2021年6月15日(火)

集積回路システム工学　第9回講義

アナログ集積回路 調査研究事例

ビジョンチップ、抵抗ネットワーク

小林春夫

群馬大学大学院理工学府　電子情報部門

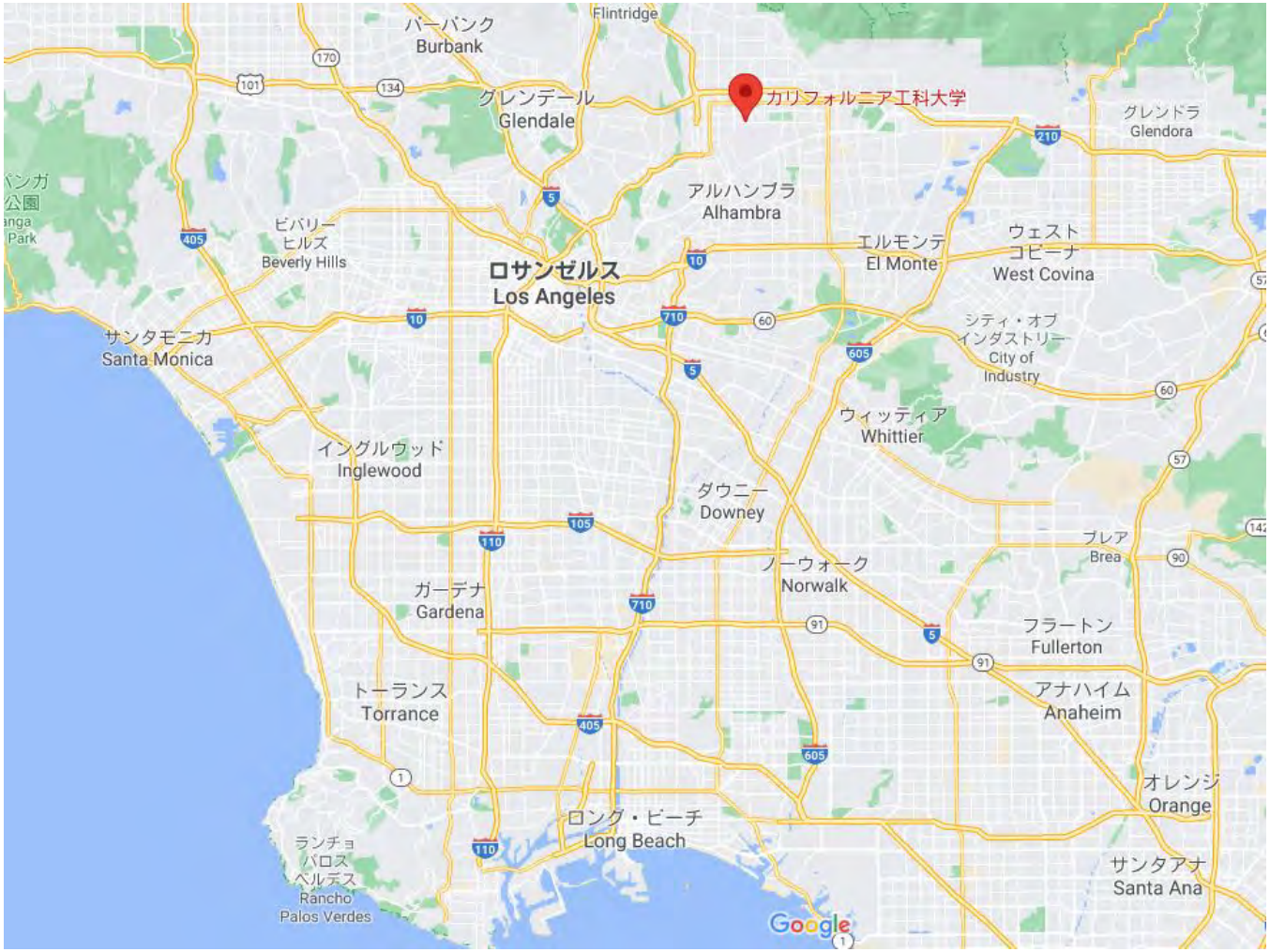
koba@gunma-u.ac.jp

下記から講義使用　pdfファイルをダウンロードしてください。

出席・講義感想もここから入力してください。

https://kobaweb.ei.st.gunma-u.ac.jp/lecture/lecture.html

ロサンゼルス地区


バーバンク
Burbank
Flintridge
カリフォルニア工科大学
グレンデール
Glendale
グレンドラ
Glendora
アルハンブラ
Alhambra
ビバリーヒルズ
Beverly Hills
ロサンゼルス
Los Angeles
エルモンテ
El Monte
ウェストコビーナ
West Covina
サンタモニカ
Santa Monica
シティ・オブインダストリー
City of Industry
ウィッティア
Whittier
イングルウッド
Inglewood
ダウニー
Downey
ブレア
Brea
ノーウォーク
Norwalk
ガーデナ
Gardena
フラートン
Fullerton
トーランス
Torrance
アナハイム
Anaheim
オレンジ
Orange
ロング・ビーチ
Long Beach
ランチョパロスベルデス
Rancho Palos Verdes
サンタアナ
Santa Ana
Google

TECHNICAL SURVEY

解 説

ビジョンチップ(I)
——アナログ画像処理用ニューロチップ——

松本 隆 小林春夫 八木哲也

松本 隆：正員 早稲田大学理工学部電気工学科
小林春夫：正員 横河電機株式会社エレクトロニクス研究所
八木哲也：正員 九州工業大学情報工学部制御システム工学科

Vision Chip (I) : Analog Image-Processing Neuro Chip. By Takashi MATSUMOTO, Member (School of Science and Engineering, Waseda University, Tokyo, 169 Japan), Haruo KOBAYASHI, Member (Electronics Laboratory, Yokogawa Electric Corp., Musashino-shi, 180 Japan) and Tetsuya YAGI, Member (Faculty of Information Engineering, Kyusyu Institute of Technology, Iizuka-shi, 820 Japan.

