

卒業論文

ノードの行列状配置構造による  
文字列間の高速類似度判定法と  
その回路構成の検討

指導教官

鈴木正國教授  
北川章夫助教授  
秋田純一助手  
深山正幸助手

提出者

金沢大学 工学部  
電気・情報工学科

集積回路工学研究室

佐々木 勝光  
平成 12 年 2 月

# 目次

第 1 章	はじめに .....	4
1-1	本研究の背景および目的 .....	4
第 2 章	遺伝子解析におけるホモロジー検索 .....	5
2-1	遺伝子に起こる突然変異 .....	5
2-1-1	遺伝子に起こる突然変異 .....	5
2-1-2	突然変異によって引き起こされる塩基配列の変化 .....	5
2-2	塩基配列の一致比較 .....	6
2-2-1	完全一致 .....	6
2-2-2	塩基置換の場合の一致比較 .....	7
2-2-3	塩基欠落の場合の一致比較 .....	8
2-2-4	塩基混入の場合の一致比較 .....	8
2-3	ノードの行列状配置構造による類似度判定法と .....	9
	その評価 .....	9
2-3-1	セルの変化を元にした類似度判定法 .....	9
2-3-2	ノードの行列状配置構造による類似度判定法 .....	10
2-3-3	点数処理 .....	12
2-3-4	並列処理 .....	13
2-3-5	類似度判定法の評価 .....	13
第 3 章	高速類似度判定回路の構成 .....	15
3-1	各ノードの回路構成 .....	15
3-1-1	ノードのブロック図 .....	15
3-1-2	ノードの回路図 .....	16
3-2	全体の回路構成 .....	18
3-3	buffer の特性 .....	19
3-4	HSPICE によるシミュレーションと結果 .....	21
3-4-1	HSPICE と C 言語とによるシミュレーション結果の比較 .....	21
3-4-2	遅延時間 .....	23
第 4 章	まとめ .....	26

4-1	まとめ .....	26
	謝辞 .....	27
	参考文献 .....	28

# 第1章 はじめに

## 1-1 本研究の背景および目的

現在、ゲノムプロジェクトと称して4種類の塩基、アデニン (A)、チミン (T)、シトシン (C)、グアニン (G) の配列から成る遺伝子の配列解析が、コンピュータを用いて盛んに行われており、ここ数年のうちに終了しようとしている。配列解析の次のステップとして、配列情報をデータベースとした遺伝子の機能推定がある。ここで、塩基配列は突然変異によって変異することがあり、類似している配列にも重要な情報が含まれている場合が多いとされているため、機能推定を行ううえでは必ずしも完全に一致しているものだけではなく、類似している配列を探し出すという事も非常に重要となってくる。

しかしながら、コンピュータによる解析では一致、不一致は容易に判別できても、「どのぐらい合っているか」といった類似の判別という点ではソフトウェアに頼らざるをえない。また、配列比較を行う際には膨大な塩基対の配列情報が納められているデータベースから、数百個の塩基対からなる塩基配列に類似した塩基配列を探し出すというようなことがあるため、従来のソフトウェアによる比較では、逐次処理を行っているため処理時間が非常に長くなってしまいう問題が生じている。それを改善しようと各種の高速化アルゴリズムが提案されているが、それでもなお比較処理に時間がかかってしまう。

そこでハードウェアによる、比較処理の並列性を生かした高速類似度判定法を提案し、それに基づいた回路を用いることで高速に類似度判定を行うことを目的としている。

## 第2章 遺伝子解析におけるホモロジー検索

### 2-1 遺伝子に起こる突然変異

#### 2-1-1 遺伝子に起こる突然変異

親から子へ、祖先から子孫へと遺伝情報を伝達する役目を果たすものが、遺伝子の本体である DNA である。通常この遺伝情報は正確に次の世代へと伝えられるのだが、外的もしくは内的な様々な要因によって突然変異が生じ、遺伝情報に変化が起こる。突然変異は自然偶発的或いは突発的に発生するものと、放射能汚染に代表されるように人為的に発生するものがある。

#### 2-1-2 突然変異によって引き起こされる塩基配列の変化

突然変異が生じると、DNA を構成している塩基の配列に様々な変化が生じる。突然変異が生じた DNA から複製される DNA は、すべて正常な DNA から塩基配列が変化したものとなる。塩基が変化するパターンとしては3種類あり、本研究ではその変化を「塩基置換」、「塩基欠落」、「塩基混入」として表現する。

図 2-1 の例を用いて説明する。正常な塩基対に対して点線で囲ってある T-A という塩基対が置換、欠落、混入する場合を考えると、置換した場合は T-A という塩基が C-G という塩基に置き換わっている。欠落した場合は欠落部分から塩基対列が左にずれる。そして混入した場合は混入部分から塩基対列が右にずれることとなる。

前述したように、突然変異によって「ずれ」が生じた塩基配列に対して「どのくらい似ているか」、という微妙な判断をコンピュータを用いて行うとするとその判断を下すことは困難である。そのため現在はソフトウェアを用いる解析が考えられているが処理に非常に時間がかかってしまう。そこで我々は、以下に示すハードウェアを用いて実現できる類似比較のアルゴリズムを提案することにより、この問題を解決できるのではないかと考えた。

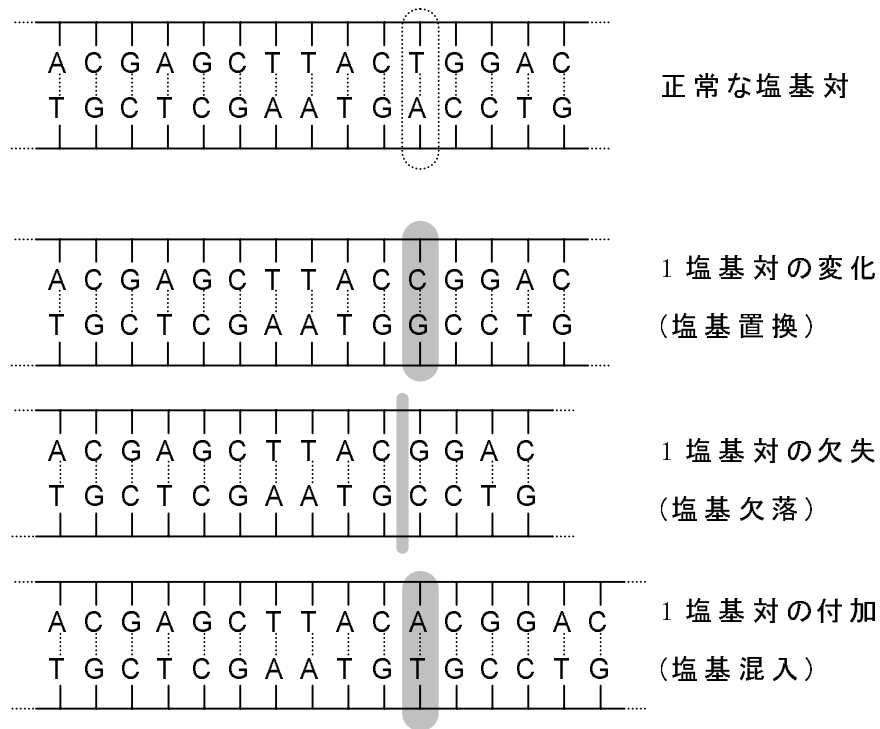


図 2-1 遺伝子の突然変異による塩基配列変化の例

## 2-2 塩基配列の一致比較

### 2-2-1 完全一致

実際に図 2-2 のように、最左列に正常な塩基配列、最上行に変化した塩基配列を並べることにより、各々の塩基に対して比較処理を行うセルを行列状に並べることができる。比較する 2 つの塩基が一致していれば「○」、不一致ならば「×」がセルに表示されている。塩基は片方が A ならばもう一方は T、或いは C ならば G というように必ず決まった対を成す性質「ハイブリダイゼーション」があるので、簡単のために表中では片方の塩基だけを示している。表を見て判るように、塩基配列が完全に一致している場合は対角線上に一致を示す「○」が並ぶことが判る。

↖	A	C	G	A	G	C	T	T	A	C
A	○	×	×	○	×	×	×	×	○	×
C	×	○	×	×	×	○	×	×	×	○
G	×	×	○	×	○	×	×	×	×	×
A	○	×	×	○	×	×	×	×	○	×
G	×	×	○	×	○	×	×	×	×	×
C	×	○	×	×	×	○	×	×	×	○
T	×	×	×	×	×	×	○	○	×	×
T	×	×	×	×	×	×	○	○	×	×
A	○	×	×	○	×	×	×	×	○	×
C	×	○	×	×	×	○	×	×	×	○

