

ニューラルネットワークを用いた論理故障診断と欠陥種類推定法

日大生産工(院) ○太田 菜月 日大生産工 細川 利典 明治大学 山崎 浩二
日大生産工 山内 ゆかり 日大生産工 新井 雅之

1. はじめに

近年、半導体微細化技術の進歩に伴い、超大規模集積回路（Very Large Scale Integrated circuits: VLSI）の高集積化・複雑化が急速に進展している[1]。それによって、異常動作の物理的な原因を特定する物理的故障解析（Physical Fault Analysis: PFA）[2]は歩留まりの向上のために、より重要となっている。PFAでは、電子顕微鏡などを用いて故障 VLSI 内部の観測を行うため、多大なコストを要する。それゆえ、故障 VLSI に存在する可能性のある故障（被疑故障）の数を事前に絞り込む故障診断[3]が、故障解析コスト低減のために重要となる。故障診断では、故障 VLSI の異常な外部出力応答を裏付けることのできる故障箇所および欠陥種類を推定する。

組合せ回路やフルスキャン設計された順序回路における単一縮退故障モデルの故障診断は様々な手法が提案されており、被疑故障数も大きく削減できることが報告されている[4][5]。また、順序回路の故障診断でも故障モデルを縮退故障などに限定すれば、故障シミュレーションを実行することで被疑故障を絞り込むことが可能である[6]。しかしながら、特定の故障モデルに限定した故障診断では、故障 VLSI の異常な外部出力応答を裏付ける故障箇所が存在しない場合[7]や、誤った箇所を被疑故障として診断する[7]可能性がある。それゆえ、特定の故障モデルに限定しない故障診断法が重要となる。

文献[8]では、すべての論理故障を表現できるユニバーサル論理故障モデルを用いた論理故障 VLSI に対する故障診断法が提案されている。この手法では、誤診断がなく正確に故障箇所を推定できることが確認されている。しかしながら、この手法では被疑故障信号線数の多さや欠陥種類の推定が不可能なことが課題として残る。

故障診断における欠陥種類の特定は、二次テストで分解能の高い追加テストを用いた場合にのみ可能であることが多く、多くの場合 PFA によって補われる[2]。PFA では、システムティック欠陥のようなレイアウトに依存する欠陥やその位置を特定するアプローチを基に欠陥種類の特定を行う。しかしながら、故障によって引き起こされる誤った振舞いの正確な論理モデルの不足によって PFA の分析が非常に複雑化する。また、故障の励起や検出は物理的なものや設計のトポロジー因子に依存し、それらは予測不可能なものも存在する。

人工ニューラルネットワーク（Artificial Neural Network: ANN）は、人間の脳内神経回路網を模した数理モデルであり、ANN を多層にして用いることで、データに含まれる特徴を段階的に深く学習することができる。それゆえ、故障 VLSI のテスト結果と故障メカニズムを関連づける予測可能なルールを基にした欠陥種類推定では判定しきれない欠陥種類の推定は、ANN のアプローチに最適である[9]。

文献[10]では、故障診断の早い段階でシステムティック欠陥を分類するために、ANN を用いた欠陥種類推定法が提案されている。この手法では、特別な診断テストパターンを用いずに、故障 VLSI のテストや初期故障診断から得られた情報を基にした論理シミュレーション結果から、分類対象のそれ

ぞれの故障の特徴量を求める。そして、それらの特徴量を基に、ANN を用いて欠陥種類を推定する。

本論文では、文献[10]の手法を先行研究とし、故障 VLSI のテストや文献[8]の手法を用いた故障診断から得られた情報を基に、ANN を用いて欠陥種類を推定する手法を提案する。提案手法では、文献[10]で提案された ANN を用いた欠陥種類推定法の基本的な枠組みを利用する。文献[10]では、スキャンベーステストに基づいた 1 パターンテストと 2 パターンテストを対象に論理故障やタイミング故障の特徴量を求めている。一方、本論文では、マルチサイクルキャプチャテストにおけるユニバーサル論理故障シミュレーションを用いた故障診断結果から得られた情報を基に、主要な論理故障モデルである縮退故障[1]、支配型ブリッジ故障、オープン故障[11]を表すそれぞれの特徴量を求める。

第 2 章ではユニバーサル論理故障診断法について述べる。第 3 章では提案手法である人工ニューラルネットワークを用いた欠陥種類推定法について述べる。第 4 章ではベンチマーク回路を用いた実験結果を示す。最後に第 5 章では、結論と今後の課題について述べる。

