for fast and accurate reliability evaluation, which is crucial in enabling effective reliability-

aware design and logic synthesis, or run-time reliability based prediction/diagnosis analysis required in Dynamic Reliability Management frameworks [16].

The remaining of the paper is organized as follows: Section II presents the general framework formalism. The proposed reliability assessment framework is introduced in Section III. Section IV concludes the paper with a summary of this work.

II. General framework formalism

One common approach is to quantify the reliability of a device/gate/circuit in terms of its primary output(s) probability of failure (i.e., the probability of obtaining an erroneous logic/voltage level for the primary output(s)). For a better adherence to faulty circuit stochastic behavior, instead of using a single probability value for reliability assessment, we propose to employ a range of probabilities, i.e., a Probability Density Function (PDF). In order to obtain an accurate circuit reliability estimate, the reliability assessment methodology should take into account several aspects, such as: (i) the circuit topology and implicitly the temporal and spatial (e.g., reconvergent fan-outs) correlations, (ii) the reliability of individual circuit gates, which may be implemented using different logic styles, number of devices, and thus require individual treatment, (iii) the input vectors applied to the circuit (e.g., the logic/voltage level and the afferent probability of being in that level, for each primary input), as well as the correlations between these inputs, and (iv) the environmental conditions and the different fault types - with varying frequency of occurrence - to which the circuit can be exposed at run-time.

In light of the above considerations, the problem statement is formulated as follows: Given a circuit, with known topology and possibly layout, its workload (e.g., known primary inputs vectors logic/voltage levels and their associated PDFs), and an aggression profile (e.g., temperature, supply voltage, fault scenarios - fault types and their expected probabilities), one is interested to determine the PDF of obtaining erroneous logic/voltage levels for the circuit primary outputs.

To this end, we model the circuit as a directed acyclic graph, specifically as a Bayesian network. The graph nodes correspond to circuit gates and wires, and are represented by continuous random variables (PDFs). The graph arcs represent the causal relationships between the nodes, thus accounting inherently for the circuit reconvergent fan-outs. For instance a gate output is conditionally dependent on the values applied at its input. Conversely, the absence of an arc between two nodes, signifies their conditional independence (e.g., the state of one node does not directly depend on the state of the other node). To serve our purpose, we discriminate the graph nodes into the following three sets:

- E, the set of evidence nodes, which are associated to the PDFs of the circuit primary inputs;
- Y, the set of latent (hidden) nodes, which are associated to the PDFs of the circuit primary outputs. These are the PDFs that we would like to infer.
- X, the latent (hidden) intermediary nodes, which correspond to the remaining nodes PDFs.

Thus, given the PDFs of the evidence nodes from E (circuit primary inputs) and the prior PDFs of the intermediary nodes from set X (i.e., the initial belief about the gates/wires PDFs of failure), we are interested to infer the PDFs of the primary output nodes from set Y. In the following we shall coarsely describe the methodology to obtain the prior PDFs, followed by an outline of the primary outputs PDFs inference approach. These two steps augment the traditional synthesis approach to achieve design-for-reliability closure [17], as graphically illustrated in Figure 1. The prior PDFs reflect the initial belief

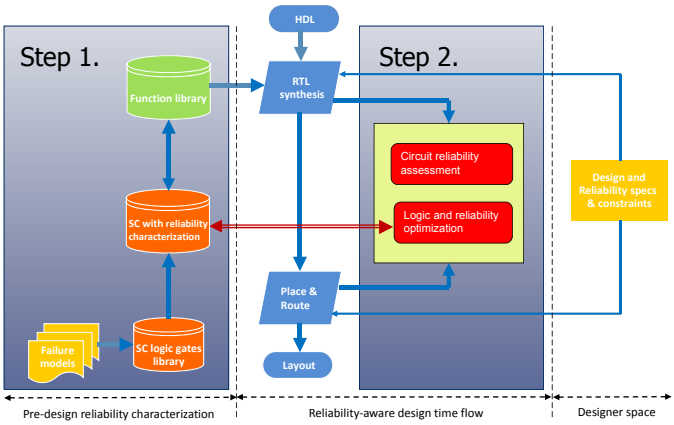

Fig. 1. Design-for-Reliability Closure.