図 2-2 完全一致

### 2-2-2 塩基置換の場合の一致比較

次に塩基置換が1カ所起こった場合と3カ所起こった場合について、一致比較を図2-3の(a)と(b)に示す。この図から、置換が1カ所の場合も3カ所の場合も対角線上のセルは変化が起こった箇所のみ「×」に変化し、それ以外の対角線上のセルは変化せずに「○」を示していることが判る。このことから置換が起こった場合は、塩基の変化量に関わらず対角線上の一致を表すセルにはずれが生じないことが判る。

↖	A	C	G	A	T	C	T	T	A	C
A	○	×	×	○	×	×	×	×	○	×
C	×	○	×	×	×	○	×	×	×	○
G	×	×	○	×	×	×	×	×	×	×
A	○	×	×	○	×	×	×	×	○	×
G	×	×	○	×	×	×	×	×	×	×
C	×	○	×	×	×	○	×	×	×	○
T	×	×	×	×	○	×	○	○	×	×
T	×	×	×	×	×	○	○	○	×	×
A	○	×	×	○	×	×	×	×	○	×
C	×	○	×	×	×	○	×	×	×	○

(a) 1箇所の塩基置換

↖	A	C	G	C	A	C	G	T	T	C
A	○	×	×	×	○	×	×	×	×	×
C	×	○	×	○	×	○	×	×	×	○
G	×	×	○	×	×	×	○	×	×	×
A	○	×	×	×	○	×	×	×	×	×
G	×	×	○	×	×	○	×	×	×	×
C	×	×	×	○	×	○	×	×	×	○
T	×	○	×	×	×	×	○	○	○	×
T	×	○	×	×	×	×	×	○	○	×
A	○	×	×	×	○	×	×	×	×	×
C	×	×	×	○	×	○	×	×	×	○

(b) 3箇所の塩基置換

図 2-3 塩基置換の例

### 2-2-3 塩基欠落の場合の一致比較

次に、塩基が1カ所欠落した場合と3カ所欠落した場合について、一致比較を図2-4の(a)と(b)に示す。ここで、欠落が生じると塩基数が減少するため、塩基列の最後尾に任意の塩基を付け加えることとした。1カ所変化させた場合を示す図2-4(a)の対角線上の変化から、欠落が生じたセルは「×」に代わり、欠落によって起こった塩基列の「ずれ」によって、一致を示している「○」が対角線から平行に「下」へとずれていることが判る。同様に、3カ所変化させた場合を示す図2-4(b)の変化から、(a)の場合と同様なずれが生じ、そのずれの回数は塩基の変化数と一致していて、変化数と同じセル数だけ対角線から離れていくことが判る。

	A	C	G	A	C	T	T	A	C	G
A	○	×	×	○	×	×	×	○	×	×
C	×	○	×	×	○	×	×	×	○	×
G	×	×	○	×	×	×	×	×	×	○
A	○	×	×	○	×	×	×	○	×	×
G	×	×	○	×	×	×	×	×	×	○
C	×	○	×	×	○	×	×	×	○	×
T	×	×	×	×	×	○	○	×	×	×
T	×	×	×	×	×	○	○	×	×	×
A	○	×	×	○	×	×	×	○	×	×
C	×	○	×	×	○	×	×	×	○	×

(a) 1箇所の塩基欠落

	A	G	A	T	T	A	C	G	A	T
A	○	×	○	×	×	○	×	×	○	×
C	×	×	×	×	×	×	○	×	×	×
G	×	○	×	×	×	×	×	○	×	×
A	○	×	○	×	×	○	×	×	○	×
G	×	○	×	×	×	×	×	○	×	×
C	×	×	×	×	×	×	○	×	×	×
T	×	×	×	○	○	×	×	×	×	○
T	×	×	×	○	○	×	×	×	×	○
A	○	×	○	×	×	○	×	×	○	×
C	×	×	×	×	×	×	○	×	×	×

(b) 3箇所の塩基欠落

図 2-4 塩基欠落の例

### 2-2-4 塩基混入の場合の一致比較

最後に、塩基が1カ所混入した場合と3カ所混入した場合について、一致比較を図2-5の(a)と(b)に示す。ここで混入される塩基は任意とし、混入が生じると塩基数が増加するため右端の塩基を切り捨てている。一カ所変化させた場合を示す図2-5(a)の対角線上の変化から、置換の場合と同様に、混入によって起こった塩基列の「ずれ」によって、一致を示している「○」が対角線から平行に「右」へとずれていることが判る。同様に、3カ所変化させた場合を示す図2-5(b)の変化から、欠落の場合と同様に、そのずれの回数は塩基の変化数と一致していて、変化数と同じセル数だけ対角線から離れていくことが判る。



	A	C	G	T	A	G	C	T	T	A
A	○	×	×	×	○	×	×	×	×	○
C	×	○	×	×	×	×	○	×	×	×
G	×	×	○	×	×	○	×	×	×	×
A	○	×	×	×	○	×	×	×	×	○
G	×	×	○	×	×	○	×	×	×	×
C	×	○	×	×	×	×	○	×	×	×
T	×	×	×	○	×	×	×	○	○	×
T	×	×	×	○	×	×	×	○	○	×
A	○	×	×	×	○	×	×	×	×	○
C	×	○	×	×	×	×	○	×	×	×

(a) 1 箇所の塩基混入

	A	C	G	T	A	C	T	G	C	T
A	○	×	×	×	○	×	×	×	×	×
C	×	○	×	×	×	○	×	×	○	×
G	×	×	○	×	×	×	×	○	×	×
A	○	×	×	×	○	×	×	×	×	×
G	×	×	○	×	×	×	×	○	×	×
C	×	○	×	×	×	×	×	○	×	×
T	×	×	×	○	×	×	○	×	×	○
T	×	×	×	○	×	×	○	×	×	○
A	○	×	×	×	○	×	×	×	×	×
C	×	○	×	×	×	○	×	×	○	×

(b) 3 箇所の塩基混入

図 2-5 塩基混入の例

## 2-3 ノードの行列状配置構造による類似度判定法とその評価

### 2-3-1 セルの変化を元にした類似度判定法

以上の一致比較の例から、置換、欠落、混入の三種類の変化に伴ってセルに以下のような特徴があることが判った。

- 1 一致が続いた場合は、対角線と平行に「○」が並ぶ
- 2 対角線と平行に並ぶ「○」の列に「×」があると、その「×」の数だけ置換が生じたことになる。また、対角線と平行に「○」が並ぶ列が、対角線から下にずれている場合は欠落、右にずれている場合は混入が生じたと判断でき、対角線からずれたセル数分だけ欠落、混入が生じたことになる。
- 3 対角線と平行にずれるパターンとしては「左上」、「下」、「右」の3種類がある

以上より、変化が生じた箇所、回数が判れば二つの塩基配列に関する類似の度合いを求めることができるのではないかと考えた。そこで、前節での一致比較の表を用いて、最初に与えた点数を変化が生じた箇所で減点することにより、類似の度合いを導き出す判定法を考えた。その例を図 2-6 に示す。