ABSTRACT

ビジョンチップは画像入力センサを持ち，エッジ検出等の初期視覚アルゴリズムの超並列処理を行う画像処理用アナログ CMOS VLSI である．Caltech の Mead によって 1980 年後半に提唱されて以来，米国の大学を中心に研究開発が行われ，さまざまなビジョンチップが提案・実現されてきており，近年ベンチャー企業からは本格的な製品も現れている．ここではビジョンチップの原理・アルゴリズム，従来技術に対する特長，新しく導かれた回路網理論上の定理，モデルになっている網膜の生理学的背景，実現のための VLSI 技術の背景，各研究機関での研究活動，開発されたいくつかのチップの紹介を行う．

キーワード：ビジョンチップ，ニューロチップ，画像処理，画像センサ，網膜

1. はじめに

近年，従来のディジタルコンピュータが不得手とした問題に対して有効性が期待されているニューラルネットワークが関心を集めている．また，VLSI 技術の進歩により，このアルゴリズムを高速に実行するニューロチップの研究開発も盛んになされている．この中で Caltech の Mead により提唱・試作された，脊椎動物の網膜をモデルにしたビジョンチップの研究開発を契機として，生体系をモデルとしたアルゴリズムばかりでなく Computer Vision アルゴリズムをアナログ VLSI 上に実現する試みも活発に行われ，実用化に近づいている[1]．これらの研究は Caltech，MIT，CMU，UCLA などの米国の大学を中心に行われているが，大学とビジョンチップの共同研究をする企業も現れ，ベンチャー企業からは本格的な製品も発売されている．またこの開発過程でいくつかの理論的側面の問題提起がなされ，その結果回路網理論上の新しい定理も導き出されている．

2. 原 理

2.1 画像認識のための前処理とビジョンチップ

画像認識のための処理の流れは，画像入力，画像の前処理，画像認識からなる．画像の前処理は，平滑化によるノイズ除去，エッジ検出などの基本的な処理で初期視覚問題（Early Vision Problem）とよばれている[2]．Caltech の Mead は脊椎動物の網膜の機能の一部（画像入力と初期視覚に対応する画像処理）を調べ，アナログ CMOS VLSI 上にそれを実現した[1]．アルゴリズムの新規性，光センサ埋蔵，MOS トランジスタをいわゆるサブスレッショルド領域で動作させていること等の理由で大きなインパ

クトを与えている．これらの画像センサと初期視覚画像処理機能を持つアナログVLSIはビジョンチップとよばれている．

2.2 画像処理システムの構成

(a) 従来のシステム構成

従来の画像処理システムでは図1に示すように，CCDセンサ内蔵のTVカメラで画像を入力し，A-D変換器で逐次的にディジタル信号に変換し，汎用ディジタル画像処理プロセッサで処理を行う．この方式はさまざまなアルゴリズムの画像処理ができるものの，大規模ハードウェアが必要で，逐次的ディジタル計算方式は大量データの画像を高速に処理するのに不向きである．

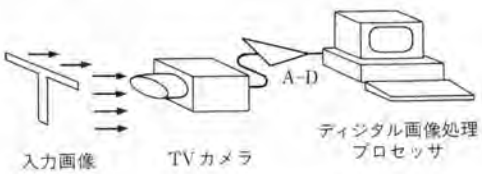


図1 従来の画像処理システム　CCDで画像入力しA-D変換を行った後ディジタル画像処理する．

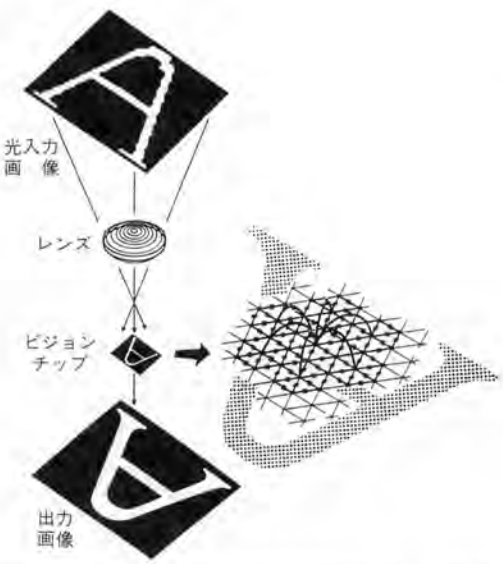


図2 ビジョンチップを用いた画像処理システム　光入力画像がレンズでビジョンチップ上の光センサアレーに集光され，ビジョンチップはその入力信号に対し初期視覚アルゴリズムを高速にアナログ並列処理する．©1990 IEEE.

(b) ビジョンチップを用いた構成

ビジョンチップを用いた画像処理システムは図2に示すようになる．光入力画像をレンズでビジョンチップ上の光センサアレーに集光し，その入力信号に対し初期視覚アルゴリズムを高速にアナログ**並列処理**する．その出力をA-D変換し汎用ディジタル画像処理プロセッサで認識その他の処理を行う．ビジョンチップは画像入力だけでなく画像の前処理を超高速に行うので後段のプロセッサの負荷を軽減する．

2.3 従来技術に対する特長

(a) ビジョンチップの計算基本原理

ビジョンチップの計算基本原理はディジタル信号処理の計算原理と本質的に異なる．この解説で紹介するチップはすべて**アナログ超並列抵抗回路網**†をCMOSで実現したものであり，計算は超並列抵抗回路網の**物理現象**が行う．より具体的には，(i) 和 (addition) は**キルヒホッフ電流則**が行い，(ii) 計算の実行は**ダイナミックス**が遂行する．入力は各抵抗ノードに加えられた電圧または電流であり，出力は**ダイナミカルシステムの安定平衡点の電圧分布**として与えられる．またダイナミカルシステムが安定平衡点に達するまで全ノードの電圧・電流が並列に動作するので，画像処理が完全に並列に行われる．

(b) ビジョンチップvsディジタルシステム

ビジョンチップは従来のディジタル画像処理システムに比べ，次のような特長がある．

(i) **小規模ハードウェア**　より具体的には配線の複雑さが低減される．例えば入力画像に大きな画素数のコンボリューションを施す場合，逐次計算方式のディジタルプロセッサでこれを実行すると極めて計算量が多くなる．一方，並列抵抗回路でコンボリューションをとることは圧倒的に単純である．直感的にいうと，あるノードに入力された電圧または電流の影響は

† ここでいう抵抗はオームの法則に従う受動線形素子のみではなく，能動素子，非線形素子を含めた広い意味である．

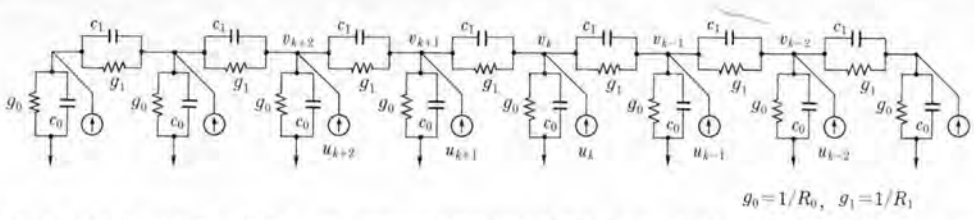


図3 1次正則化問題と抵抗回路網　第1近傍への正抵抗結合 R_1 のネットワークで1次正則化問題が解ける．入力は各ノードへの電流で，出力は平衡状態に達したときの各ノード電圧である．