regarding the PDF of failure for the case of the considered circuit aggression profile (i.e., for given fault scenario - fault types and probabilities of occurrence, and environmental conditions - temperature, supply voltage). These prior PDFs are obtained during the pre-design reliability characterization step, which consists of standard cells reliability characterization. Specifically, for several physical and electrical parameters variation (e.g., gate oxide thickness, threshold voltage, supply voltage, temperature), and failure models (e.g., stuck-at-fault, transient), which can transcend different architectural levels, the standard cell ancillary reliability information is obtained by spanning the probability failure space via a Monte Carlo analysis [18]. This pre-design step is performed only once for a specific technology. At design-time a gate prior PDF, is selected from the ensemble of ancillary PDFs attached to the gate under consideration during the pre-design reliability characterization step. The PDF selection is made based on the resemblance of the pre-design aggression profile (environmental, fault scenarios) to the current design-time aggression profile.

An outline of the proposed circuit reliability assessment - which is performed at design-time during the second step from Figure 1) - may now be given. Given the evidence nodes PDFs and the prior PDFs for the remaining nodes, we are interested in inferring the posterior PDFs of the output nodes Y. As concerns the probabilistic inference method, we opt for variational inference [19] [20], since it serves better for developing fast, potentially on-line algorithms in large-scale settings. The general idea of variational inference is to

recast the probabilistic inference problem into an optimization-based formulation, and express the posterior distribution of interest as the optimization problem solution. In order to provide a means of approximating the posterior distribution of interest (i.e., the PDF of the circuit primary outputs), the optimization problem is relaxed. Such relaxations can be carried out in various ways, either by approximating the function to be optimized, or by approximating the set over which the optimization is performed [20]. One common approach is to approximate the posterior distribution of interest (i.e., the true posterior) with a family of distributions from which the distribution that is closest to the true posterior is seeked. The measure of closeness between two distributions is commonly the Kullback-Leibler (KL) divergence [20]. As the true posterior is unknown, and in consequence also the KL divergence, the optimization of the KL divergence is then usually cast as an optimization of another quantity (i.e., the evidence lower bound, that is the lower bound on the logarithm of the marginal probability of the observations), which does not directly depend on the true posterior. The resulting optimized distribution constitutes the approximated posterior of interest. Figure 2 coarsely illustrates the variational inference basic principle.

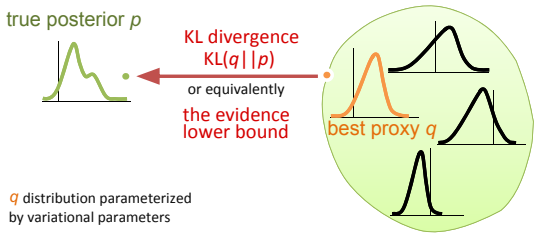

Fig. 2. Variational Inference Concept.