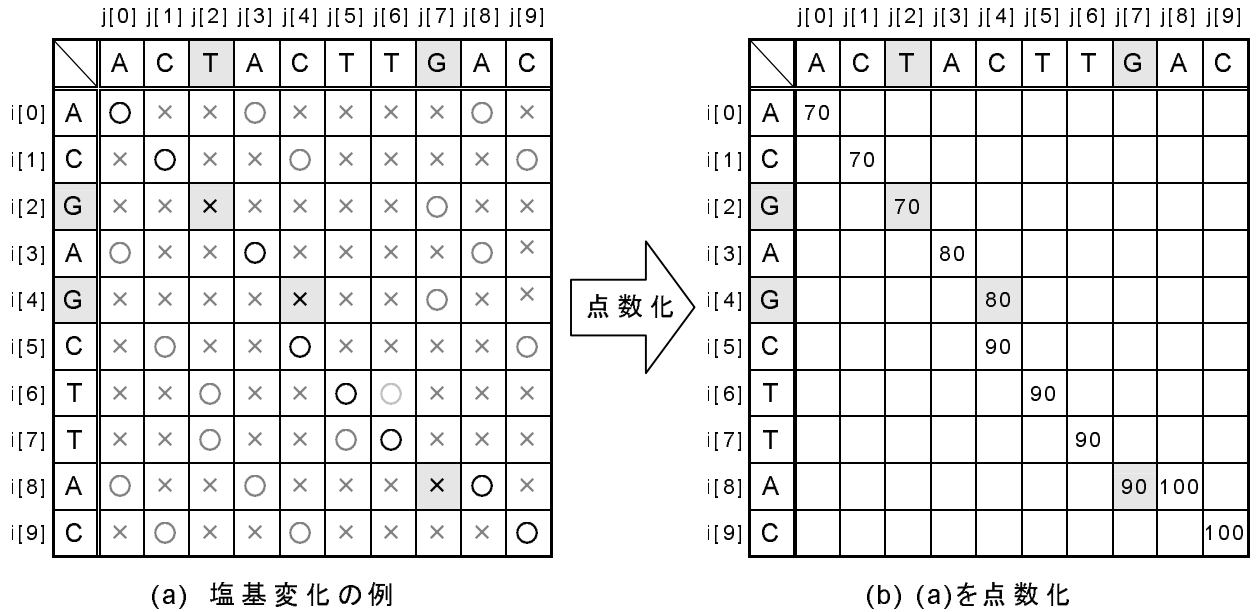


図 2-6 塩基の変化による一致度の点数化

図 2-6(a)には、最左列 (i 列) に一つの塩基配列、最上行 (j 行) にそれから変化が生じた塩基配列が示されていて、j[2]で i[2]が「G」から「T」に置換、j[3]と j[4]の間で i[4]の「G」が欠落、j[7]で「G」が混入されている場合が示されている。我々が提案する判定法では、一致していればそのセルに入力された点数をそのまま引き継ぎ、不一致ならば入力された点数を減点して引き継ぐとしているため、この表のように右下端のセルに 100 点を与えると、変化が生じた箇所に対応したセルで 10 点ずつ減点されながら左上に向かって点数が順次決定されて行き、最終的な点数が左上端に出力される。それを点数で表した表が図 2-6(b)である。ここでは簡単のために図 2-6(b)では、図 2-6(a)で濃く示されたセルの点数のみを表示している。このように今回提案する判定法は、前のセルの一致情報を考慮に入れて点数を決定していくため、図 2-6(b)で最終的に左上端に出力される「70」という点数は、2 つの塩基配列全体の一致情報を含んだ点数であると考えることができる。我々はこの点数を「2 つの塩基配列の類似度」と考えることとした。

### 2-3-2 ノードの行列状配置構造による類似度判定法

前項で類似度を定義したがこれを回路として実現するためには、一致比較と減点処理を行うセルに置き換わる回路を考案し、その回路を行列状に配置して塩基変化によって引き起こされる「ずれ」のパターンを生かした接続をすることが必

要である。以上を考慮に入れて接続された図 2-7 のような回路ならば類似度を求めることができるのではないかと考えた。以降セルに置き換わった回路をノードと呼ぶこととする。

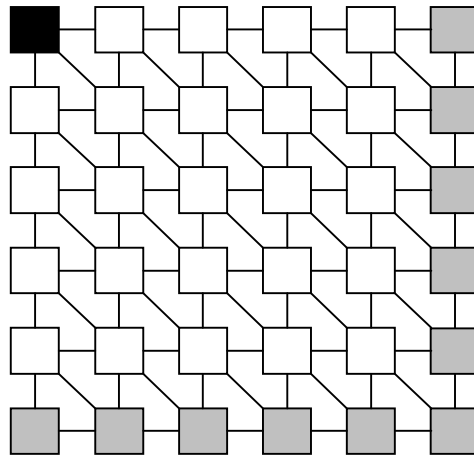


図 2-7 6×6 のノードの行列状配置構成

実際に図 2-7 を用いて比較処理を行う際には、以下のようなアルゴリズムを用いる。

- 1 最右列、最下列（図 2-7 の灰色のノード）の各々のノードに一致状況に応じて、決まった点数を入力する
- 2 各ノードでの点数を決定する際には、右、右下、下の 3 方向からの入力を以下のように処理して、最も高い点数をそのノードでの点数とし、上、左上、左の次のノードへと出力する。
  - 処理を行うノードが「○」ならば、右下からの点数はそのまま受け取り、右、下からの点数は減点して受け取る
  - 処理を行うノードが「×」ならば、3 方向からの点数を減点して受け取り、2-1 の場合よりも減点量を多くする。
  - 2-1, 2-2 いずれの場合も右、下からの点数の減点量に対して右下からの点数の減点量を多くすることで、対角線方向からの点数を優先する。
- 3 以上の操作が行列状に配置した全ノードにおいて独立に行われ、最右列、最下行から左上に向かって順次点数が決定されていき、最終的に行列の左

上端（図 2-7 の黒色のノード）に一致度に応じた点数が出力される。それを 2 つの塩基配列の類似度とする。

ここで 2 において 3 方向からの点数を減点した後、そのうちで最も高い点数を選択し次のノードへと出力するとしたが、その理由として、比較的点数が高いということはそのノードへとたどり着くまでにより多く一致したノードを経由してきたということが言え、信頼度が高いと考えたからである。また、右下からの点数を優遇する理由としては、一致が続く場合には対角線と平行に「○」が並ぶ傾向があることから、右もしくは下から入力される点数は塩基変化があった場合の点数であると考えることができたからである。

### 2-3-3 点数処理

各ノードでの点数の入力と出力の関係を表したものを図 2-8 に示す。図のように A、B<sub>1</sub>、B<sub>2</sub> の 3 方向のノードからの点数がノード N に入力され、各々に対して一致情報に応じた減点が行われ、そのうちで最も高い点数がノード N の点数となりノード C へと出力されることとなる。ここで右下からの点数を優遇するために、右または下からの点数は共に塩基の変化に影響を受けるものであるから、共通のパラメータを用いることとした。ノード N とノード A、B<sub>1</sub>、B<sub>2</sub> の一致情報から図 2-9 に示されているように 8 通りの組み合わせが考えられ、減点量は 6 通りに分けることができる。

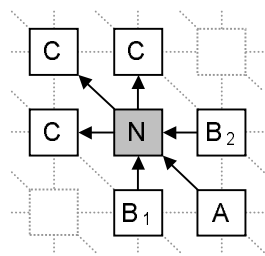


図 2-8 ノードの入出力の関係

入力元	ノード N	減点量
A : ○	○	P-1 (減点なし)
A : ×	○	P-1 (減点なし)
A : ○	×	P-2
A : ×	×	P-3
B <sub>1</sub> , B <sub>2</sub> : ○	○	P-4
B <sub>1</sub> , B <sub>2</sub> : ×	○	P-4
B <sub>1</sub> , B <sub>2</sub> : ○	×	P-5
B <sub>1</sub> , B <sub>2</sub> : ×	×	P-6

図 2-9 ノードの一致度による組み合わせ

本研究は「高速な類似度判定ができる回路の実現」が目標であることから、回路設計を行うことを考えると減点には除算法を用いることが適当であると判断した。

### 2-3-4 並列処理

ここまで一致比較を行列で表現した図をいくつか紹介してきたが、これらのノード数は比較する塩基数の2乗となっている。そこで、これらのノードが逐次処理を行うとすると、塩基数の増加に伴って処理時間が非常に長くなってしまいうという問題が生じてくる。その問題を改善するために、図2-7のように各ノードを接続することによって並列処理の実現を試みた。図2-10の例において並列処理が可能となれば、灰色のノードの点数が決定していれば、黒色のノードは独立して処理を行うことができることになる。そのため、処理時間は、右下から左上の対角線と直行する対角線に平行な方向に並ぶノードの列数で決まり、 $n$ 個からなる二つの塩基配列の点数処理を行う場合のこのノードの列数は $(2n-1)$ 列となり、ほぼ塩基数の2倍の列数だと見なせる。

よって、実際にかかる計算時間は、比較する塩基配列の塩基数に比例した時間で済むことになる。

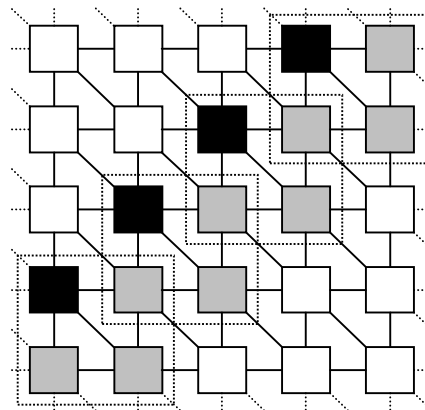


図 2-10 ノードの並列処理

### 2-3-5 類似度判定法の評価

以上のアルゴリズムに対して C 言語を用いたシミュレーションを行った。図2-11に示されているパラメータを用いた時に、置換、欠落、混入が混合して発生した場合の各ノードでの点数を表したものを図2-12に示す。ここで図2-11にあるしきい値とは、変化数によって減点の割合を変えるためのものである。

