Chapter 5: Correlative Derivative

November 9, 2003

1 The Metric theory

Suppose, we are given a circuit function C[$\mathbf{X}$] and we are trying to optimize fault-coverage. Instead of optimizing fault-coverage, we may be, minimizing, measurable metrics, like distance at which a fault is unable to propogate.

Now, suppose, we are given a combinational circuit. The circuit is memoryless and static. Now, to manipulate the circuit, we supply it with an input vector and run it. Now, suppose, if we get the same metric value, independent of the time the input vector, is presented to the circuit, then the following is true of that circuit.

Note: It is important to note this point that the metric is memoryless. Suppose, the metric, does depend on the sequence, of input vectors, our following arguements are valid, over the local time frame.

Definition 1

$$\mathrm{M}(\mathbf{X}) = \sum_{i}^{2^n} \mathrm{M_i} \prod_{j} \mathrm{Y_{ij}} \tag{1}$$

where $\prod_i \mathrm{Y_i}$ *is the term the produces the effect* $\mathrm{M_i}$.

For example, for an OR gate, we could have

$$\mathrm{M_{01}} x_0(1 - x_1) + +\mathrm{M_{10}}(1 - x_0)x_1 + \mathrm{M_{11}} x_0 x_1 \tag{2}$$

So, when the input vector, is 01, we get the output metric, $\mathrm{M_{01}}$. Now, one can prove that given a function M($\mathbf{X}$), and the pdf(x_i), or the input probabilities, then the statistical average of the metric is given as

$$\left[\prod_i \int_{-\infty}^{\infty} dx_i \mathrm{pdf}(x_i)\right] \mathrm{M}(\mathbf{X}) \tag{3}$$

Now, since the $\mathrm{pdf}(x_i) = (1 - p_i)\delta(1 - x_i) + p_i\delta(x_i)$, Parker muclusky rules $p_i^2 = p_i$ and $(1 - p_i)pi = 0$ all apply. So, what becomes, true,is that the second order, third order, 4th order, etc, effects of an input x_i are not present. So, the metric function can be written as

$$\mathrm{M}(\mathbf{X}) = m_0 + \sum_i m_i x_i + \sum_{i \neq j} m_{ij} x_i x_j + ... \tag{4}$$

Given that, the metric function can also be written as

$$\mathrm{M}(\mathbf{X}) = \mathrm{M}(x_i = 0) + x_i \frac{\partial \mathrm{M}}{\partial x_i}(\mathbf{X}) \tag{5}$$

From the fact, that second order effects are not present, one can deduce that x_i and $\frac{\partial \mathrm{M}}{\partial x_i}(\mathbf{X})$ behavely indpendently,for any metric.

Take correlation function $\mathrm{Cor}(\mathrm{M}(\mathbf{X}), x_i)$

$$\mathrm{Corr}(\mathrm{M}, x_i) = \frac{\langle \mathrm{M} x_i \rangle - \langle \mathrm{M} \rangle \langle x_i \rangle}{\sigma_M \sigma_{xi}} \tag{6}$$

Evatuating, numerator part of the expression,

$$\begin{aligned}
&= \langle \mathrm{M} x_i \rangle - \langle \mathrm{M} \rangle \langle x_i \rangle && (7)\\
&= \left\langle \left(\mathrm{M}(x_i = 0) + x_i \frac{\partial \mathrm{M}}{\partial x_i} \right) x_i \right\rangle - \left\langle \mathrm{M}(x_i = 0) + x_i \frac{\partial \mathrm{M}}{\partial x_i} \right\rangle \langle x_i \rangle && (8)\\
& && (9)
\end{aligned}$$