III. Reliability assessment framework via variational inference

In Subsection III-A, the optimization objective, i.e., the evidence lower bound, is derived. The commonly employed evidence lower bound bound in variational inference, denoted as $\mathcal{L}_{XY}(q)$, is a function of both sets of hidden variables (the intermediary nodes X and output nodes Y). However, for our circuit settings, this bound is not suitable, because of its associated convergence speed drawbacks and its strong simplifying assumptions (specifically, it neglects the correlations between the intermediary nodes X and the output nodes Y). For our purpose, we analytically integrate one set of hidden variables (e.g., Y), and thus obtain the bound as a function of only one set of variables $\mathcal{L}_X(q)$. The benefits of this approach are improved convergence speed and a more accurate estimate of the true posterior, as the bound to be optimized in this case, $\mathcal{L}_X(q)$, is tighter than the classical $\mathcal{L}_{XY}(q)$ [21], [22]. As simplifying assumption, we suppose that the output nodes Y are conditionally independent among them, given the PDFs of the evidence nodes E, and the PDFs of the inferred nodes Y. We note that for circuit specific scenarios, this assumption is always satisfied, as each circuit output has its own cone of influence and can be computed independently from the rest of the circuit outputs, given all primary inputs, and all intermediary nodes. Subsection III-B and III-C address the optimization algorithm, i.e., a nonlinear Conjugate Gradient (CG) method, which exploits the geometry of the curved q distributions space. Specifically, (i) instead of the classical gradient, the natural gradient is employed, as it gives the direction of the highest increase in the optimization objective $\mathcal{L}_X(q)$, and (ii) the CG method is performed along curves, instead of straight lines. As the space of all q distributions is curved, the commonly employed flat space Euclidian approach can result in a sequence of updates which significantly deviate from the true curve whose end-point is the seeked optimization solution. The employed CG method exhibits a global convergence property and superlinear convergence speed. The optimization solution yields the best estimate of the posterior, and thus the circuit primary outputs reliability.

A. The evidence lower bound

Subsequently, we shall employ the following notations:

- $\mathbb{E}_{g(x)}[f(x)] = \int g(x) f(x)\, dx$ denotes the expectation of distribution f under distribution g;
- $KL[f(x)||g(x)] = -\int f(x) \log \frac{g(x)}{f(x)} dx$ denotes the KL divergence between the distributions f and g.

In variational inference, we would like to minimize the KL divergence from the approximated distribution (i.e., the variational distribution over the latent variables/parameters X and Y), $q(X,Y)$ and the true posterior, $p(X,Y|E)$. To this end, a lower bound on the logarithm of the marginal likelihood (i.e., the model evidence E), $\log p(E)$ is derived, by employing Jensen's inequality ($\log \mathbb{E}_g[f(a)] \geq \mathbb{E}_g[\log f(a)]$, $\forall a$ random variable, f and g distributions), as follows:

$$\begin{aligned}
\log p(E) &= \log \int p(E,X,Y)\, dX\, dY \\
&= \log \int p(E,X,Y) \frac{q(X,Y)}{q(X,Y)}\, dX\, dY \\
&\geq \int q(X,Y) \log \frac{p(E,X,Y)}{q(X,Y)}\, dX\, dY \\
&= \mathbb{E}_{q(X,Y)}[\log p(E,X,Y)] - \mathbb{E}_{q(X,Y)}[\log q(X,Y)] \\
&\triangleq \mathcal{L}_{XY}(q). \qquad (1)
\end{aligned}$$

Since $\log p(E) = \mathcal{L}_{XY}(q) + KL[q(X,Y)\,||\,p(X,Y|E)]$, minimizing the KL divergence (which is ≥ 0) is equivalent to maximizing the lower bound $\mathcal{L}_{XY}(q)$, as illustrated in Figure 3. In the commonly and most straightforward variational inference (Variational Bayes (VB)) framework, the family of approximate distributions are restricted to a tractable family, by positing that the variational distribution q factorizes over the latent variables, e.g., $q(X,Y) = q(X)\, q(Y)$, i.e., X and Y are conditionally independent, where each factor of q has a free functional form. The objective is to determine the variational distribution which maximizes the evidence lower bound

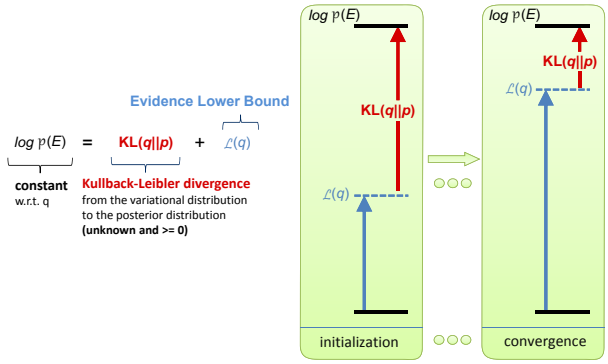

Fig. 3. Evidence Lower Bound optimization.