“瞬間的に”他のすべてのノードに伝えられるからである．式による説明のほうを好む読者のため図3のような並列抵抗回路を考える．平衡点における各ノード i でのキルヒホッフ電流則は，

$$-(g_0+2g_1)v_i+g_1(v_{i+1}+v_{i-1})+u_i=0 \quad (1)$$

である．いま式（1）の v_i を中心に考えてみると，そこに現れるのは v_{i+1}，v_{i-1} すなわち**最近傍**の変数のみである．にもかかわらず大きな画素数のコンボリューションがとれるのは空間変数 i に関して IIR（Infinite Impulse Response）構造をしているからである．これを FIR（Finite Impulse Response）型ネットワークで実現しようとすると，膨大なノード間結合が必要となる．式（1）は v_i，v_{i+1}，v_{i-1} の線形結合の計算であるが，キルヒホッフ電流則と抵抗特性の物理現象がそれを自然に遂行していることがわかる．このような構造はここで紹介するビジョンチップ全体に共通していえることである．信号処理チップを実現する際に大きな問題になるのは演算素子間の配線の複雑さであり，それはシリコンでだけでなく生体（網膜はもとより脳の他の部分）においても同様の拘束条件になっており，Mead は，これが“the single most important”な事実であると主張している[11]．

（ii） **高速（リアルタイム）処理**　ダイナミカルシステムが平衡点に収束する時間が処理時間に当たるが，現在の CMOS 技術での寄生容量は極めて小さいのでせいぜい数マイクロ秒で完了する．

（iii） **低消費電力**　例えばサブスレッショルド領域の CMOS 回路を用いれば著しい電力の節約となる．

しかしながら，ディジタルシステムは，（i）**高精度**，（ii）**柔軟なプログラミングが可能**という点が優れている．ビジョンチップは自動走行システム，ロボットビジョン，ファクトリオートメーションでの目視検査などの，上記の長所が発揮しうるところに実用化できよう．

2.4 正則化問題と抵抗回路網

（a） 正則化問題（Regularization Problem）

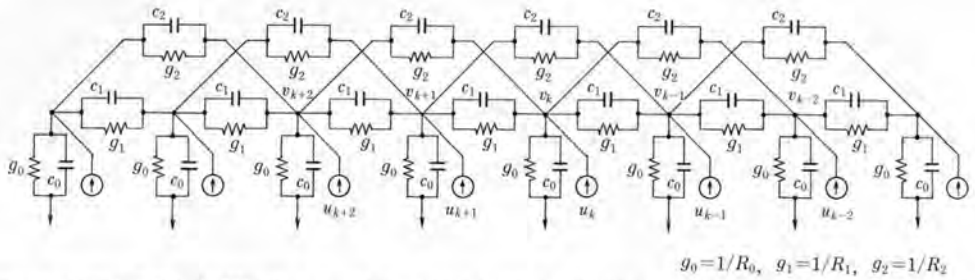
MIT 人工知能研究所の Poggio らは，エッジ検出，ステレオ視などの初期視覚問題を“不良設定問題”（Ill-Posed Problem）としてとらえ，それを正則化（Regularize）することにより，well-posed な問題になると考えた[3),4)]．すなわちノルム空間上の汎関数

$$G_p(v, d)=\|Av-d\|^2 + \sum_{k=1}^{p}\int \lambda_k(x)\left(\frac{d^k v(x)}{dx^k}\right)^2 dx$$

の最小化問題に帰着させた．但し，A は考えている正規化問題に固有の作用素，$\|\cdot\|$ は問題に固有のノルム，そして第2項以後は滑らかさに対するペナルティである．簡単のため1次元の場合を書いたが，2次元の場合は特有の問題も生じる．あるクラスの正規化問題が並列抵抗回路で自然に解けることを示すため，空間変数を離散化し，また A が恒等写像の場合を考えると，$p=1$ の場合は $\boldsymbol{v}=(v_1, \cdots, v_n)$，$\boldsymbol{d}=(d_1, \cdots, d_n)$ として

$$G_1(\boldsymbol{v}, \boldsymbol{d})=\sum_{k=1}^{n}(v_k-d_k)^2+\lambda_1\sum_{k=2}^{n}(v_k-v_{k-1})^2 \quad (2)$$

の最小化問題になる．これはノイズが含まれたデータ d_k に平滑化を施す問題と考えてよいが，有限次元の2次形式最小化問題なので v_k に関



$g_0=1/R_0$, $g_1=1/R_1$, $g_2=1/R_2$

図4　2次正則化問題と抵抗回路網　第1近傍への正抵抗結合 R_1，第2近傍への負性抵抗結合 R_2 のネットワークで2次正則化問題が解ける．

して微分し零とおけばよい．

$$\frac{1}{2}\frac{\partial G_1}{\partial v_k}=v_k-d_k-\lambda_1(v_{k-1}+v_{k+1}-2v_k)=0 \tag{3}$$

また $p=2$ で $\lambda_1=0$，$\lambda_2>0$ のときは，

$$G_2(\boldsymbol{v},\boldsymbol{d})=\sum_{k=1}^{n}(v_k-d_k)^2+\lambda_2\sum_{k=2}^{n-2}(v_{k+1}-2v_k+v_{k-1})^2$$

$$\frac{1}{2}\frac{\partial G_2}{\partial v_k}=v_k-d_k-\lambda_2(-v_{k-2}-v_{k+2}+4v_{k-1}+4v_{k+1}-6v_k)=0 \tag{4}$$

を得る．

（b）　線形並列抵抗回路網

データ d_k が与えられたとき，上の式（3）や式（4）を満たす v_k を何らかの手順で得ることができれば正規化問題の解になっていることに注意し，図3の並列回路の各ノードでキルヒホッフ電流則を書き下すと

$$(c_0+2c_1)\frac{dv_k}{dt}-c_1\left(\frac{dv_{k+1}}{dt}+\frac{dv_{k-1}}{dt}\right)=-g_0v_k+u_k+g_1(v_{k-1}+v_{k+1}-2v_k)$$

を得る．g_0，g_1，c_0，$c_1>0$ ならば明らかにこの回路は安定であり，平衡点

$$-g_0v_k+u_k+g_1(v_{k-1}+v_{k+1}-2v_k)=0 \tag{5}$$

に収束する．式（5）は $\lambda_1=g_1/g_0$，$d_k=u_k/g_0$ としたときの式（3）にほかならない．Mead のいわゆる“Silicon Retina”はこのような回路網である[1],[9]．寄生容量 c_0 は数百 fF のオーダであり，$1/g_0$，$1/g_1$ を数 kΩ のオーダとすると抵抗回路網が安定平衡点に整定する（すなわち1次正則化問題を解く）までの時間は数 μs のオーダである．これをディジタルコンピュータで解く場合は，時間 t を離散化して全ノードに対して収束するまで計算するので，膨大な計算量になることがわかる．