Since $\mathrm{M}(x_i = 0)$ and x_i are indepedent and $\frac{\partial \mathrm{M}}{\partial x_i}$ and x_i are independent,

$$\begin{aligned}
&= \langle \mathrm{M}(x_i = 0) \rangle \langle x_i \rangle + \left\langle \frac{\partial \mathrm{M}}{\partial x_i} \right\rangle \langle x_i^2 \rangle - \langle \mathrm{M}(x_i = 0) \rangle \langle x_i \rangle - \left\langle \frac{\partial \mathrm{M}}{\partial x_i} \right\rangle \langle x_i \rangle^2 && (10)\\
&= \left\langle \frac{\partial \mathrm{M}}{\partial x_i} \right\rangle \left(\langle x_i^2 \rangle - \langle x_i \rangle^2 \right) && (11)\\
&= \left\langle \frac{\partial \mathrm{M}}{\partial x_i} \right\rangle \sigma_{xi}^2 && (12)\\
&= \sigma_{xi}^2 \frac{\partial}{\partial x_i} \langle \mathrm{M} \rangle && (13)
\end{aligned}$$

Now,

$$-\mathrm{Corr}(\mathrm{M}, x_i) = -\frac{\sigma_{xi}}{\sigma_M} \left\langle \frac{\partial \mathrm{M}}{\partial x_i} \right\rangle \tag{14}$$

$$-\mathrm{Corr}(\mathrm{M}, x_i) = -\sigma_{xi} \left\langle \frac{\partial}{\partial x_i} \left(\frac{\mathrm{M} - \mu_M}{\sigma_M} \right) \right\rangle \tag{15}$$

Now, as you know, already, if $\left\langle \frac{\partial \mathrm{M}}{\partial x_i} \right\rangle = 0$, the metric is locally a maxima or a minima. However, what is counter-intuitive, is that, for that to happen, the correlation between the the input x_i and the metric, has to be 0. That result make strike you, as suprising. But, in fact, there is a perfectly resonable explanation, as to why this is the case. Now, we can expand the derivative, as follows

$$\left\langle \frac{\partial \mathrm{M}}{\partial x_i} \right\rangle = \mathrm{M}(x_i = 1) - \mathrm{M}(x_i = 0) \tag{16}$$

Now, when that happens to be 0, $\mathrm{M}(x_i = 1) = \mathrm{M}(x_i = 0)$. If M_i is positive, or can be made positive, by adding some finite constant c. What becomes true is the following.

That is to say, statiscally, there is no difference, when x_i is a 1 or a 0. Now, notice that when $\mathrm{M}(x_i = 1)$ or $\mathrm{M}(x_i = 0)$, it is controlled by every input, other than x_i.

So, what it actually means is that, for the metric to be maximal or minimal, the input probability, should make the circuit behave as though, the inputs are independent of each other. To use Agarwal's terminology, there should be minimal turbulence.

1.1 Gradient Descent

Now, if we correct x_i by the correlation coefficient, we are in fact, just doing a simple gradient descent. However, since the method is statistical, we end up with a method, that simulated annealing, implicitly, and thus, are able to overcome, problems of being stuck in local maximas or minimas.

1.2 Metric and $\langle \mathrm{lt} \rangle$

In the previous chapter, we introduced terms $\langle lt \rangle$, which is average of the log of each term. From the definition of the metric,

$$\mathrm{M}(\mathbf{X}) = \sum_i^{2^n} \mathrm{M_i} \prod_j Y_{ij} \tag{17}$$

we can see a direct relationship between metric function and $\langle lt \rangle$. Take each term in the above sum and suppose we take the log of that

$$\log \left(\mathrm{M_i} \prod_j Y_{ij} \right) = \log \mathrm{M_i} + \sum_j \log Y_{ij} \tag{18}$$

The distribution of that, is gaussian, and the expected value of that, is given by $\langle \log \mathrm{M_i} \rangle + \langle \log \prod_i Y_i \rangle = \langle \log \mathrm{M_i} \rangle + \langle \mathrm{lt} \rangle$. And there is an explicit relationship between the variance of $\log \mathrm{M_i}$ and σ_{lt}.