2. ユニバーサル論理故障診断手法

本章では、被疑故障信号線を絞り込むための故障診断手法であるユニバーサル論理故障診断手法[8]について述べる。第 2.1 節では、すべての論理故障を表現できるユニバーサル論理故障モデルについて説明する。第 2.2 節では、ユニバーサル論理故障モデルを対象とした故障シミュレーション法について説明する。

2.1. ユニバーサル論理故障モデル

本論文では、ゲートの論理関数を変化させる論理故障を分類せず、まとめて取り扱う。なお、スキャンテストのキャプチャモードに複数サイクル間の順序動作を組み込んだマルチサイクルキャプチャテスト[12]の故障診断では、被検査回路を時間展開して考えるため、単一故障であっても複数の時刻で誤りが生じ得る。本論文では、マルチサイクルキャプチャテストの時間展開数 k の各時刻における故障励起の有無に基づいて故障をモデル化する。この故障モデルをユニバーサル論理故障モデル[8]と呼ぶ。

表 1 の 1 列目から 4 列目に、3 サイクルキャプチャテストの場合のユニバーサル論理故障モデルの誤り動作例を示す。表 1 は、ある信号線 n に対し、各時刻で誤りが生じるときを F、生じないときを P で示している。例えば表 1 の誤り 4 は、1 時刻目に信号線 m, n の値が期待値と異なる値になり、2 時刻目、3 時刻目では期待値と等しい値になる誤り動作を表している。表 1 より、 k サイクルキャプチャテストの場合、ある信号線 m, n に対して、誤りが生じる時刻の組合せ（誤り動作）は、 $2^k - 1$ 通りであることがわかる。本論文の診断手法では、この $2^k - 1$ 通りの組合せすべてに関して振舞いをシミュレーションする。

A Logical Fault Diagnosis Method and an Estimation Method of Defect Types Using Artificial Neural Networks

Natsuki OTA, Toshinori HOSOKAWA, Kouzi YAMAZAKI, Yukari YAMAUCHI and Masayuki ARAI

表 1. ユニバーサル論理故障診断例($k = 3$)

誤り動作	時刻			m		n		観測値	
	1	2	3	TS_1	TS_2	TS_1	TS_2	TS_1	TS_2
誤り1	F	F	F	11	10	01	11	00	11
誤り2	F	F	P	01	10	00	10		
誤り3	F	P	F	10	01	10	01		
誤り4	F	P	P	01	00	11	10		
誤り5	P	F	F	00	01	11	01		
誤り6	P	F	P	00	00	00	00		
誤り7	P	P	F	11	10	11	01		

2.2. ユニバーサル論理故障モデルを対象とした故障シミュレーション法

単一ユニバーサル論理故障モデルでは故障の励起条件が不明であるため、パステストでの故障の励起の有無を判断できない。したがって、ユニバーサル論理故障モデルを対象とした故障シミュレーション法では、フェイルテスト系列のみを用いて故障診断を行う。

本論文では単一ユニバーサル論理故障を対象として故障診断を行う。単一ユニバーサル論理故障に故障シミュレーション法を適用するためには、表 1 の 1 列目から 4 列目に示したような、考えうるすべての誤り動作について故障シミュレーションを実行し、外部出力および疑似外部出力におけるシミュレーション結果と観測値を比較する。すべてのフェイルテスト系列で両者が完全に一致したものを被疑故障とする。また、本論文では、シミュレーション結果（外部出力、疑似外部出力）がテストの観測値と一致する誤り動作を、「観測値を説明できる誤り動作」と表現する。

表 1 に、3 サイクルキャプチャテストにおいて、ある信号線 m , n に関する故障診断の例を示す。表 1 の TS_1 , TS_2 はフェイルテスト系列、 m , n の列は TS_1 , TS_2 での故障シミュレーション値、観測値の列は TS_1 , TS_2 での観測値を示す。 TS_1 では、信号線 m , n は共に観測値 00 を説明できる誤り動作が存在する。しかしながら、 TS_2 では、信号線 n は誤り動作 FFF で観測値 11 を説明できるが、信号線 m は観測値 11 を説明できる誤り動作が存在しない。それゆえ、信号線 n において、 TS_1 で誤り動作 FPP もしくは PFP として振る舞い、 TS_2 で誤り動作 FFF として振る舞う論理故障が被疑故障と判定される。したがって、信号線 n が被疑故障信号線と判定される。