	しきい値以上	しきい値以下	しきい値
P-1	0.00	0.00	80点
P-2	0.60	2.40	
P-3	0.45	1.80	
P-4	0.60	2.40	
P-5	1.00	4.00	
P-6	0.75	3.00	

図 2-11 パラメータ (図 2-12)

ここで、最上行と最左列の灰色の塩基は変化が生じた箇所を表していて、灰色のノードは変化によって「ずれ」が生じた箇所、つまり減点箇所を示している。また、太い枠で囲ってあるノードは信頼度の高い点数を伝達するノードの流れを示している。図 2-12 から、提案したアルゴリズムに基づいて比較処理が行われていることが判る。

	A	C	T	C	A	G	C	T	G	C
A	77	78	76	81	90	82	83	90	91	96
G	61	77	82	79	81	90	86	87	96	96
T	73	74	82	86	82	83	90	91	91	96
A	81	75	76	82	91	82	87	94	91	96
G	84	81	78	76	82	91	83	90	100	96
C	82	90	82	83	79	84	91	87	91	100
T	86	86	90	86	83	83	88	91	91	96
T	90	90	91	90	90	87	87	94	91	96
A	100	91	94	91	96	94	91	91	94	96
C	96	100	96	100	96	96	100	96	96	100

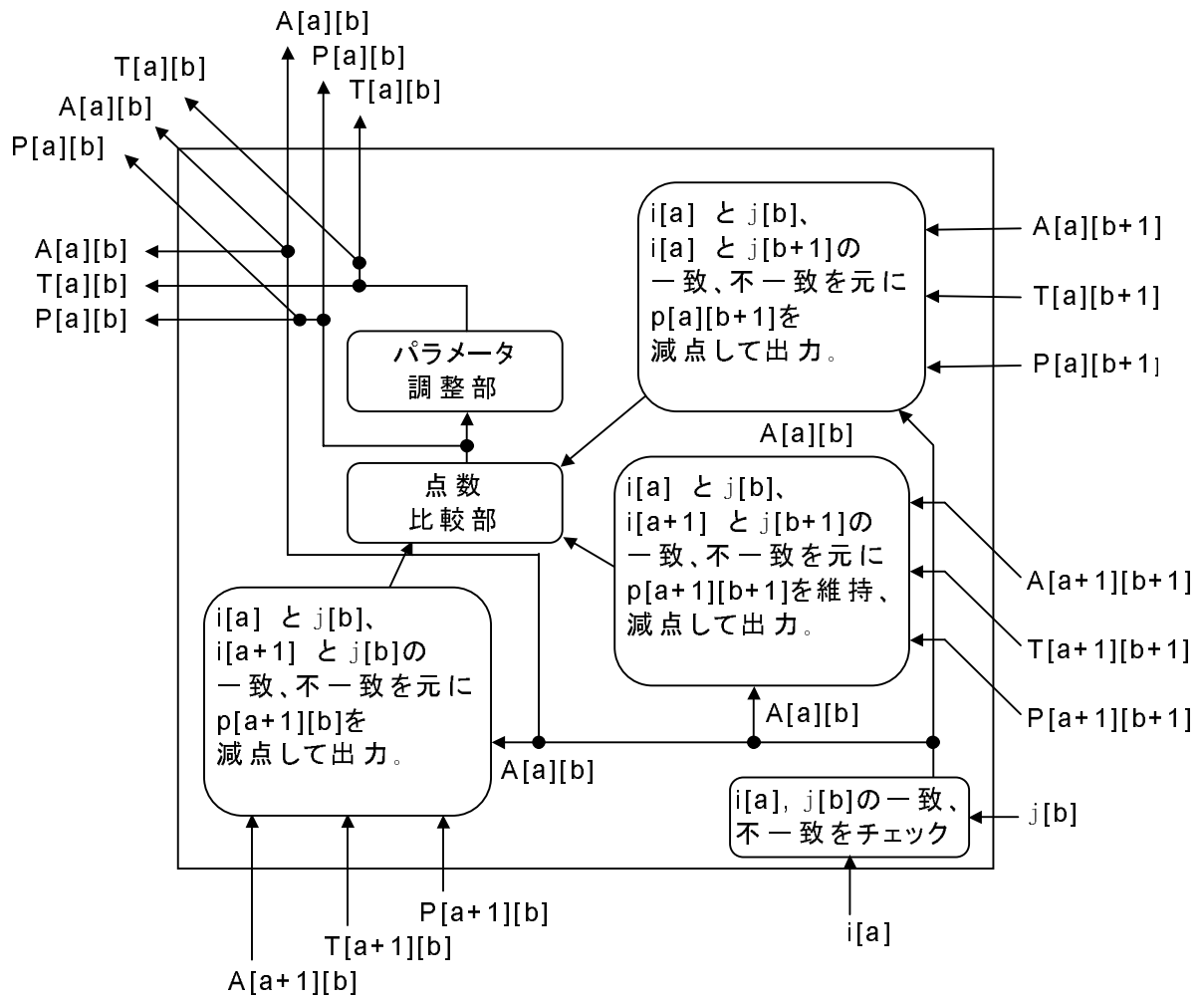
図 2-12 各ノードの出力(塩基変化(3種混合))

## 第3章 高速類似度判定回路の構成

### 3-1 各ノードの回路構成

#### 3-1-1 ノードのブロック図

一つのノードの構成をブロック図で表現すると図 3-1 のようになる。



- $i[x], j[y]$ : 塩基情報  
 $A[x][y]$ :  $i[x], j[y]$ の一致情報  
 $P[x][y]$ :  $i[x], j[y]$ のノードにおける点数  
 $T[x][y]$ :  $P[x][y]$ がしきい値以上かどうかの情報

図 3-1 ノードのブロック図

ここで一致情報を A、点数を P、パラメータ調整情報を T としている。この回路は対応する塩基の一致、不一致を調べる「一致比較部」と減点を行う「減点部」、最高点を選択する「点数比較部」、そして「パラメータ調整部」から成っている。

A[x][y]と T[x][y]によって減点する際のパラメータを決定して P[x][y]の減点を行い、点数比較部で3つの減点部から出力された点数のうち最も高い点数を選択し次のノードへと出力するものとなっている。また、パラメータ調整部で選択された点数としきい値との比較を行い、その情報も次のノードへと出力している。

### 3-1-2 ノードの回路図

ノードを電子回路と論理回路で表現したものを図 3-2、3-3 に示す。ここで図の都合上 2 つに分けてあるが Out1、2、3 で繋がっており、入力を赤色、出力を青色で表してある。図 3-2 の黒枠で囲ってある部分は一致比較部であり、塩基が 4 種類であるため 4 値の一致、不一致を調べることができるものとなっている。黒色点線枠で囲ってある部分は減点部であり、3 方向からの入力に対応して減点部は大きく見て 3 つあるが、2 種類のパラメータ用にそれぞれが 2 種類の減点部分から構成されている。減点処理には抵抗を用いており、減点量を抵抗の比で決定することで除算を実現している。灰色点線枠で囲ってある部分はスイッチ部分であり、2 種類の減点部から出力された点数をパラメータ調整情報によって選択している。また図 3-3 では、1 つ目の比較器で減点部から出力された 3 つの点数のうち、2 つの点数比較を行いスイッチによって高い方の点数を選択し、その点数と残りの点数を次の比較器で比較処理を行い、同様にスイッチで処理をし、そのノードでの点数を決定して次のノードへと出力している。以上が点数比較部の処理となる。また、最終的に出力される点数としきい値電圧とを比較器で処理することによりパラメータ調整情報を出力している。図 3-3 の黒色点線枠で囲ってある部分は buffer であり、他のノードとの影響を取り除くために出力部分に設置してある。