2 Spectral Test Set

Now, one can prepare a spectral test set,that changes in time, as follows. Suppose $[x_{\mathrm{k0}}, x_{\mathrm{k1}}, x_{\mathrm{k2}}, ..., x_{\mathrm{kn}}]$ is a real weightset, in time instant k. Then, one can form a binary input vector y_{ki}, by taking,

$$y_{\mathrm{ki}} = \mathrm{rand}() < x_{\mathrm{ki}} \tag{19}$$

This way of doing business of keeping a real testset, has several advantages over keeping track of the sequences of binary vectors themselves. For one, because of the randomness, it allows for, it incorporates a variation of the simulated annealing technique,which has an advantage of overcoming, local maximas and minams. Secondly, suppose, we are able to take the fourier transform/wavelet transform of the input spectral set, we can decompose them into orthogonal components, and correlate, and isolate temporal spectra that is both present in the metric , and the input variables.

3 Spectral Testing

Previous work by G. Giani, S. Sheng, M. Hsaio and V.D.Agarwal has shown the viability of spectral testing methods using Haradmard transform. In the following sections, we will try to provide a general framework for spectral testing.

3.1 Basis

We start by assuming that the relavent spectra, are computed by the orthormal basis $\phi_n(t)$. Othornormal basis have the following property

$$\langle \phi_i^*(t)\phi_j(t)\rangle = \delta_{ij} \tag{20}$$

Now, without the loss of generality, we can write the input vector sequence as

$$x_i = \sum_j a_{ij}\phi_j(t) \tag{21}$$

where $0 \leq \sum_j |a_{ij}| \leq 1$ by virtue of x_i being a binary random variable.See previous chapter for proofs.
Now, suppose, we are given a time-based metric function $\mathrm{M}(\mathbf{X}, t)$. Without the loss of generality, we can write that function in the basis as

$$\mathrm{M}(\mathbf{X}, t) = \sum_k b_k\phi_k(t) \tag{22}$$

3.2 Results

Now, we will prove a very interesting result. Given that M is function of binary variables x_i, without the loss of generality, the function is expandable as

$$\mathrm{M}(\mathbf{X},t) = m_0(t) + \sum_i m_i(t)x_i + \sum_{i\neq j} m_{ij}(t)x_ix_j + \tag{23}$$

$$\sum_{i\neq j\neq k} m_{ijk}(t)x_ix_jx_k + ... \tag{24}$$

Also can be written as,

$$\mathrm{M}(\mathbf{X},t) = \mathrm{M}(\mathbf{X}|x_i=0,t) + x_i(t)\frac{\partial \mathrm{M}}{\partial x_i}(\mathbf{X},t) \tag{25}$$

Now, expanding x_i we, would get the following,

$$\mathrm{M}(\mathbf{X},t) = m_0(t) + \sum_i\sum_p m_i(t)a_{ip}\phi_p(t) + \sum_{i\neq j}\sum_p\sum_q m_{ij}(t)a_{ip}a_{jq}\phi_p(t)\phi_q(t) + ... \tag{26}$$

Or equivalently,

$$\mathrm{M}(\mathbf{X},t) = \mathrm{M}(\mathbf{X}|x_i=0,t) + \left[\sum_i a_{ip}\phi_p(t)\right]\frac{\partial \mathrm{M}}{\partial x_i}(\mathbf{X},t) \tag{27}$$

Now suppose, we differentiate the above by a_{ie}.

$$\frac{\partial \mathrm{M}}{\partial a_{ie}} = \phi_e(t)\frac{\partial \mathrm{M}}{\partial x_i}(\mathbf{X},t) \tag{28}$$

Suppose, we try to calculate, $\frac{\partial}{\partial a_{i0}}\langle \mathrm{M}\phi_e(t)\rangle$. We will see the following

$$\frac{\partial\left(\mathrm{M}\phi_e(t)\right)}{\partial a_{i0}} = \phi_e(t)\frac{\partial \mathrm{M}}{\partial x_i}(\mathbf{X},t) \tag{29}$$