次に $p=2$ の場合を考える．図4の回路の（安定平衡点の）のキルヒホッフ電流則を書き下すと $-R_2=4R_1$ の場合，式（4）と同じになる．従って図4の抵抗回路網で2次正則化問題が解ける．ここで注意することは図4では図3に比べ二つ隣りのノードに対する結合が必要で，しかもその抵抗値 R_2 は負である[10]．

（c）　非線形並列抵抗回路網

上に述べた問題は2次形式の最小化問題であって正確には**標準**正則化問題とよばれている．いま，式（2）の代りに

$$G(\boldsymbol{v},\boldsymbol{l},\boldsymbol{d})=\sum_{k=1}^{n}(v_k-d_k)^2+\lambda_1\sum_{k=2}^{n}(v_k-v_{k-1})^2(1-l_k)+\lambda_1'\sum_{k=2}^{n}l_k \tag{6}$$

を考える．$\boldsymbol{l}=(l_1,\cdots,l_n)$ で，l_k は0または1しか値をとらず，line variable とよばれる．$l_k=0$ とすると式（6）は式（2）と同じで，データ d_k に平滑化を行う．一方 $l_k=1$ とすると第2項は零となり第3項が現れる．これは与えられたデータ d_k を平滑化した電圧分布 v_k に対して，ノード間の差 $(v_k-v_{k-1})^2$ が大きくないときはそのまま平滑化を行うが，$(v_k-v_{k-1})^2$ が大きいときは $l_k=1$ とし，$(v_k-v_{k-1})^2$ の項を零にするかわりに $\lambda_1'l_k$ のペナルティを受けることを意味する．与えられた画像に**線形演算**を行

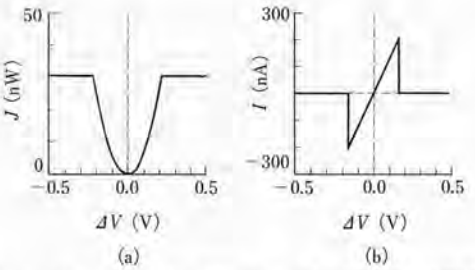


図5 （a）非標準正則化問題のペナルティ関数 $f(\cdot)$ の特性
式（7）の非線形項 $f(\cdot)$ の特性を表している．
（b） 抵抗ヒューズ（関数 $g(\cdot)$）の特性　両端の電位差が小さいときは電位差に比例した電流が流れるが，大きいときは電流が流れなくなる．©1991 Harris.

う限り，ノイズが除去されるかわり，エッジは必ずボケてしまいエッジの正確な場所も検出できない．式（6）は入力データに平滑化を施しながら，内在するエッジを検出する強力なアルゴリズムである．ところで式（6）は $\boldsymbol{v}$ と $\boldsymbol{d}$ だけでなく $\boldsymbol{l}$ にも依存し，そのままでは並列抵抗回路にならないので，第2，第3項の l_k について最小値をとると，計算は省略するが

$$G(\boldsymbol{v}, \boldsymbol{d}) = \sum_{k=1}^{n} (v_k - d_k)^2 + \sum_{k=2}^{n} f(\lambda_1, \lambda_1'; v_k - v_{k-1}) \tag{7}$$

となる．$f(\lambda_1, \lambda_1', v)$ は図5（a）に示すグラフをもつ．微分を v_k についてとり零とおくと，

$$v_k - d_k + g(\lambda_1, \lambda_1'; v_k - v_{k-1}) + g(\lambda_1, \lambda_1'; v_k - v_{k+1}) = 0$$

を得る．$g(\lambda_1, \lambda_1', v)$ は図5（b）で与えられる．従ってこの $g(\cdot)$ のような特性を持った非線形素子（**抵抗ヒューズ**）を作れば並列抵抗回路で式（6）の最小化問題を解ける．MIT の Harris はこれを実現し，印象的な実験を行った[5]−[8]（後述）．但し，図5（a）の $f(\cdot)$ は微分可能でない点があるので，理論的正当化には $f(\cdot)$ を1パラメータの関数族で摂動する必要がある．

2.5 抵抗回路網の時間空間安定性

多くのビジョンチップに内蔵される抵抗回路網は一般に，（i） m 個隣りのノードまで抵抗結合，寄生容量をもち，（ii）抵抗値は負の値をとりえる．このような抵抗回路網が不安定に

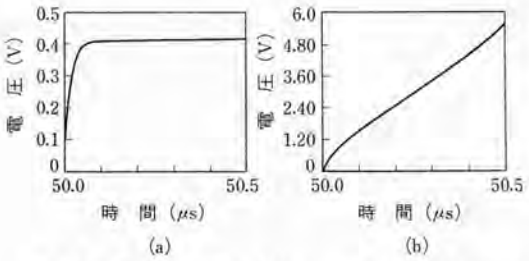


図6 抵抗回路網の時間安定性　図4のネットワークのノード31にステップ入力を与えたときのノード31の時間応答．ネットワークは負性抵抗，寄生容量を含むので時間不安定になりえる．
（a） $R_0 = 100\,\mathrm{k}\Omega$，$R_1 = 5\,\mathrm{k}\Omega$，$R_2 = -20\,\mathrm{k}\Omega$ の場合（時間安定）．
（b） $R_0 = 100\,\mathrm{k}\Omega$，$R_1 = 5\,\mathrm{k}\Omega$，$R_2 = -17\,\mathrm{k}\Omega$ の場合（時間不安定）．©1990 IEEE.

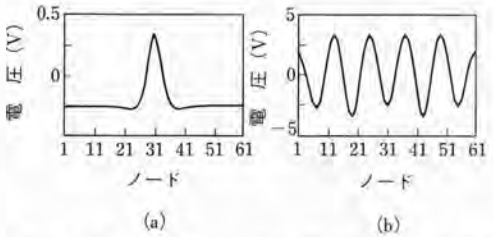


図7 抵抗回路網の空間安定性　図4のネットワークのノード31に一定入力を与えたときの平衡状態での各ノード応答（空間インパルス応答）．ネットワークは負性抵抗を含むので空間インパルス応答が不安定になりえる．
（a） $R_0 = 100\,\mathrm{k}\Omega$，$R_1 = 5\,\mathrm{k}\Omega$，$R_2 = -20\,\mathrm{k}\Omega$ の場合（空間安定）．
（b） $R_0 = 100\,\mathrm{k}\Omega$，$R_1 = 5\,\mathrm{k}\Omega$，$R_2 = -17\,\mathrm{k}\Omega$ の場合（空間不安定）．©1990 IEEE.