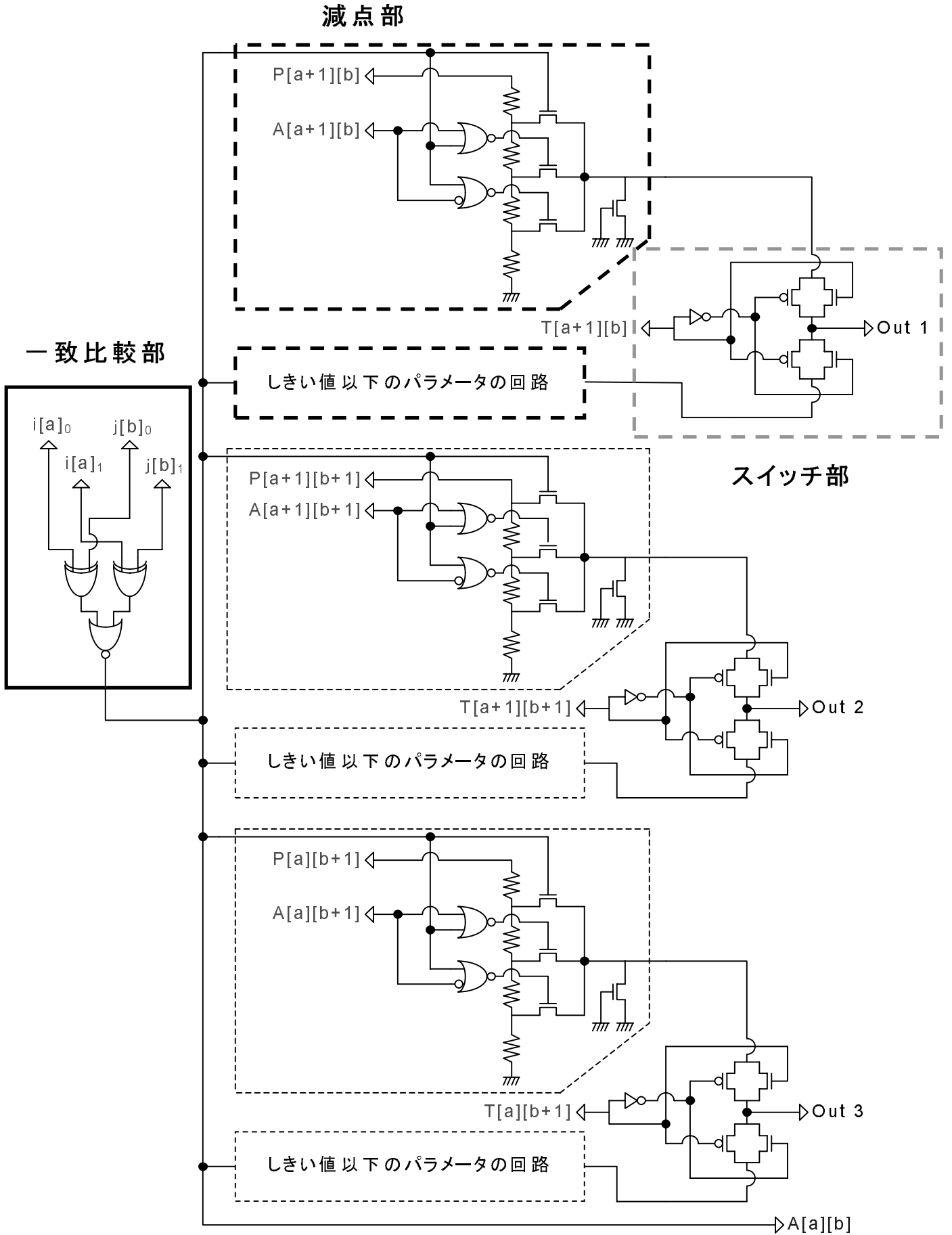


図 3-2 類似度判定回路ノード部 (a)

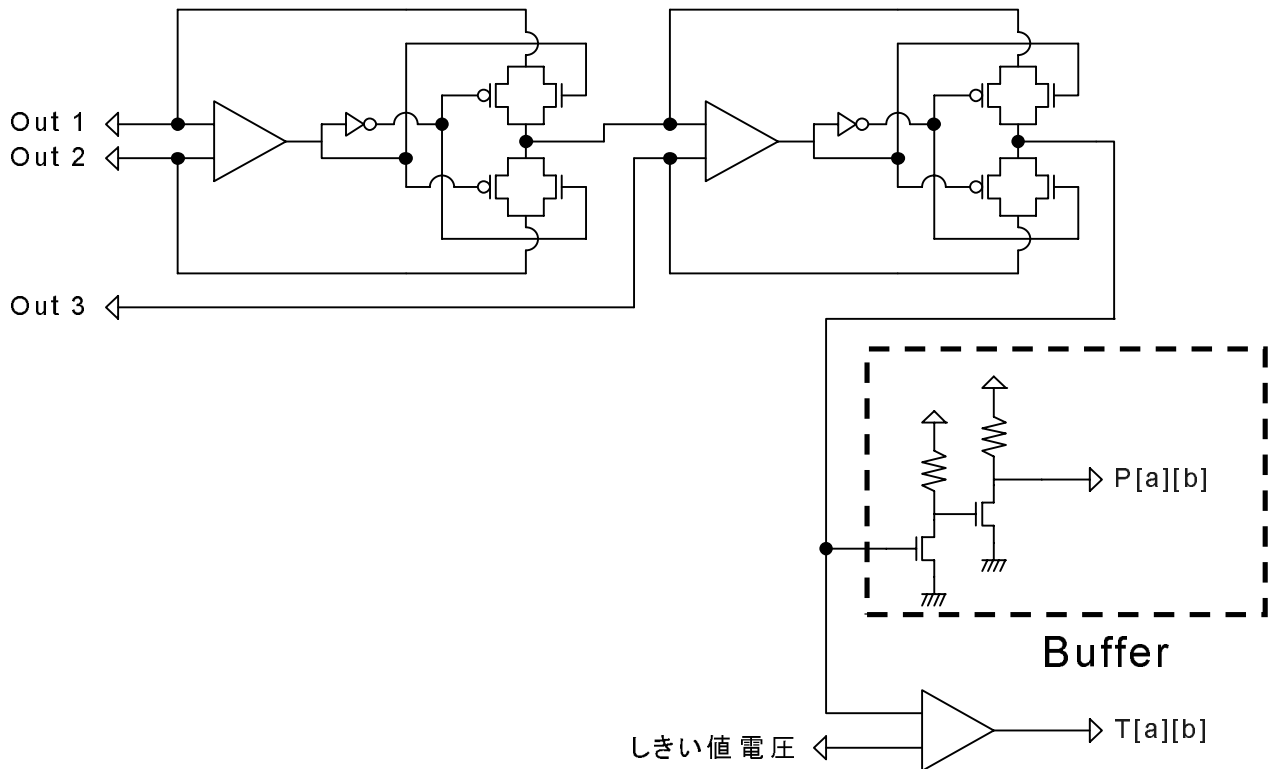


図 3-3 類似度判定回路ノード部 (b)

### 3-2 全体の回路構成

以上のノードを行列状に配置し、塩基配列情報を入力する部分にはシフトレジスタを配置し、それらをアルゴリズムに基づいて配線したものが図 3-4 である。まずデータベースから比較する塩基数分だけ CLK2 によって入力し、もう一方の塩基配列を CLK1 によって入力する。次に、この類似度判定回路が処理を始めてから安定するまでの時間を CLK2 の 1 クロックとし、クロック毎に 1 塩基毎シフトして入力させる。この回路は「塩基情報の入力から、左上端のノードの出力が安定するまで」を 1 回の処理とする回路であるため、類似度を計算するという点ではクロックに依存しないものとなっている。また、ノード 1 つにつきトランジスタ数は 206 個であり、回路全体で使用するピン数は 7 本である。