表 2 に、表 1 に示した故障診断例で得られる故障検出情報を表す。故障検出情報は、被疑故障信号線と判定された信号線に対してのみ得られる。被疑信号線 n の TS_1 における故障検出情報は、観測値 00 を説明できる誤り動作である FFP および PFP、 TS_2 における故障検出情報は、観測値 11 を説明できる誤り動作である FFF である。被疑故障信号線 n が真の故障信号線であれば、 TS_1 の 1 時刻目は F もしくは P、2 時刻目は F、3 時刻目は P でなければならない。このような情報が被疑故障信号線ごと、テスト系列ごとに得られる。故障検出情報については、第 4.1 節で説明する。

表 2. ユニバーサル論理故障診断例($k = 3$)

テスト系列	TS_1			TS_2		
	1	2	3	1	2	3
信号線 n	F	F	P	F	F	F
	P	F	P			

3. 人工ニューラルネットワークと故障検出情報を用いた欠陥種類推定手法

本章では、故障 VLSI のテストやユニバーサル論理故障診断[8]から得られた情報を基に、ANN を用いて欠陥種類を推定する手法を提案する。第 3.1 節では、提案手法の概要を説明する。第 3.2 節では、特徴量を求める際に用いる故障検出情報について説明する。第 3.3 節では、本提案手法の特徴量について説明する。第 3.4 節では、ANN の出力となるラベル値を求めるための出力符号化について説明する。

3.1. 概要

ANN の学習やテストに使用されるデータは、特徴量とラベル値のペア (*features, labels*) から成る集合で構成される。提案手法では、ユニバーサル論理故障モデルに対する故障診断結果とそのテスト集合の論理シミュレーション結果から主要な論理故障モデルである縮退故障、支配型ブリッジ故障、オープン故障を表すそれぞれの特徴量を求め、ANN を用いて欠陥種類を推定する。製造テストからはフェイルテスト系列およびパステスト系列を、故障診断からは被疑故障信号線や故障検出情報を得て、それらを用いて特徴量を求める。1 つの欠陥に対する被疑故障信号線が複数得られた場合、すべての被疑故障信号線に対しての特徴量をそれぞれ求める。

図 3 に提案手法のアルゴリズムを示す。文献[10]の手法同様に、アルゴリズムは大きく分けられる。特徴量は、第 2 章で説明したユニバーサル論理故障モデルに対する故障診断結果とそのテスト集合の論理シミュレーション結果から計算される。

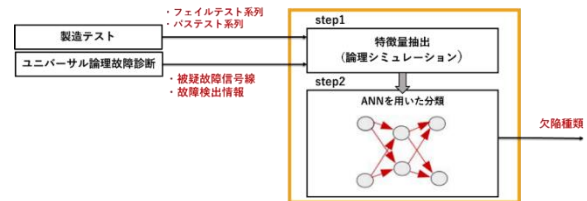


図 1. 提案手法のアルゴリズム

3.2. 故障検出情報

提案手法では、製造テストからフェイルテスト系列およびパステスト系列を、ユニバーサル論理故障診断から被疑故障信号線や故障検出情報を得て、それらを用いて特徴量を求める。

フェイルテスト系列の故障検出情報とは、第 2 章で述べたユニバーサル論理故障診断においてシミュレーション結果（外部出力、疑似外部出力）がテストの観測値と一致する、すなわち観測値を説明できる誤り動作の集合のことである。例えば、表 1 および表 2 より、被疑故障信号線 n のテスト系列 TS_1 における故障検出情報は FFP と PFP、テスト系列 TS_2 における故障検出情報は FFF である。

第2章で述べたように、ユニバーサル論理故障診断ではパステスト系列での故障の励起の有無を判断できないため、フェイルテスト系列のみを用いて故障診断を行う。しかしながら、パステストの故障検出情報からは、故障が励起されてはいけないテスト系列、および時刻が分かる。それゆえ、提案手法では特徴量を抽出する際にフェイルテスト系列の故障検出情報だけでなく、パステスト系列の故障検出情報も用いる。パステスト系列の故障検出情報とは、ユニバーサル論理故障診断においてシミュレーション結果（外部出力、疑似外部出力）がテストの観測値（すなわち期待値）と一致する誤り動作の集合のことである。