なり得ることは Poggio 等も指摘しており，次の二つの安定性が問題になる．

（a） **時間安定性**　例えば図4の抵抗回路網は負性抵抗と寄生容量を持つので，時間的に不安定になる可能性がある．図6（a）は負性抵抗値が大きい場合のあるノードの電圧の時間推移で，一定値に収束している（時間安定）．図6（b）は負性抵抗値が小さい場合で，時間と共に発散している（時間不安定）．

（b） **空間安定性**　図4の抵抗回路網は空間フィルタとして使われるが，このコンボリューション核（空間インパルス応答）は負性抵抗の値が小さくなると空間的に発振する．図7（a）は負性抵抗値が大きい場合の空間イン

パルス応答である（空間安定）．図7（b）は負性抵抗値が小さい場合で，空間インパルス応答が激しく振動しており，画像処理に使用するのに適していない（空間不安定）．

（c） **時間空間安定性の一致**　これら二つの安定性は別々の概念であり，先験的には両者は無関係であるが，予想に反し両者の条件は一致することが数値実験で，後に厳密な形で証明された[11],[12]．この負性抵抗を含む抵抗回路網の時間空間安定性一致の定理は，回路網理論における新しい結果となり得よう．

2.6 生理学的背景

脊椎動物の網膜は発生過程からみると脳の一部が突出してできたものであり，**脳そのもの**と考えてよい．この記事で紹介されているビジョンチップのいくつかは網膜神経回路の中のごく一部を再現するものであって，網膜の機能から見ればまだまだ貧弱なものである．これは網膜の多彩な視覚情報処理機能およびそれを実現する神経回路構造に，未知な部分が多いことと，ある程度解明されている部分があってもそれを理解し“カナモノ”にのせる力を持った研究者（群）が余りいないことによる．ここでは網膜に関する生理学的な知見についてスペースが許す範囲で大まかに解説してみたい．詳しい内容に興味のある読者のためには，網膜についての著書[13]がある．

網膜は眼底に位置する厚さ200～300ミクロンの神経組織である．その組織の中には驚くほど緻密に細胞が配列されている（図8（a））．ビジョンチップへの応用を考えたとき，脊椎動物の中でも魚類，両生類等の下等動物の網膜は興味ある対象である．その理由は，下等動物では高次中枢いわゆる脳が高等動物に比べ貧弱なため，視覚の基本的機能がむしろ網膜に存在するからである．実際カエルなどでは，餌となる虫を確認すると思われる細胞が網膜で見つかっている．また下等動物の網膜細胞は比較的大きく，生理学的な実験も容易であることから，ビジョンチップに結びつく重要なヒントが直接得られる．

網膜には大きく分けて5種類の細胞がある．図の下方が眼球の前面すなわちレンズ側であり，光はこの方向から入射する．脊椎動物の場合，光は網膜組織を透過し入射方向から見て一番奥にある視細胞（薄緑色）により吸収され，電気信号へと変換される．視細胞および後で述べる水平細胞，双極細胞は光に対し**アナログ信号**で応答する．驚くべきことに視細胞は**photon 1個**をとらえて反応することができる．視細胞には，桿体とよばれる細胞と錐体とよばれる細胞がある．細胞の上部が円柱型のものが桿体で光感度が高く，カメ網膜では1 photon当たり平均130 μV程度の電位変化により応答する．細胞の上部が円錐型のものが錐体で比較的光感度が低く同じく平均25 μV程度の電位応答をする．それぞれ暗がり，明るい場所で働き，合わせて5 log単位以上の光強度の範囲をカバーする広いダイナミックレンジを実現している．更に視細胞の応答特性は，網膜が明るい環境に連続的に置かれた場合，感度が下がる方向にシフトする．光強度変化に対し最も感度の高いレンジで光をとらえる，すなわち受容器レベルでの**順応**である．

細胞が網膜上のある領域に与えられた光刺激に反応したとき，その領域を受容野とよぶ．視細胞の受容野は，視細胞1個に比べ大きい．これは同じタイプの視細胞同士が電気的に結合し，視細胞間に側方のリークが存在するためである．リークにより空間解像度が落ち一見不都合に思われるが，各細胞に存在する内在的な雑音を平滑化するという意味で重要な構造である．錐体には，異なる光波長に対して最高感度を示す三つのサブタイプが存在し，これらは赤，緑および青錐体とよばれる．色覚における三原色仮説を生理学レベルで証明する重要な事実である．

視細胞は，二次ニューロンである**水平細胞**（青色）および**双極細胞**（赤色）と外網状層（図8（b）参照）とよばれる部位で，化学シナプスを介し連絡し合っている．水平細胞も視細胞と同様に近傍同士で電気的に結合している．この