$\mathcal{L}_{XY}(q)$. The obtained optimized distribution q^* constitutes the approximation of the true posterior distribution p over the latent variables/parameters. This approach convergence speed can be very slow, prohibiting its utilization in large scale settings. Furthermore, the strong independence assumptions between the latent variables may place unrealistic or questionable factorizations, disregarding important correlations among the graph nodes, and as a consequence resulting in a significant bias of the posterior distribution estimate.

Rather than assuming the factorization of q over the latent variables, an approach for a closer adherence to reality would be to analytically marginalize (i.e., to integrate out) a subset of the latent variables, and thus to perform the optimization of the evidence lower bound only with respect to the remaining latent variables. Besides the improvement in convergence speed, this approach also provides a more accurate estimate of the true posterior distribution with a tighter lower bound, as achieved in several other probabilistic models contexts [22], [21], [23]. Integrating analytically the latent variables Y, a lower bound on the model evidence with respect to the variables X can be derived by applying the Jensen inequality, in a similar manner to the derivation in (1):

$$\begin{aligned}\log p(E) &\geq \int q(X) \log \frac{\int p(E,X,Y)\,dY}{q(X)}\,dX \\ &= \int q(X) \log \frac{\int p(E,X|Y)p(Y)\,dY}{q(X)}\,dX \\ &= \int q(X) \log \frac{\mathbb{E}_{p(Y)}\left[p(E,X|Y)\right]}{q(X)}\,dX \\ &= \mathbb{E}_{q(X)}\left[\log \mathbb{E}_{p(Y)}\left[p(E,X|Y)\right]\right] - \mathbb{E}_{q(X)}\left[\log q(X)\right] \\ &\triangleq \mathcal{L}_X(q). \end{aligned} \tag{2}$$

For tractability reasons, a first-order (linear) approximation [21] of $\mathbb{E}_{q(Y)}\left[g(X)\right]$ with $g(\mathbb{E}_{q(Y)}\left[X\right])$ is applied to (2), yielding the following expression for the bound $\mathcal{L}_X(q)$:

$$\mathcal{L}_X(q) = \log \mathbb{E}_{p(Y)}\left[e^{\mathbb{E}_{q(X)}\left[\log \frac{p(E,X|Y)}{q(X)}\right]}\right]. \tag{3}$$

The expression of the optimal approximated posterior is then given by:

$$q^*(Y) = \frac{e^{\int q(X) \log \frac{p(E,X|Y)\,p(Y)}{q(X)}\,dX}}{C}$$

$$= \frac{p(Y)\; e^{\mathbb{E}_{q(X)}\left[\log \frac{p(E,X|Y)}{q(X)}\right]}}{C}, \tag{4}$$

where C is the normalization constant, and is obtained by integrating the denominator in (4) with respect to Y.

We index the distribution $q(X)$ by a set of variational parameters θ, and seek the configuration of θ which optimizes the lower bound $\mathcal{L}_X(q)$, rendering therefore q^* which is the closest to the true posterior p.

The convergence of the algorithm can be monitored by evaluating if the difference between the previous lower bound (for the previous θ) and its current value (for the current update of θ) is sufficiently small. The optimal distribution q^* for the parameters θ at bound convergence is given by (4), with $p(Y)$, the prior distribution of Y. We note that, the lower bound $\mathcal{L}_{XY}(q)$, depends on two sets of variables, whose updates are interlocked (the bound optimization with respect to each set is performed while holding the other set fixed). In our case however, the lower bound $\mathcal{L}_X(q)$, is expressed only as a function of one set (the variational parameters) and as such the lower bound optimization is performed only with respect to one set of variables. One may further note that the only assumed factorization in this approach is among the output latent nodes Y, conditioned by the evidence nodes E and the approximated nodes X. This can be easily determined using an independence criterium. The d-separation topological criterium [24] determines whether a set of nodes A are conditionally independent of another set of nodes B given a set of evidence nodes V. In particular the set of nodes A is d-separated from the set of nodes B by the set V iff at least one of the following three axioms holds true: (i) every undirected path contains a sequential node in V ($\longrightarrow V_i \longrightarrow$), (ii) every undirected path contains a divergent node in V ($\longleftarrow V_j \longrightarrow$), and (iii) every undirected path contains a convergent node ($\longrightarrow T_k \longleftarrow$) such that neither the convergent node, nor any of its descendants are in V. One can observe that for circuit specific scenarios, as is our case, the independence among the output nodes conditioned by the rest of the circuit nodes (input nodes - E, and the inferred nodes X) is always satisfied, since case (i) always holds true.