3.3. 特徴量

ANNの構造や学習と同様に、特徴量抽出は本論文で提案する欠陥種類推定法の正答率を向上するために非常に重要である。現時点での分類対象の論理故障モデルは0縮退故障、1縮退故障、支配型ブリッジ故障、オープン故障の4種類の論理故障である。

特徴量抽出では、まず故障検出情報集合の要素の積集合を求める。本論文では、論理圧縮アルゴリズム Espresso[13]を用いて故障検出情報の単純化式を算出する。故障検出情報の単純化式は、励起によりある時刻で誤りが生じるという事象を表す F、生じないという事象を表す P、生じていても生じていなくても故障検出に関係のないという事象を表す X から構成される。例えば、表2の被疑故障信号線 n のテスト系列 TS_1 における故障検出情報の単純化式は XFP である。次に、フェイルテスト系列およびパステスト系列による論理シミュレーション結果と故障検出情報の単純化式を用いて、以下の7つの特徴量を取得する。

縮退故障を表す特徴量を特徴量1～特徴量4に、支配型ブリッジ故障を表す特徴量を特徴量5に、オープン故障を表す特徴量を特徴量6～特徴量7に示す。

特徴量1 故障検出情報が F のとき、故障信号線の論理値が 1 であるテスト系列の割合

式(2)に特徴量1を求める式を示す。 $|TS|$ はフェイルテスト系列とパステスト系列を合わせた全テスト系列数を、 $|TS_{F1}|(TS_{F1} \subset TS)$ は故障検出情報が F であるすべての時刻で故障信号線の論理値が 1 であるテスト系列数を表す。

$$feature_1 = |TS_{F1}|/|TS| \quad (2)$$

特徴量2 故障検出情報が P のとき、故障信号線の論理値が 0 であるテスト系列の割合

式(3)に特徴量2を求める式を示す。 $|TS_{P0}|(TS_{P0} \subset TS)$ は故障検出情報が P であるすべての時刻で故障信号線の論理値が 0 であるテスト系列数を表す。

$$feature_2 = |TS_{P0}|/|TS| \quad (3)$$

特徴量3 故障検出情報が F のとき、故障信号線の論理値が 0 であるテスト系列の割合

式(4)に特徴量3を求める式を示す。 $|TS_{F0}|(TS_{F0} \subset TS)$ は故障検出情報が F であるすべての時刻で故障信号線の論理値が 0 であるテスト系列を表す。

$$feature_3 = |TS_{F0}|/|TS| \quad (4)$$

特徴量4 故障検出情報が P のとき、故障信号線の論理値が 1 であるテスト系列の割合

式(5)に特徴量4を求める式を示す。 $|TS_{P1}|(TS_{P1} \subset TS)$ は故障検出情報が P であるすべての時刻で故障信号線の論理値が 1 であるテスト系列数を表す。

$$feature_4 = |TS_{P1}|/|TS| \quad (5)$$

特徴量5 故障信号線の論理値に影響を与える可能性のある信号線の存在を表すスコア

すべてのテスト系列において、故障信号線の隣接信号線集合内に全時刻で以下の条件のいずれかを満たす隣接信号線 Aggressor が存在する可能性を基に、任意のスコアを表す。

【条件】

1. 故障検出情報が X である
2. 故障検出情報が P のとき、故障信号線の論理値と隣接信号線の論理値が等しい
3. 故障検出情報が F のとき、故障信号線の論理値と隣接信号線の論理値が異なる

式(6)に、特徴量5を求める式を示す。Aggressor は故障信号線の論理値に常に影響を与える可能性のある信号線、Adj は故障信号線の隣接信号線集合、 x は任意のスコア ($0 < x \leq 1$) を表す。本論文では、 $x = 1.0$ とする。

$$feature_5 = \begin{cases} x, & \exists adj \in Adj : adj \text{ is Aggressor} \\ 0, & \text{others} \end{cases} \quad (6)$$