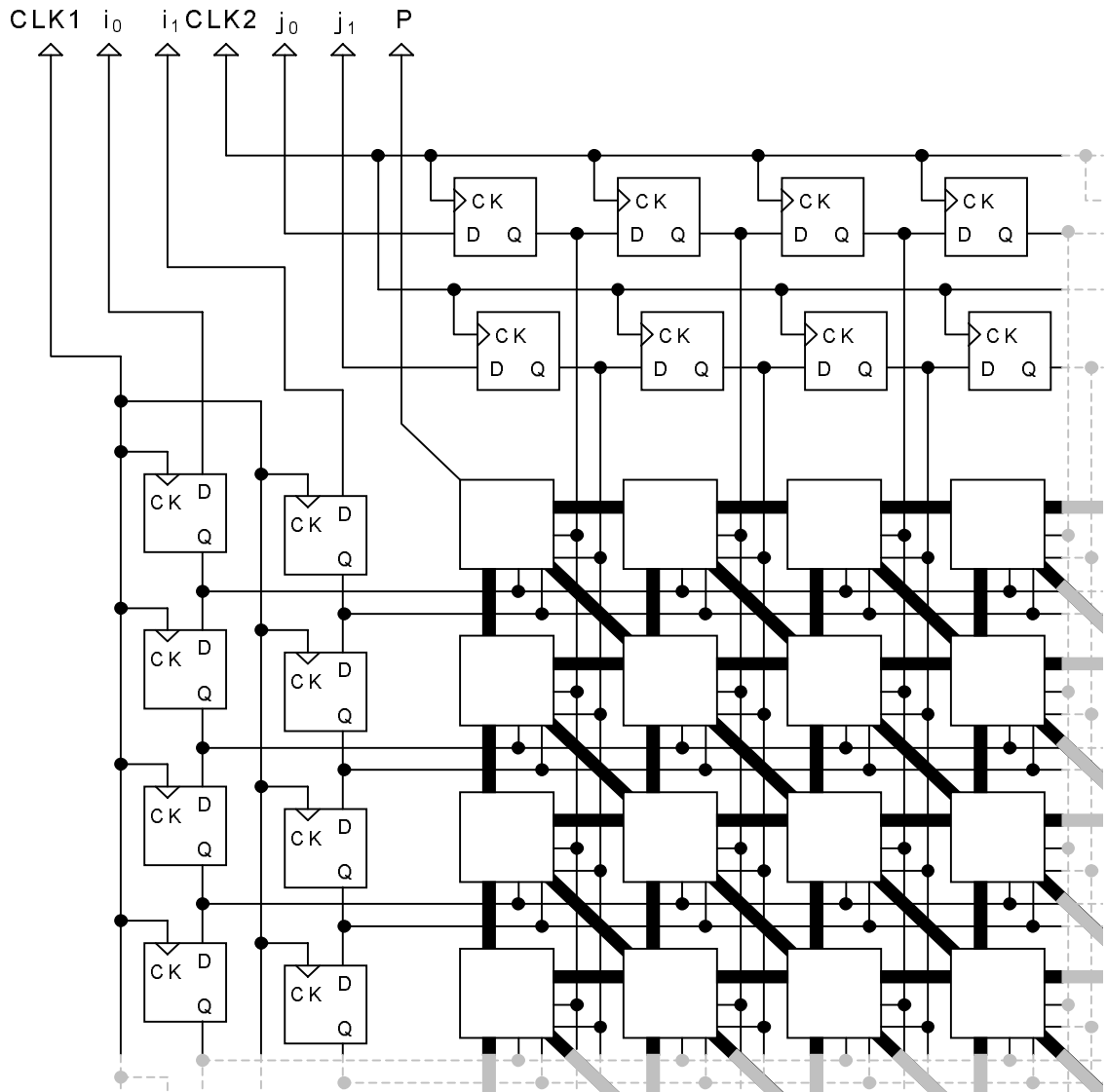


図 3-4 ノードの行列状配置構造による類似度判定回路

### 3-3 buffer の特性

前述したように、この回路には他のノードとの影響を除くために buffer を用いている。buffer は低入力インピーダンスかつ高出力インピーダンスでなければならず、入力と出力が線形、つまり電圧利得が 1 に近くなければならない。そこで今回我々が使用した buffer の特性を図 3-5 に示す。図に示されているように、利得が 1 として見なすことができる範囲は 2.25V~3.75V である。ここで、2.25V~3.0V 間では入力に対して出力が降下していて、3.0V~3.75V 間では上昇している。この入出力差は約 0.019V となっていて、高い点数はより高く、低い点数はより低くなる傾向となり、類似を判定するという目的から、大きな問題はないと考えた。以上より 0

点を 1.75V、100 点を 3.75V、つまり回路で用いる電圧の領域幅を 2V とした。ここで 2.25V、点数で表すと 25 点を下回らないように C 言語のプログラムを調整し、それに従った減点部の回路構成を行った。

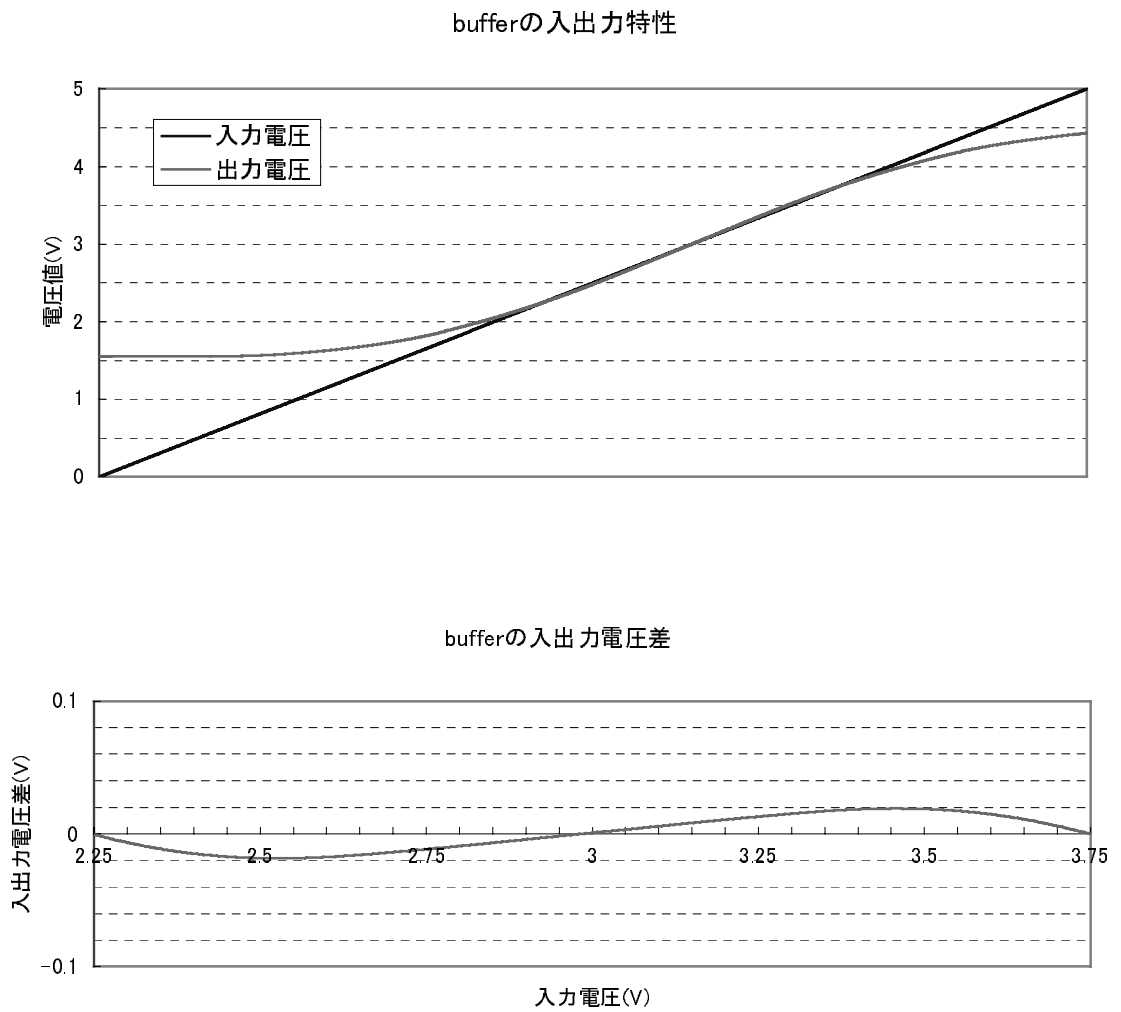


図 3-5 buffer の入出力特性

### 3-4 HSPICE によるシミュレーションと結果

#### 3-4-1 HSPICE と C 言語とによるシミュレーション結果の比較

これらの要素を持つノードを行列状に 9×9 で配置し、図 3-6 のパラメータを用いて HSPICE で回路シミュレーションを行った。図 3-7 に HSPICE での結果、図 3-8 に同じ条件で行った C 言語でのシミュレーション結果、図 3-9 にその差、図 3-10 に出力される点数に対する差の割合を示す。

	しきい値以上	しきい値以下	しきい値
P-1	0.00	0.00	80 点
P-2	0.30	1.20	
P-3	0.225	0.90	
P-4	0.30	1.20	
P-5	0.50	2.00	
P-6	0.375	1.50	

図 3-6 パラメータ(回路シミュレーション)

	A	C	T	G	A	C	T	T	G
A	85.84	79.16	81.09	86.34	89.11	88.10	91.22	99.79	96.20
G	85.82	84.88	81.28	84.15	85.43	93.69	91.24	92.57	99.83
T	83.15	88.90	89.40	82.80	87.30	88.28	93.15	94.66	96.10
A	84.50	86.25	92.18	88.64	83.71	91.89	91.53	92.57	99.83
G	92.62	83.71	88.64	97.09	88.45	86.09	91.24	96.39	96.10
C	88.38	91.89	86.09	88.45	96.88	88.23	89.27	94.66	96.10
T	89.37	91.53	91.24	89.27	89.27	96.65	89.27	92.57	99.83
T	94.76	92.57	96.39	94.66	92.57	92.57	96.39	92.57	96.10
A	96.20	99.83	96.10	96.10	99.83	96.10	96.10	96.10	96.10