B. The gradient of the lower bound $\mathcal{L}_X(q)$

Subsequently, we concern ourselves with finding the best configuration of the variational parameters θ which optimizes the objective function $\mathcal{L}_X(q)$. In [25], the authors proved that the coordinate ascent algorithm is equivalent to the natural gradient method. Therefore, an update via taking a step in the steepest direction in the space of variational parameters θ, using a Riemannian metric (e.g., the natural gradient [26] which accounts for the space information geometry), is equivalent to performing a coordinate ascent update. To this end, we evaluate the gradient of the lower bound with respect to the variational parameters, which gives us the direction of the coordinate ascent update for the variational parameters θ.

The space of all probability distributions $S = \{q(X|\theta)\}$ is not Euclidian with an orthonormal coordinate system θ, but a curved space, namely a Riemannian manifold. In such spaces,

the shortest distance between two points does not correspond anymore to an Euclidian line, but to a geodesic (i.e., a curve) following the space curvature. The immediate consequence is that the steepest descent direction along a manifold path (as given by the iterative updates of the parameters θ) is different than the steepest descent direction in the classical Euclidian parameter space. Specifically, in the case of statistical manifolds, the natural gradient corresponds to the direction which maximizes the objective function $\mathcal{L}_X(\theta)$, such that the KL divergence $KL\left[q(X|\theta)\,||\,q(X|\theta+\delta\theta)\right]$ is not changed through the optimization (otherwise stated, the natural gradient gives the direction of the highest increase in the objective function, for the smallest change in the KL divergence):

$$\tilde{\nabla}\mathcal{L}_X(q(\theta)) = \underset{\delta\theta}{\operatorname{argmax}}\ \mathcal{L}_X(q(X|\theta+\delta\theta))$$
$$s.t. \quad KL\left[q(X|\theta)\,||\,q(X|\theta+\delta\theta)\right] \leq \epsilon. \tag{5}$$

It follows then, the expression of the natural gradient:

$$\tilde{\nabla}\mathcal{L}_X(q(X|\theta)) = W^{-1}(\theta)\cdot\nabla\mathcal{L}_X(q(X|\theta)), \tag{6}$$

where W is the Fisher information matrix with $w_{ij}(\theta) = \mathbb{E}_q\left[\nabla_{\theta_i}\log q(X|\theta)\cdot\nabla^T_{\theta_j}\log q(X|\theta)\right]$. A chief advantage is the KL-invariance with respect to the re-parametrization of the family of variational distributions $q(X|\theta)$, i.e., the update direction depends only on $q(X|\theta)$, and not on a particular transformation of the θ parameters; the followed optimization trajectory in the parameters space is the same, regardless of the re-parametrization of θ. As a result, as opposed to the vanilla gradient, the natural gradient exhibits fast isotropic convergence properties. Additionally, it circumvents the slow or early convergence proneness of the vanilla gradient, avoiding over-aggressive steps on ridges and too small steps on plateaus, and hence being able to cope in an efficient manner with ill-shaped $\mathcal{L}_X(\theta)$.