特徴量6 故障検出情報が F のとき、故障信号線の論理値と隣接信号線の論理値が異なるテスト系列の割合

式(7.1)および式(7.2)に、あるテスト系列 TS_j を印加した際の故障励起関数 $E_{j,k}(f)$ を表す式を示す。 $TD_k(TS_j)$ は TS_j を印加した際に時刻 k で論理値が 1 である隣接信号線の並走距離の総和を、 D_{ai} は隣接信号線 a_i の並走距離を、 $v_{ai}(TS_j, k)$ は TS_j を印加した際の隣接信号線 a_i の時刻 k における論理値を表す。故障励起関数では、故障信号線の隣接信号線のうち、論理値が 1 である隣接信号線の並走距離の総和が全隣接信号線の並走距離の総和の 1/2 以上のとき 1 を、それ以外のとき 0 を出力する。

$$TD_k(TS_j) = \sum_i D_{ai} \cdot v_{ai}(TS_j, k) \quad (7.1)$$

$$E_{j,k}(f) = \begin{cases} 0, & TD_k(TS_j) < \frac{1}{2} \sum_i D_{ai} \\ 1, & TD_k(TS_j) \geq \frac{1}{2} \sum_i D_{ai} \end{cases} \quad (7.2)$$

式(8)に特徴量6を求める式を示す。 $|TS_{open1}|(TS_{open1} \subset TS)$ は故障検出情報が F であるすべての時刻で、故障信号線の論理値と故障励起関数 $E_j(f)$ の結果が異なるテスト系列数を表す。

$$feature_6 = |TS_{open1}|/|TS| \quad (8)$$

特徴量7 故障検出情報が P のとき、故障信号線の論理値と隣接信号線の論理値が等しいテスト系列の割合

式(9)に特徴量7を求める式を示す。 $|TS_{open2}|(TS_{open2} \subset TS)$ は故障検出情報が P であるすべての時刻で、故障信号線の論理値と故障励起関数 $E(v_f)$ の結果が等しいテスト系列数を表す。

$$feature_7 = |TS_{open2}|/|TS| \quad (9)$$

3.4. 出力符号化

本論文では、論理故障の欠陥種類を 0 縮退故障, 1 縮退故障, 支配型ブリッジ故障, オープン故障の 4 つに分類する。出力は、分類する故障モデル数の長さのベクトル $[y_0, y_1, y_2, y_3]$ ($y_0, y_1, y_2, y_3 \in \{0,1\}$) として符号化される。 y_0 は 0 縮退故障のラベル値, y_1 は 1 縮退故障のラベル値, y_2 は支配型ブリッジ故障のラベル値, y_3 はオープン故障のラベル値を表す。訓練データの場合ラベル値は既知の値であり、教師信号として分類したい故障モデルを表すノードのラベル値を 1 に、その他のノードのラベル値を 0 に設定する。例えば、分類したい故障モデルが 1 縮退故障の場合、教師信号は $[0,1,0,0]$ に設定され、オープン故障の場合 $[0,0,0,1]$ に設定される。

4. 実験結果

本章では、提案手法の有用性を示すために 2 つのベンチマーク回路を用いた実験結果を示す。本実験は、故障 VLSI のテストやユニバーサル論理故障診断法[8]を用いた故障診断から得られた情報を基に、ANN を用いた欠陥種類推定法を適用する。欠陥種類は、主要な論理故障モデルである 0 縮退故障, 1 縮退故障, 支配型ブリッジ故障, オープン故障の 4 つに分類する。テスト系列は、縮退故障検出用 k サイクルキャプチャテストを用いる。故障は、各回路に各故障モデルの故障を 500 つずつ挿入し、故障を挿入する信号線は、回路のレイアウト情報を基に、隣接信号線の存在する信号線のうち、複数の種類の故障と等価な振舞いをしない信号線の中からランダムに選択する。ANN の入力データである特徴量は、全データの 70% を訓練データ、残りの 30% をテストデータとする。このとき、データの振分けはランダムに行う。また、ANN の学習では最適化関数は Adagrad[14]を用い、学習率は 0.1 とする。

表 4 に、隠れ層の層数を 2 層、ノード数を 9 つに固定して実験を行った際の実験結果を示す。1 列目は実験対象回路名を表す。2 列目は正しい答えの故障モデル名を、3 行目から 6 行目は推定結果を表す。再現率とは、答えが A であるもののうち、正しく A と推測されたものの割合を指す。正解率とは、全予測に対する正答率を表す。

最適な ANN の構造は実験対象回路ごとに異なるが、そのような最適な選択を行うことは、回路に関する知識と利用可能なデータにより初めて可能となり、困難である。しかしながら、最適な構造ではなくとも、多くの回路で問題なく動作する ANN の構造を設定することは可能である。本実験では、すべての回路で 95% 以上の高い精度を得ることができた。