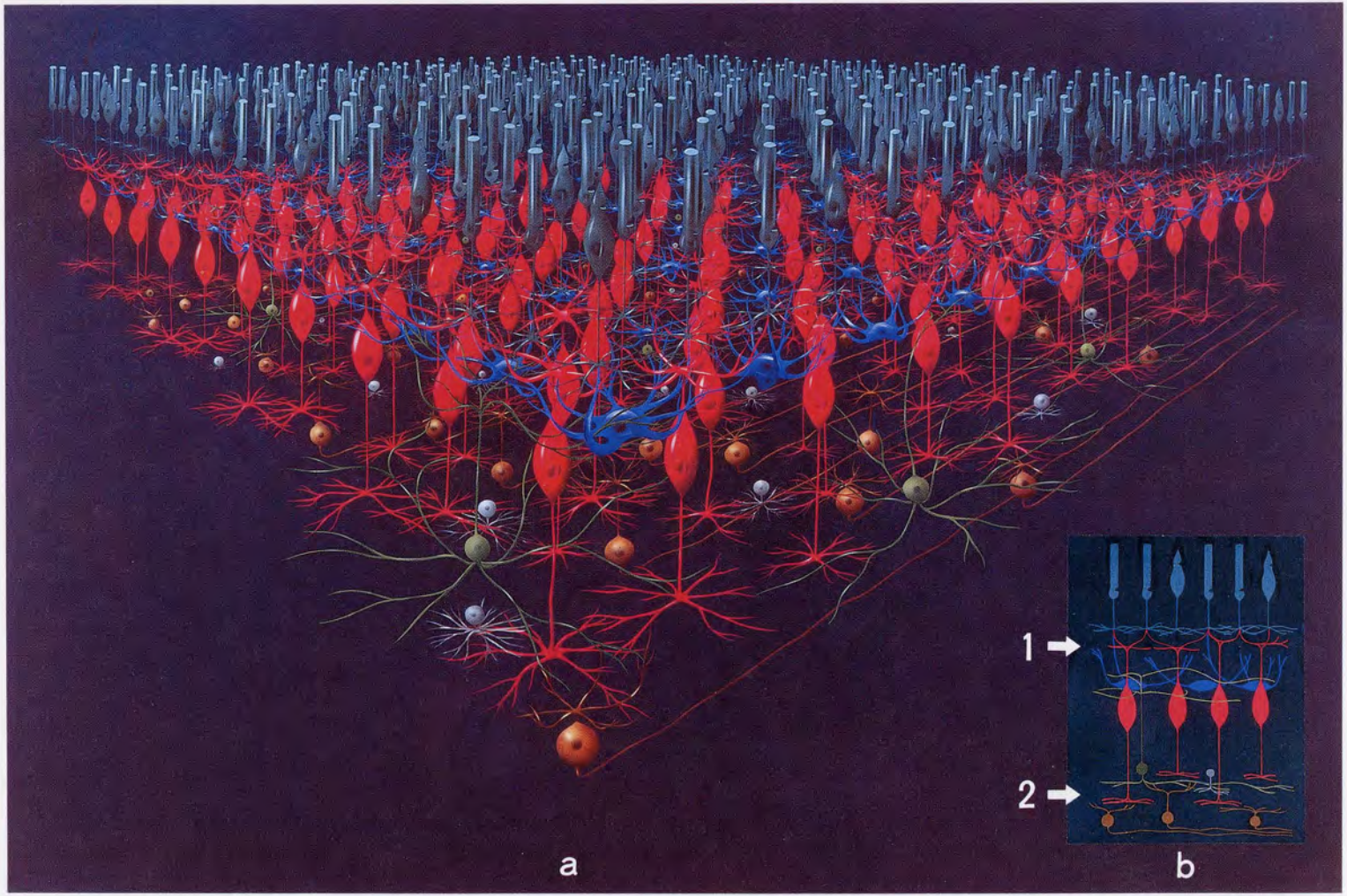


図8　脊椎動物網膜の基本構造　（a）網膜の全体図（詳細は本文）．（b）網膜の断面図．細胞の色分けは（a）と対応している．矢印1および2で示した層がそれぞれ外網状層および内網状層である．

電気的結合は視細胞に比べはるかに強く，水平細胞の受容野は網膜全体にも及ぶことがある．視細胞と水平細胞の成すネットワークは，二層の抵抗回路網により表現できる[14]．この構造は，初期視覚問題における標準正則化理論と深くかかわることが指摘され，ビジョンチップにも応用された（次号）．水平細胞には，桿体から入力を受けるもの，錐体から入力をうけるものがある．また異なる波長感度を持ったいくつかのサブタイプも存在し，色覚発現にも重要な役割を果たしていると考えられる．

双極細胞は**外網状層神経回路**の出力細胞である．双極細胞では，細胞の直上に照射された光に対する応答と，直上を除いてその周辺部に照射された光に対する応答の符号が反対となる．このような受容野は，**中心—周辺拮抗型**とよばれ，入力画像に対し平滑化とコントラスト強調を同時に実行するフィルタである．後に紹介する（5.7）いわゆる $\nabla^2 G$-like ビジョンチップはこの双極細胞の出力にヒントを得ている．双極細胞には，受容野に関しオン型とオフ型とよばれる二つのサブタイプが存在する．オン型は受容野の中心でプラス応答，周辺でマイナス応答をする．オフ型はこの反対である．なぜ二つのミラーイメージのようなサブタイプが存在するのかはよくわかっていない．双極細胞の出力は以下に述べる内網状層の細胞へと送られ，内網状層では，刺激の形状や速度などの特徴抽出が分化した経路（チャネル）で計算されている．従って双極細胞に至るまでの神経回路は，このように分化した視覚情報処理機能を実現するための共通の前処理フィルタと考えられ，ビジョンチップにおいても基本的な構成部となる．

双極細胞，**アマクリン細胞**（灰色）および**神経節細胞**（黄土色）は**内網状層**（図8（b）参照）において化学シナプスを介して複雑に結合し合う．内網状層の情報処理は生体の種により多少異なり，生体の棲む環境に対する機能の適応が見られる．アマクリン細胞は光刺激が与えられたとき，あるいは光刺激が切られたときに一過性に応答するという特徴がある．この応答特性は刺激光が細胞の受容野の入ったとき，あるいは受容野から出ていくときをとらえることができるので，視野の中で**動いている刺激**をとらえると考えられる．アマクリン細胞は，速度検出以外でも何らかの重要な役割を果たしている可能性もある．またこのアマクリン細胞と同じ層に位置する細胞で，**インタプレキシフォルム細胞**（IP細胞）とよばれる細胞がある．この細胞は内網状層で処理された信号を**外網状層にフィードバックする**細胞で，最近では網膜の中の6番目の細胞として分類されることもある．IP細胞は，網膜の**明るさ**に対する**順応レベル**に応じて水平細胞の**受容野の大きさを変化させる**という報告がある[15]．網膜が神経回路網のパラメータを変化させ，入出力特性をシステムレベルで環境に順応させる（これを神経順応とよぶ）のは興味深い．このような順応機能をビジョンチップで応用するためには，更に詳細な生理学的解析が必要であろう．

神経節細胞は網膜の出力細胞で，その神経繊維（軸索）は高次中枢（脳）へと投射される．神経節細胞の高度な情報処理を示す実験結果がカエルから得られている．カエルの神経節細胞の応答は，いくつかのサブタイプに分類されているが，中でも明るいバックグラウンドに小さな動く暗い部分があると良く反応する細胞は，“虫検出ニューロン”とよばれている．残念ながら，内網状層の神経回路構造はまだよくわかっていない．今後のより詳しい研究成果が期待される．

（次号につづく）

文　献

（1） Mead C.：“Analog VLSI and Neural Systems”, Addison-Wesley, Reading, MA (1989).

（2） Marr D.：“Vision”, W.H. Freeman, San Francisco, CA (1982).

（3） Poggio T., Torre V. and Koch C.：“Computational Vision and Regularization Theory”, Nature, 317, pp.314-319 (Sept. 1985).

（4） Poggio T., Voorhees H. and Yulle A.：“A Regularized Solution to Edge Detection”, AI Memo, MIT, Chambridge, MA (May 1985).

（5） Harris J.：“An Analog VLSI Chip for Thin-Plate Surface Interpolation”, IEEE Conf. Neural Info.