図 3-7 類似度判定回路のシミュレーション結果

	A	G	T	A	G	C	T	A	C
A	84.92	77.60	79.89	85.11	88.19	87.32	90.64	100.18	95.62
T	84.93	83.98	79.99	83.28	84.16	93.26	90.82	91.78	99.99
C	81.94	88.38	88.91	81.58	86.69	87.39	92.52	94.48	95.62
T	83.40	85.38	91.91	88.02	82.63	91.71	90.90	91.78	99.99
C	92.45	82.63	88.02	97.34	87.85	84.81	90.82	96.13	95.62
G	87.39	91.71	84.81	87.85	97.15	87.37	88.26	94.48	95.62
A	88.26	90.90	90.82	88.26	88.26	96.64	88.26	91.78	99.99
C	94.48	91.78	96.13	94.48	91.78	91.78	96.13	91.78	95.62
T	95.62	99.99	95.62	95.62	99.99	95.62	95.62	95.62	95.62

図 3-8 C 言語でのシミュレーション結果

	A	G	T	A	G	C	T	A	C
A	0.92	1.57	1.20	1.23	0.92	0.78	0.57	-0.39	0.58
T	0.89	0.90	1.30	0.87	1.27	0.42	0.42	0.79	-0.16
C	1.21	0.51	0.49	1.22	0.60	0.89	0.62	0.17	0.47
T	1.10	0.87	0.27	0.62	1.08	0.18	0.64	0.79	-0.16
C	0.16	1.08	0.62	-0.24	0.60	1.29	0.42	0.26	0.47
G	0.99	0.18	1.29	0.60	-0.27	0.86	1.02	0.17	0.47
A	1.12	0.64	0.42	1.02	1.02	0.01	1.02	0.79	-0.16
C	0.28	0.79	0.26	0.17	0.79	0.79	0.26	0.79	0.47
T	0.58	-0.16	0.47	0.47	-0.16	0.47	0.47	0.47	0.47

図 3-9 HSPICE でのシミュレーションと C 言語でのシミュレーションとの差

	A	G	T	A	G	C	T	A	C
A	1.08	2.02	1.27	1.44	1.04	0.89	0.63	-0.39	0.61
T	1.04	1.07	1.63	1.04	1.51	0.45	0.46	0.86	-0.16
C	1.48	0.58	0.55	1.50	0.69	1.02	0.67	0.18	0.49
T	1.31	1.02	0.29	0.70	1.30	0.20	0.70	0.86	-0.16
C	0.17	1.30	0.70	0.27	0.68	1.52	0.46	0.27	0.49
G	1.01	0.20	1052	0.68	-0.28	0.98	1.16	0.18	0.49
A	1.27	0.70	0.46	1.16	1.16	0.01	1.16	0.86	0.16
C	0.30	0.86	0.27	0.18	0.86	0.86	0.27	0.86	0.49
T	0.61	0.16	0.49	0.49	0.16	0.49	0.49	0.49	0.49

図 3-10 出力される点数に対する差の割合 (%)

この結果より、差は最終出力、つまり類似度で 0.92 点、各ノードでは最高で 1.57 点であることが判り、出力される点数に対しての割合はそれぞれ 1.08%、2.02%であった。これは類似度を判定することが本研究の目的であることや、ノードごとに生じる buffer や比較器などの素子による影響を考慮すると、この差には問題はないと考えられる。

### 3-4-2 遅延時間

この高速類似度判定回路の処理時間を求めた。対角線上のノードの出力電圧から処理時間が求められることができるので、完全一致の場合と塩基置換の場合とのを比較行った。完全一致の場合を表したものを図 3-11 に示す。完全一致の場合は node99 に多少のぶれが生じ、その他のノードにも若干のぶれが生じてはいるものの電圧値で 0.002V、点数にすれば 0.1 点のぶれであるため影響はないと判断した。次に塩基置換が生じた場合の出力電圧を図 3-12 に示し、置換、欠落、混入の 3 パターンが混合して生じた場合のシミュレーション結果を図 3-13 に示す。図から判るように、いずれのノードでも電圧降下と時間との傾きはほぼ一致していて、図 3-12 に点線で示されているようにこの回路で使用する最低電圧である 2.25V へ低下し、安定するまでには約 230ns かかる。このことと、ノードの伝搬遅延が 5ns 程度であること

から $9 \times 9$ にノードを配置した場合には、電圧降下にかかる時間は約 $150\text{ns}$ であることが判った。そこで100塩基からなる塩基配列を比較した場合、この回路は約 $1\mu\text{s}$ つまり $1\text{MHz}$ で動作するため、1秒間に約100万塩基対の類似度判定が可能であると考えられる。

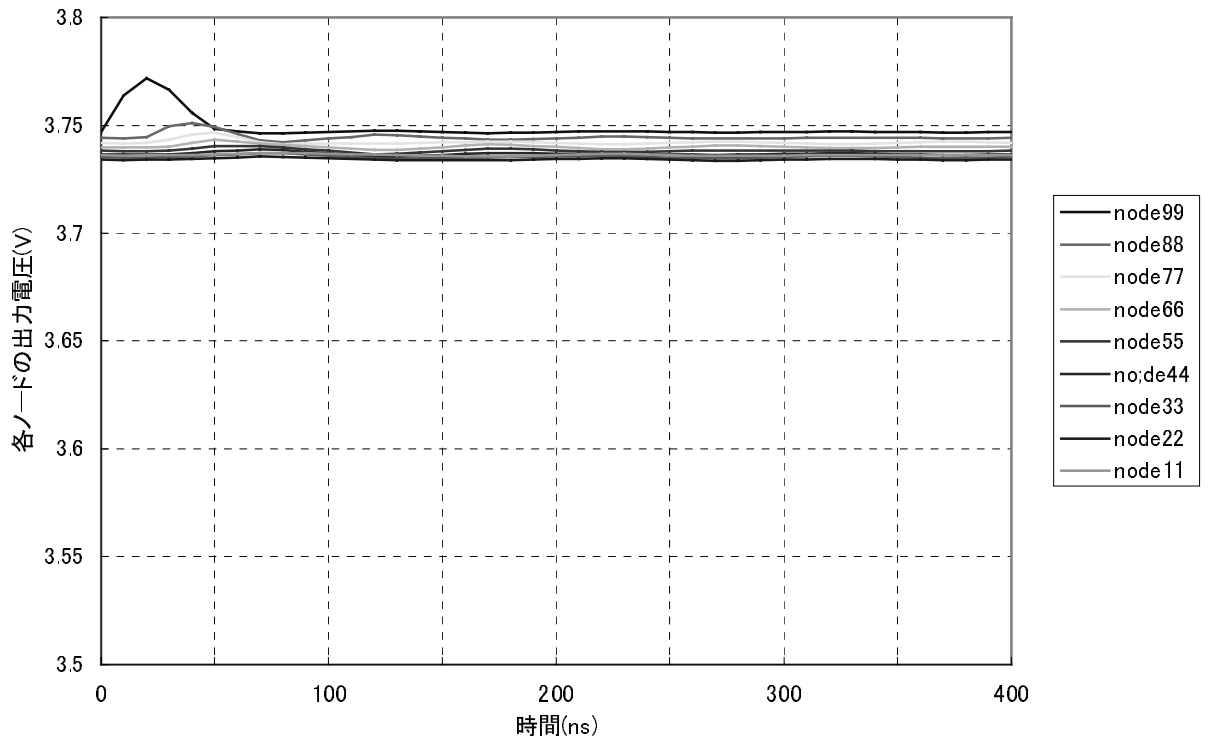


図 3-11 対角線上にあるノードの出力電圧(完全一致)