表 3. 欠陥種類推定結果

回路名	答え	推定結果				再現率 [%]	推定精度 [%]
		sa0	sa1	bridge	open		
GATE_MUL	sa0	150	0	0	0	100	97.14
	sa1	0	149	0	1	99.33	
	bridge	0	1	99	6	93.39	
	open	0	0	7	113	94.16	
GATE_DIV	sa0	150	0	0	0	100	96.83
	sa1	0	149	0	1	99.33	
	bridge	5	2	134	9	89.33	
	open	0	0	2	148	98.06	

5. 結論

本論文では、故障 VLSI のテストや文献[8]の手法を用いた故障診断から得られた情報を基に、ANN を用いて欠陥種類を推定する手法を提案し、評価した。提案手法では、ユニバーサル論理故障モデルに対する故障診断結果とそのテスト集合

の論理シミュレーション結果から、主な論理故障モデルである縮退故障, 支配型ブリッジ故障, オープン故障を表すそれぞれの特徴量を求め、人工ニューラルネットワークを用いて、被疑故障信号線の欠陥種類を推定する。ベンチマーク回路を用いた実験では、すべての回路で 95% 以上の推定結果を達成した。

今後の課題として、さらに正答率を上げるために ANN の構造や活性化関数等の変更と評価を行うことや、セル内故障やタイミング故障も含めた欠陥種類推定法の提案が挙げられる。

参考文献

- [1] 藤原秀雄, "ディジタルシステムの設計とテスト," 工学図書株式会社, 2004
- [2] S.-Y. Liu, Y.-C. Hou, C.-C. Chang, and J.-C. Lin, "Sige profile inspection by using dual beam fib system in physical failure analysis," in 20th IEEE Int'l Symposium on the Physical and Failure Analysis of Integrated Circuits, pp.490-492, 2013.
- [3] H.Y.Chang, E.Manning, and G.Metze "Fault Diagnosis of Digital Systems," John Wiley & Sons, Inc., 1970.
- [4] Ismed Hartanto, Srikanth Venkataraman, W. Kent Fuchs, Elizabeth M. Rudnick, Janak H. Patel, and Sreejit Chakravarty, "Diagnostic simulation of stuck-at faults in sequential circuits using compact lists," ACM Transactions on Design Automation of Electronic Systems, pp.471-489, 2001.
- [5] V.Boppana and W. K. Fuchs, "Fault dictionary compaction by output sequence removal," Proc. ICCAD, pp.287-296, 2001.
- [6] T. Yamada, K. Yamazaki, and E. J. McCluskey, "A Simple Technique for Locating Gate-Level Faults in Combinational Circuits," Proc. ATS, pp.65-70, 1995.
- [7] D. B. Lavo, T. Larrabee, and B. Chess, "Beyond the Byzantine Generals: Unexpected Behavior and Bridging Fault Diagnosis," Proc. ITC, pp.611-619, 1996.
- [8] H. Takano, T. Hosokawa, H. Yamazaki and K. Yamazaki, "A Fault Diagnosis Method for a Single Universal Logical Fault Model Using Multi Cycle Capture Test Sets," Proc 16th IEEE Workshop on RTL and High-Level Testing, pp.74-79, 2015.
- [9] C. M. Bishop, "Pattern Recognition and Machine Learning (Information Science and Statistics)," Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2006.
- [10] L.R.Gómez and H-J.Wunderlich, "A Neural-Network-Based Fault Classifier," IEEE 25th Asian Test Symposium, pp.144-149, 2016.
- [11] 山崎浩二, 堤利幸, 高橋寛, 樋上喜信, 相京隆, 四柳浩之, 橋爪正樹, 高松雄三 "故障励起関数を利用したオープン故障の診断法," 電子情報通信学会論文誌 D Vol. J93-D No.11, pp. 2416-2425, 2010
- [12] J.Abraham, U.Goel and A.Kumar, "Multi-cycle sensitizable transition delay faults," VLSI Test Symposium, pp.306-313, 2006.
- [13] 笹尾勤, "論理設計 スイッチング回路理論," 近代科学社, pp.195-196, 2005
- [14] John Duchi, Elad Hazen, Yoram Singer. "Adaptive Subgradient Methods for Online Learning and Stochastic Optimization," Journal of Machine Learning Research 12, 2011