In other words,

$$\left\langle\frac{\partial \mathrm{M}}{\partial a_{ie}}\right\rangle = \left\langle\frac{\partial\left(\mathrm{M}\phi_e(t)\right)}{\partial a_{i0}}\right\rangle \tag{30}$$

which is the same as

$$\frac{\partial\langle \mathrm{M}\rangle}{\partial a_{ie}} = \frac{\partial\langle \mathrm{M}\phi_e(t)\rangle}{\partial a_{i0}} \tag{31}$$

since a_{ie} and a_{i0} are constants of the simulation and does not vary in time, in that run. The RHS is true, because we are dealing with boolean random variables, and they do not have second order effects. Since $b_k = \langle \mathrm{M}\phi_k(t)\rangle$, we have a very interesting relationship,

$$\frac{\partial\langle \mathrm{M}\rangle}{\partial a_{ik}} = \frac{\partial b_k}{\partial a_{i0}} \tag{32}$$

It could be interpreted as follows. Suppose, we are trying to find spectral component ϕ_k in x_i that maximizes the average M, it would be equivalent to maximimizing the spectral component ϕ_k in M by manipulating a_{i0}.

3.3 Model of b_k: Output Component Metric

Definition 2

$$\mathrm{M}(\mathbf{X},t) = \sum_k \mathrm{M_k}\left(\mathbf{X}(t)\right)\phi_k(t) \tag{33}$$

where we model $\mathrm{M_k}(\mathrm{X})$, as

$$\mathrm{M_k}(\mathbf{X}) = \sum_i^{2^n} \mathrm{B_{ki}} \prod_j \mathrm{Y_{ij}} \tag{34}$$

One way to interpret this model is that, every time vector $\prod_j \mathrm{Y_{ij}}$, is presented at the input at time instant t, it contributes $\mathrm{B_{ki}}\phi_k(t)$ to metric M. For example, take the metric, for a 2-bit circuit,

$$\begin{aligned}\mathrm{M}(\mathbf{X},t) &= \mathrm{B}_{1,01}(1-x_0)x_1\phi_1(t) + \mathrm{B}_{1,11}x_0x_1\phi_1(t) + \mathrm{B}_{2,10}x_0(1-x_1)\phi_2(t) \quad (35)\\ &\quad + \mathrm{B}_{2,00}(1-x_0)(1-x_1)\phi_2(t) + \mathrm{B}_{3,11}x_0x_1\phi_3(t) \quad (36)\end{aligned}$$

Everytime, vector 11 is presented to the circuit,at time t, the metric M, at that instance is $\mathrm{B}_{1,11}\phi_1(t) + \mathrm{B}_{3,11}\phi_3(t)$.Strickly speaking, for a sequential circuit, it is not required,for a vector, presented at different time instants t to contribute a constant $\mathrm{B_{ki}}$.
Instead,the definition of $\mathrm{B_{ki}}$ is as follows.

Definition 3 *Given a time instant t, is picked randomly, from a uniform distribution of ts,* $\mathrm{B_{ki}}$ *is statistical average value of component* $\phi_k(t)$*, at time instant t, caused by the term* $\prod_j \mathrm{Y_{ij}}$.

3.4 Conditions and Requirements

If $\mathrm{B_{ki}}$ exists , any set of sequential simulations, should converge onto $\mathrm{B_{ki}}$, independent of the location, or number of repetitions of the vector $\prod_j \mathrm{Y_{ij}}$.

4 Properties

Given a constant weightset $x = a_{i0}$. From

$$\mathrm{M} = \sum_k \mathrm{M_k}(\mathbf{X}(t))\phi_k(t) \tag{37}$$

Now, $\mathrm{b_k} = \langle \mathrm{M}\phi_k(t)\rangle = \mathrm{M_k}(a_{i0})$. We also know that,

$$\begin{aligned}\frac{\partial \mathrm{M}_0}{\partial a_{ie}} &= \frac{\partial b_k}{\partial a_{i0}} \quad (38)\\ &= \frac{\partial \mathrm{M_k}}{\partial a_{i0}}(a_{i0}) \quad (39)\end{aligned}$$

It is surprising to note that along the curve of a constant weightset $x_i = a_{i0}$, $\frac{\partial \mathrm{M}_0}{\partial a_{ie}}$ is still well defined and even though, the actual othorgonal component $\phi_e(t)$ is not present at the input.