Proc. Systems-Natural and Synthetic (1988).
(6) Harris J., Koch C., Luo J. and Wyatt J. JR.："Resistive Fuses：Analog Hardware for Detecting Discontinuities in Early Vision", Analog VLSI Implementation of Neural Systems, Mahowald M. and Mead C. ed., Kluwer Academic (1989).
(7) Harris J.："Analog Models for Early Vision", PhD Thesis, California Institute of Technology (1991).
(8) Liu S.C. and Harris J.："Generalized Smoothing Networks in Early Vision", Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp.184-191 (1989).
(9) Mead C. and Mahowald M.："A Silicon Model of Early Visual Processing", Neural Networks, 1, 1, pp.91-97 (1988).
(10) Kobayashi H., White J.L. and Abidi A.A.："An Active Resistor Network for Gaussian Filtering of Images", IEEE Journal of Solid-State Circuits, 26, 5, pp.738-748 (May 1991).
(11) Matsumoto T., Kobayashi H. and Togawa Y.："Spatial Versus Temporal Stability Issues in Image Processing Neuro Chips", IEEE Trans. on Neural Networks, 3, 4, pp.540-569 (July 1992).
(12) White J.L. and Willson A.N.："On the Equivalence of Spatial and Temporal Stability for Translation Invariant Linear Resistive Networks", IEEE Trans. on Circuits and Systems, 39, pp.734-743 (Sept. 1992).
(13) Dowling J.E.："The Retina：An Approachable Part of the Brain", Belknap Press of Harvard University Press, Cambridge, Massachusetts (1987).
(14) Yagi T., Ariki F. and Funahashi Y.："Dynamic Model of Dual Layer Neural Network for Vertebrate Retina", Proc. IJCNN, Washington, 1, pp.787-789 (1989).
(15) Shigematsu, T. and Yamada, M. (1988). Effects of Dopamine on Spatial Properties of Horizontal Cells in the Carp Retina. *Neuroscience Reveach, Supplement* 8,s69-s80.

まつもと　たかし
松本　隆　(正員)
昭41早大・理工・電気卒．昭44ハーバード大大学院・応用数学修士．昭47工博（早大）．昭52～54カリフォルニア大バークレー・電気工学-計算機科学研究員．非線形回路の分岐とカオス，ニューラルネットワークの研究に従事．現在，早大・理工・電気工学科教授．平2～3非線形問題研究専門委員会委員長．Proceedings of IEEE 編集委員．Circuits, Systems and Signal Processing 編集委員．

こばやし　はるお
小林　春夫　(正員)
昭55東大・工・計数卒．昭57同大学院修士課程了．同年横河電機（株）入社．以来，計測器，ミニスーパコンピュータの研究開発に従事．昭62から平元UCLA・電気・修士課程留学．アナログCMOS IC設計，ニューラルネットワークに関心を持つ．

やぎ　てつや
八木　哲也　(正員)
昭54名大・理・物理卒．昭60同大学院医学研究科了．学術振興会特別研究員（生理学研究所）・名工大助手を経て，平2九工大情報工学部助教授．生体の視覚情報処理についての研究に従事．医博．IEEE，日本生理学会，神経回路学会，日本宇宙航空環境医学会各会員．

TECHNICAL SURVEY

ビジョンチップ(II・完)
——アナログ画像処理用ニューロチップ——

解 説

松本 隆 小林春夫 八木哲也

松本 隆：正員 早稲田大学理工学部電気工学科
小林春夫：正員 横河電機株式会社エレクトロニクス研究所
八木哲也：正員 九州工業大学情報工学部制御システム工学科

Vision Chip [II・finish]: Analog Image-Processing Neuro Chip. By Takashi MATSUMOTO, Member (School of Science and Engineering, Waseda University, Tokyo, 169 Japan), Haruo KOBAYASHI, Member (Electronics Laboratory, Yokogawa Electric Corp., Musashino-shi, 180 Japan) and Tetsuya YAGI, Member (Faculty of Computer Science and Systems Engineering, Kyusyu Institute of Technology, Iizuka-shi, 820 Japan).

3. ビジョンチップの実現

(a) アナログCMOS VLSI技術

ディスクリート部品によるアナログ画像処理装置には先駆的仕事がいくつかある．例えばNHKの安田，山口，福島，長田は20年以上も前にMeadの回路網と同様な抵抗回路網を実現しているし[1]，東京大学の石川，吉澤はモーメント計算用並列抵抗回路網を実現している[2]-[4]．現在はLSI技術の進歩によりビジョンチップは1チップ化されるようになった．それらのほとんどがアナログCMOS回路で構成されている．

(b) 画像センサ

ビジョンチップでは画像入力のための光センサとして，(i) CCD，(ii) ホトトランジスタ，(iii) ホトダイオード，が使われている．ホトトランジスタによる方法はMeadによって提案された方法で，標準CMOSプロセスの寄生バイポーラトランジスタを利用し，80 bB程度の入力ダイナミックレンジが得られる[5]．

(c) 抵抗回路網

標準CMOSプロセスで，抵抗回路網の抵抗は，(i) MOS抵抗，(ii) ポリシリコン抵抗，(iii) 拡散抵抗，(iv) スイッチドキャパシタ，が使われている．MOS抵抗は図9に示すように，FETを二つ縦列または並列に並べて線形性を良くし，バイアス電圧を制御することで抵抗値を調整できる．

(d) サブスレッショルド領域CMOS回路

MOS FETでゲート・ソース間電圧がスレッショルド電圧より低い場合は，カットオフ（サブスレッショルド）領域として通常の回路ではドレーン・ソース間電流は流れないとして扱うが，実際はこの領域でも微小の電流が流れる．このサブスレッショルド領域のCMOS回路は，時計や心臓ペースメーカ等極めて低消費電力を要求される分野で使われてきた[6]．この領域のCMOS回路では，低消費電力は実現できるが精度は悪化する．が，生物の脳も低消費電力・低精度でも高度の情報処理を行っているというのがMeadの主張であり，彼のグループのチップ

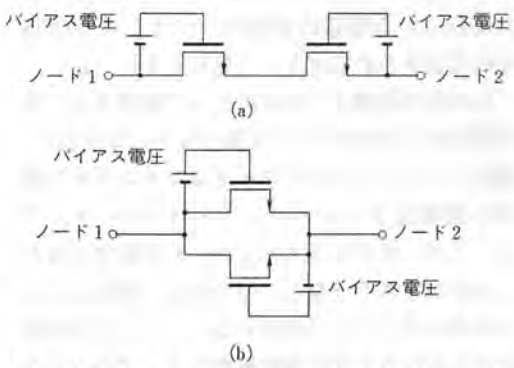


図9 MOS抵抗 (a) 縦続に二つのMOSを結合 (b) 並列に二つのMOSを結合　両方の回路とも二つのMOSを用いて非線形項をキャンセルし線形な抵抗を実現している．バイアス電圧により抵抗値を制御できる．