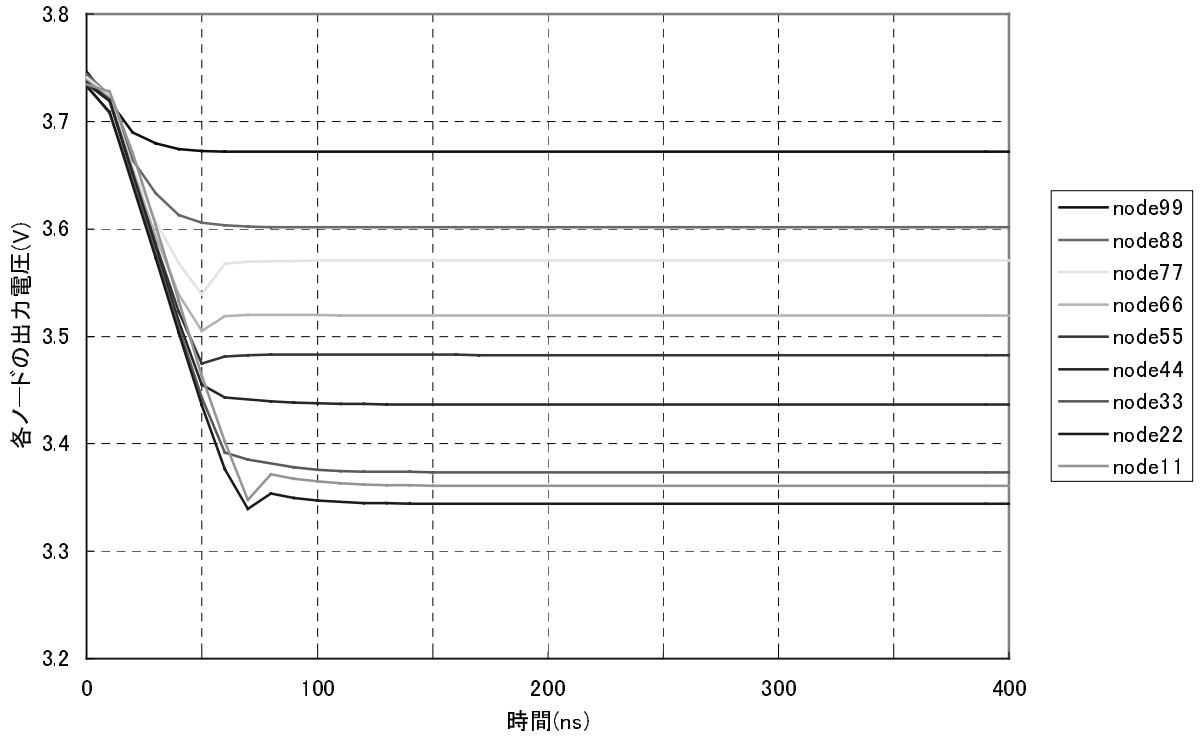


図 3-12 対角線上にあるノードの出力電圧(塩基置換)

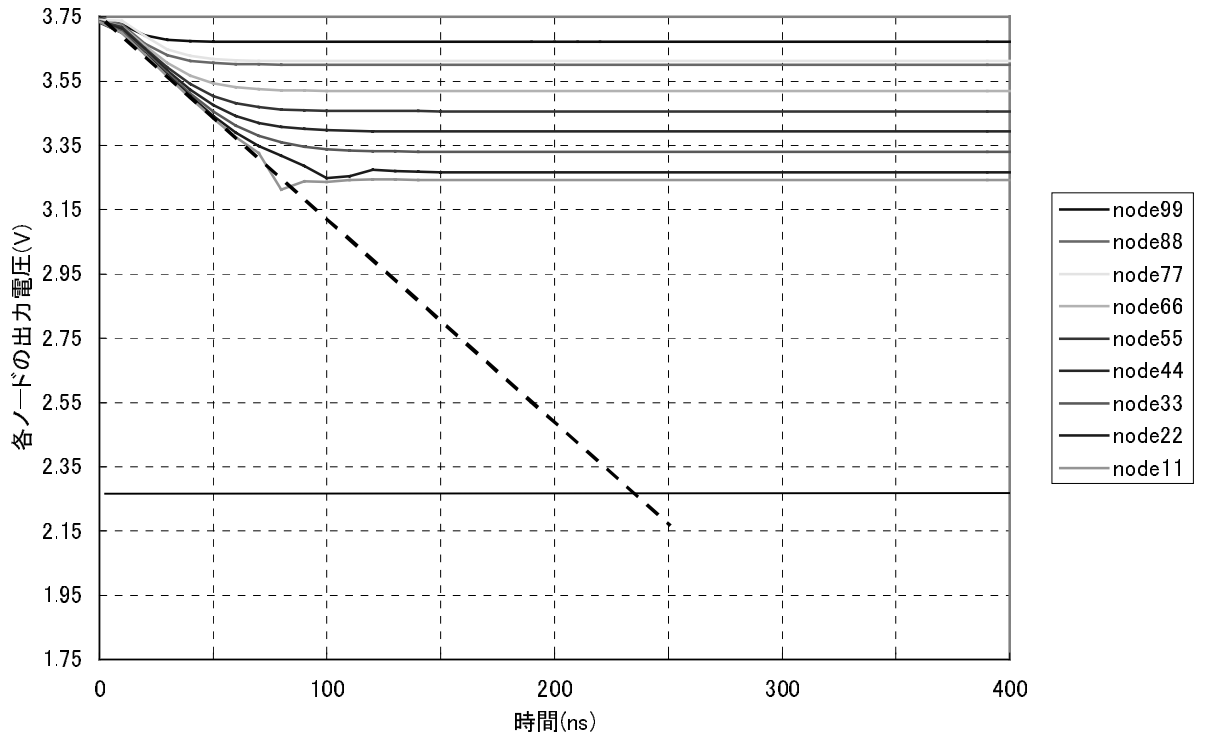


図 3-13 対角線上にあるノードの出力電圧(塩基変化(3種混合))

## 第4章 まとめ

### 4-1 まとめ

- 正常な塩基配列と置換、欠落、混入の変化が生じた塩基配列とを行列表の表を用いて比較解析を行った。その結果、対角線を中心に一致を表す「○」の列がずれており、そのずれのパターンは3種類あることが判った。解析法とずれのパターンより、処理を行うノードの配置配線を決定した。配置配線より各ノードの構造を「右」、「右下」、「下」のノードからの3入力とし、「上」、「左上」、「左」のノードへの3出力とした
- 今回提案する高速類似度判定回路では減点部分に抵抗を用いることによって、点数を電圧で表現できるようになった。また、抵抗を用いることで除算法での減点が実現した。この回路はNOT、NOR、EXOR、抵抗、比較器、bufferから構成されており、トランジスタ数は「206」である
- 同じ条件のもとで行ったC言語とHSPICEとによるシミュレーション結果の比較から、多少のずれはあるもののほぼ一致する結果が得られた。また、処理速度はノード数ではなく塩基数に比例することから高速な処理が可能であるということがHSPICEによるシミュレーション結果から判った
- プログラムおよび回路に適切な修正を加えて、C言語とHSPICEのシミュレーション結果の差をより小さくすることによって、実際の回路により近いシミュレーションを行うことが期待できる

## 謝辞

本研究を行うにあたり、様々な方々からの多くのご指導、ご協力を頂戴いたしました。この場を借りて心から感謝の意を表します。

指導教官としてご指導を頂いた金沢大学集積回路工学研究室教授 故鈴木正國先生に深く感謝すると共に、心からご冥福をお祈りいたします。

ご多忙の中、研究面や生活面に関して多くのご指導を頂いた北川章夫助教授に深く感謝いたします。研究を進めるにあたって公私を問わず、ご指導、ご協力くださった秋田純一助手に心から感謝いたします。2ヶ月と短い間でしたが、研究面において様々なご指導を頂いた深山正幸助手に感謝いたします。VLSI設計室の設備や管理作業において多くのご協力を頂いた柿本芳雄文部技官に深く感謝いたします。また、研究生活をおくるにあたり多くのご協力を頂いた金沢大学医学部保険学科の中山和也助手に心から感謝いたします。

研究面でご協力下さった東京大学大規模集積システム設計教育センター、日本モトローラ（株）、ローム（株）の関係者の方々に感謝いたします

本研究の共同研究者として研究面のみならず生活面でもご指導、ご協力頂いた素子工学研究室 博士前期課程2年の藤井直樹さんに心から感謝いたします。

半年間と短い間でしたがご指導、ご協力頂いた Dr. Nandana Fernando 氏に深く感謝いたします。

研究および学生生活においてご指導、ご協力頂いた博士前期課程2年の小川明宏さん、高瀬信二さん、中橋憲彦さん、早川史人さん、博士前期課程1年の今井豊さん、数馬晋吾さん、藤田隼人さん、水野浩樹さん、渡辺晃さん 電磁波通信工学研究室の村上崇さんに深く感謝いたします。研究室、日常生活で苦楽を共にした笠井稔彦君、大門慎治君、高松直樹君、辻川隆俊君、遠山治君、中村公亮君、蓮達弘君、水木誠君に深く感謝いたします。

その他、学生生活において様々な面で多くの方々にご指導、ご協力を頂いたことに心から感謝いたします。

## 参考文献

- 黒田行昭著「基礎遺伝学」（近代遺伝学の流れ）裳華房（1995）
- 宝来聡著「DNA 人類進化学」岩波科学ライブラリー52
- 国立遺伝子研究所「電子博物館」第1版（1999）
- 原 央 編著 「MOS 集積回路の基礎」近代科学社（1992）