C. The optimization algorithm

As concerns the nonlinear optimization techniques, we opt to employ the nonlinear Conjugate Gradient (CG) method, due to its algorithmic simplicity, superlinear (at least quadratically) convergence, and suitability for large scale optimization scenarios. In this method, one determines first the search direction, H_k, then computes a step size, α (as a result of a line search, or set adaptively), and finally one updates the parameters in the search direction using $\theta_k = \theta_{k-1} + \alpha\, H_k$. In the flat Euclidean space, evaluating the search direction amounts to computing:

$$H_k = -G_k + \gamma_k\, H_{k-1}, \tag{7}$$

with $G_k = \tilde{\nabla}\mathcal{L}_X(q(\theta_k))$, denoting the Riemannian gradient of the objective function, and with commonly employed γ_k variants such as Fletcher-Reeves, Polak-Ribiere, and Hestenes-Stiefel [27]. The majority of previous gradient-based statistical inference approaches employ the flat space approximation of the conjugate gradient [28][29] [30], based on the rationale that the minimization of functions on a Riemannian manifold is locally equivalent to the minimization on an Euclidian space (since every Riemannian manifold can be isometrically embedded in an Euclidean space). However, as the statistical space is a curved manifold, most of the Euclidean space operations become undefined. For instance, the minimization of $\mathcal{L}$ is not performed any longer along straight lines but along geodesics (i.e., the shortest curves which connect two points on the manifold). Another example is the additive rule employed for updating the search direction, which makes no longer sense in the Riemannian manifold (it has no geometrical meaning), and as a consequence, the resulting sequence of parameters updates can significantly deviate from the true geodesic whose end-point is the optimal solution. Since H_{k-1} does not reside in the same tangent space as G_k, it follows that it needs to be transported to the tangent space of θ_k to enable the addition of the two vectors. This transport operation is known as parallel translation, or as vector transport, which is basically a computationally efficient relaxation of parallel translation. Figure 4 a) illustrates the CG algorithm in Euclidean space, and Figure 4 b) depicts its counterpart in Riemannian space. When

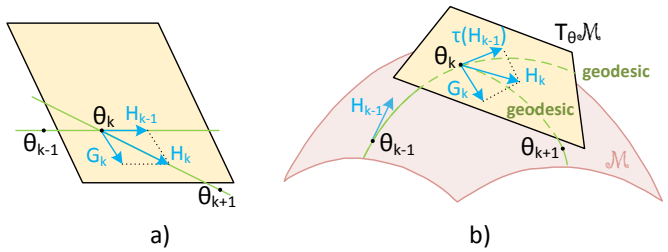

Fig. 4. The CG in: a) flat (Euclidean) space, and in b) curved (Riemannian) space.

compared to the flat space approximation, the Riemannian conjugate gradient provides significant advantages in terms of convergence speed and accuracy of solution.

In the case of statistical Riemannian manifolds, (7) is generalized to:

$$H_k = -G_k + \gamma_k\, \tau(H_{k-1}), \tag{8}$$

where $\tau(H_{k-1})$ defines the parallel translation of H_{k-1} (see Figure 4 b)). One iteration of the CG algorithm can be outlined as follows: at iteration k, the Riemannian gradient G_k of the objective function is evaluated, and the new search direction H_k for geodesic minimization is conjugate to the gradient and is evaluated to be a combination of the previous search direction H_{k-1} and the current Riemannian gradient at step G_k; finally a step is made in the direction of H_k to obtain θ_{k+1}, which is the minimum of the objective function in the direction of H_{k-1}. In Algorithm 1, we present the formalism of the CG method on Riemannian manifolds. In Algorithm 1, $\tau_{\alpha_k H_k}(H_k)$ is the vector transport associated with the differentiated retraction R [31], which is scaled when the norm of the previous search direction is increased [32] (since a parallel translation of a vector should preserve its norm). The step size α_k is chosen such that it obeys the strong Wolfe conditions [27] [33], which prevent the step length to be excessively short.

Algorithm 1 Riemannian conjugate gradient method for optimizing the lower bound $\mathcal{L}$

Input: the objective function $\mathcal{L} : \mathcal{M} \longrightarrow \mathbb{R}$

Output: global minimizer of $\mathcal{L}$

Set an initial point θ_0 on $\mathcal{M}$
Set the initial search direction $H_0 = -G_0$
$k = 0$
repeat

- Calculate the step length $\alpha_k > 0$
- Take a minimizing step by setting: $\theta_{k+1} = R_{x_k}(\alpha_k H_k)$
- Calculate γ_{k+1}
- Compute the new search direction: $H_{k+1} = -G_{k+1} + \gamma_{k+1}\tau^R_{\alpha_k H_k}(H_k)$

until θ_{k+1} sufficiently minimizes the objective $\mathcal{L}$

The global convergence of the Riemannian conjugate gradient was proved in [33]. As concerns the speed of convergence, using the rescaled vector transport proposed in [32], result in superlinear convergence of the sequence $\{\theta_k\}$ to the global minimizer.

IV. Conclusions

In this paper we proposed a PDF based IC reliability assessment framework. For a closer adherence to a faulty circuit stochastic behavior, we employed a distribution of probabilities, instead of relying on a single probabilistic value to reflect the reliability status of an IC. The framework is based on a variational inference method, and exploits the geometry of the statistical manifold to yield a fast and scalable reliability assessment approach, which can be potentially integrated in reliability aware synthesis tools. The numerical assessment of the proposed approach for medium and large scale combinational circuits is under progress, and the final manuscript will include its results.

Acknowledgement

This work was supported by the Seventh Framework Programme of the European Union, under Grant Agreement number 309129 (i-RISC project).

References

[1] M. Pecht, *Product Reliability, Maintainability, and Supportability Handbook*. 2nd edition, CRC Press, 2009.

[2] G. Horton, "A new paradigm for the numerical simulation of stochastic Petri nets with general firing times," in *Proceedings of the European Simulation Symposium (ESS)*, 2002.

[3] K. Patel, I. Markov, and J. Hayes, "Evaluating circuit reliability under probabilistic gate-level fault models." in *Proceedings of the International Workshop on Logic Synthesis*, 2003, pp. 59–64.

[4] J. Han, E. R. Taylor, J. B. Gao, and J. A. B. Fortes, "Faults, error bounds and reliability of nanoelectronic circuits," in *IEEE International Conference on Application-Specific Systems, Architecture Processors, 51*, 2005, pp. 247–253.

[5] E. Taylor, J. Han, and J. Fortes, "Towards accurate and efficient reliability modeling of nanoelectronic circuits," in *6th IEEE Conference on Nanotechnology, 1*, 2006, pp. 395–398.

[6] J. Han, H. Chen, E. Boykin, and J. A. B. Fortes, "Reliability evaluation of logic circuits using probabilistic gate models," in *Microelectronics Reliability, 51*, 2011, pp. 468–476.

[7] M.R. Choudhury, and K. Mohanram, "Reliability analysis of logic circuits." in *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, 28(3)*, 2009, pp. 392–405.

[8] S.J. Seyyed Mahdavi, and K. Mohammadi, "SCRAP: sequential circuits reliability analysis program." in *Microelectronics Reliability, 49(8)*, 2009, pp. 924–933.

[9] D.T. Franco, M.C. Vasconcelosa, L. Navinera, and J.-F. Navinera, "Reliability analysis of logic circuits based on signal probability," in *Proceedings of the 15th IEEE International Conference on Electronics, Circuits and Systems (ICECS)*, 2008, pp. 670–673.

[10] ——, "Signal probability for reliability evaluation of logic circuits," in *Microelectronics Reliability, 48(8)*, 2008, pp. 1586–1591.

[11] T. Rejimon and S. Bhanja, "Time and space efficient method for accurate computation of error detection probabilities in VLSI circuits," in *IEE Proceedings on Computers and Digital Techniques, 152(5)*, 2005, pp. 679–685.

[12] ——, "Probabilistic error model for unreliable nano-logic gates," in *6th IEEE Conference on Nanotechnology, 1*, 2006, pp. 47–50.

[13] W. Ibrahim and V. Beiu, "Using Bayesian networks to accurately calculate the reliability of Complementary Metal Oxide Semiconductor gates," in *IEEE Transactions on Reliability, 60(3)*, 2011, pp. 47–50.

[14] G. R. Roelke, R. O. Baldwin, and D. Bulutoglu, "Analytical models for the performance of von Neumann multiplexing," in *IEEE Transactions on Nanotechnology, 6(1)*, 2007, pp. 75–89.

[15] J. Han, and P. Jonker, "A system architecture solution for unreliable nanoelectronic devices," in *IEEE Transactions on Nanotechnology, 1(4)*, 2002, pp. 201–208.

[16] Y. Wang, "Aging assessment and reliability aware computing platforms."

[17] S. Cotofana, A. Schmid, Y. Leblebici, A. Ionescu et al., "CONAN - a design exploration framework for reliable nano-electronics architectures," in *Proceedings of the 16th IEEE International Conference on Application-Specific Systems, Architecture Processors*, 2005, pp. 260–267.

[18] M. Stanisavljevic, A. Schmid, and Y. Leblebici, "Fault-tolerance of robust feed-forward architecture using single-ended and differential deep-submicron circuits under massive defect density," in *Proceedings of the International Joint Conference on Neural Networks*, 2006, pp. 2771–2778.

[19] C. M. Bishop, *Pattern recognition and machine learning*. Springer Science+Business Media, LLC., 2006.

[20] M. J. Wainwright and M. I. Jordan, "Graphical models, exponential families, and variational inference," in *Foundations and Trends in Machine Learning 1, 1-2*, 2008, pp. 1–305.

[21] J. Sung, Z. Ghahramani, and S. Bang, "Latent-space variational Bayes," in *Pattern Analysis and Machine Intelligence, 30(12)*, 2008, pp. 2236–2242.

[22] N. King, and N. D. Lawrence, "Fast variational inference for Gaussian process models through KL correction," in *17th European Conference on Machine Learning*, 2006, pp. 270–281.

[23] M. Lazaro-Gredilla, and M. K. Titsias, "Variational heteroscedatic Gaussian process regression," in *Proceedings of the International Conference on Machine Learning*, 2011.

[24] T. Verma, and J. Pearl, "Causal networks: semantics and expressiveness," in *Proceedings of the 4th Workshop on Uncertainty in AI*, 1988, pp. 352–359.

[25] M. Sato, "Online model selection based on the variational Bayes," in *Neural Computation 13(7)*, 2001, pp. 1649–1681.

[26] S. Amari, "Natural gradient works efficiently in learning," in *Neural Computation 10(2)*, 1998, pp. 251–276.

[27] J. Nocedal, and S. J. Wright, *Numerical Optimization*. 2nd ed, Springer, 2006.

[28] M. Kuusela, T. Raiko, A. Honkela, and J. Karhunen, "A gradient-based algorithm competitive with variational Bayes EM for mixture of Gaussians," in *Proceedings of International Joint Conference on Neural Networks*, 2009, pp. 1688–1695.

[29] A. Honkela, T. Raiko, M. Kuusela, M. Tornio and J. Karhunen, "Apprximate Riemannian conjugate gradient learning for fixed-form variational Bayes," in *Journal of Machine Learning Research 11*, 2010, pp. 3235–3268.

[30] M. D. Hoffman, D. M. Blei, C. Wang, and J. Paisley, "Stochastic variational inference," in *Journal of Machine Learning Research 14*, 2013, pp. 3235–3268.

[31] P. A. Absil, R. Mahony, R. Sepulchre, *Optimization algorithms on matrix manifolds*. Princeton University Press, 2008.
[32] H. Sato, and T. Iwai, "A new, globally convergent Riemannian conjugate gradient method," 2013.
[33] W. Ring, and B. Wirth, "Optimization methods on Riemannian manifolds and their applications to shape spaces," in *SIAM J. Optim. 22(2)*, 2012, pp. 281–285.