は例外なくサブスレッショルド領域で動いている．

（e） シンプルな回路と複雑な配線

ビジョンチップは，入出力回路・抵抗等から成るシンプルなセル回路（画素に対応）が縦横に並んだ，メモリICに似た構成になっている．しかし，電源配線，データ読出用配線のほかに，抵抗回路網によるセル間配線が必要である．例えば文献（7），（8）のような二つ隣りまでのセル間の抵抗結合が必要な場合は，配線領域がチップ面積の40％程度にもなる．ビジョンチップのLSI化の際には配線の複雑さが問題になる．なお配線の複雑さを定量化する試みが（9）でなされている．

（f） シリコンファンダリ

米国にはMOSIS（MOS Implementation System）とよばれる組織があり，大学や研究所，企業の研究者，学生はICのマスタパターンまで設計すれば，そのデータを電子メールで送ることにより，この機関を通じ低価格（例えば2.2 mm×2.2 mmチップで550ドル程度）でICを試作できる．学生が自分のアイディアをアルゴリズム化し，それを実現するICを設計し，ICのマスクパターンを送った後8週間でチップを手に入れ，それをテストし，まとめて卒業論文とする例はいくらでもある．もちろんアナログ回路だけでなくディジタル回路も可能である．米国の半導体関連技術のいくつかが圧倒的に優れている間接的要因の一つとしてMOSISの存在はきわめて大きいと思われる．

MOSISの創設もMeadによって提唱され，米国国防省（DARPA）の支援のもとになされた．面白いことに，そのオフィスはロサンゼルス郊外の閑静なヨット・ハーバー（マリーナ・デル・レイ）のすぐ近くという，LSI研究とはおよそ縁がなさそうなところである．実際ここには普通のオフィス以外何もない．ここで各研究者から送られてきた複数種類のチップのマスクパターンデータを1枚のウェーハにアセンブリし，半導体メーカにファブリケーションを依頼する．このマルチチップ技術により各IC当りの低試作価格を実現している．欧州にもEUROCHIPとよばれる同様の組織がある．

海外の研究者が必ずと言ってよいほど放つ質問“半導体王国日本にMOSISに対応するものが何故ないのか？”の返答に窮し，くやしさをかみしめている研究者はたくさんいる．日本がDRAMの勝利だけではやっていけない時代は既に到来している．背筋が寒い思いをしている研究者は筆者等だけではないであろう．

4. 各研究機関での研究活動

4.1 大　学

ビジョンチップの研究開発は主に米国の大学，企業によってなされている．

（a） カルフォルニア工科大学（Caltech）

Caltechはロサンゼルスのダウンタウンから北に車で約30分のパサデナとよばれる地区にある．すぐ近くにNASAのJPL（Jet Propulsion Laboratory）も位置し，ニューロチッププロジェクトを含め，Caltechと共同研究を行っている．ここのMeadはディジタルLSI設計法（Mead-Conway法），シリコンコンパイラの発明者としても知られ，システム，回路，デバイスにわたって精通している．ニューロチップは自分が今まで経験した最も興味深いものであってLSIの新しいパラダイムになると位置付けている．ここで視覚の生理学，アルゴリズム，そしてそれを実現するための回路技術の研究がなされ，さまざまなビジョンチップが開発されている．例えば動き検出，移動物体追跡，盲人用視覚センサ等である．

（b） マサチューセッツ工科大学（MIT）

マサチューセッツ州ケンブリッジのMIT電気工学科Wyatt（回路網理論），Lee，Sodini（アナログMOS回路），人工知能研究所Poggio，Horn（コンピュータビジョン）らのもとで1988年から5か年計画でビジョンチッププロジェクトが始められている[10]．このプロジェクトはCaltechとも連携している（CaltechのKochがliaisonを務めている）．このプロジェクトでは，ビジョンチップの回路の安定性，物

体の位置方向検出チップ，抵抗ヒューズ，スイッチドキャパシタによる抵抗ネットワークの実現，CCDによるアナログ画像メモリの実現，ステレオ視，動き検出チップ等の研究が行われている．

またMIT Lincoln研究所ではこのグループとは別にChiangがCCDをベースにしたビジョンチップを開発している[11]～[14]．

（c） カーネギメロン大学（CMU）

米国北東部ピッツバーグのCMUの電気・コンピュータ工学科のGruss，Carley，Kanadeはレンジファインデング用ビジョンチップを開発した[16]．このチップは物体の3次元プロファイルを従来のシステムに比べ2けた以上高速に精度良く求めることができ，リアルタイムロボットビジョンに用いることができる．またCarleyのもとで，フローティングゲートMOSでアナログメモリを実現する研究が行われた[16]．これはアナログニューロチップの実現に利用できる．

（d） カルフォルニア大学ロサンゼルス校（UCLA）

ロサンゼルスのウエストウッドに位置するUCLAでも電気工学科Abidiのもとでビジョンチップの開発が行われた[17],[18]．Abidiは高速アナログIC設計が専門分野であるが，そのアナログ回路技術のバックグランドを生かし，画像平滑化用ビジョンチップを開発した．これは2次正則化問題を解くチップである．このチップは負性抵抗を含むので回路の安定性が調べられ，2.5に示したようなネットワークの時間空間安定性の一致が早稲田大学で一般的に証明され[17]，UCLAでも別の手法で証明された[18]．また，最近，早稲田大学の松本らはAbidiと層構造をもつ∇^2G-likeなSCEとよばれるビジョンチップを共同開発した[19],[20]．

4.2 インダストリ

企業の中にもビジョンチップを開発し製品化しようとする動きが始まった．

（a） ベンチャー企業

Mead，Faggin（マイクロプロセッサ4004発明者の1人）らによって創設されたニューロベンチャー企業Synapticsではビジョンチップを用いた小切手カスタマーナンバーの超小型認識システムを開発した．また，Tanner Inc.ではCaltechで開発された動き検出チップを用いてオプティカルマウスを開発している[25]．

（b） 大企業

南カルフォルニア地区にある閑静な町サウザンド・オークスにはRockwell International社サイエンス・センターがあり，GaAsの研究などで知られている．ここのMathurが率いるグループは地元のCaltech，UCLAのビジョンチップのプロジェクトを積極的に支援し，また自社でも開発を進め，ロボット等のリアルタイムの物体認識に応用しようとしている[31]．同地区にあるHughes Aircraft社でも抵抗ヒューズを用いたビジョンチップを開発している．

5. 開発されたビジョンチップの紹介

5.1 シリコン網膜チップ

このチップは多くの論文や記事に紹介されているので詳細は省略するが，前回の図3の回路網で入力と各ノードの電圧の差を出力しコントラスト強調を行うものである．但し平滑化は行っていない．当初開発された網膜チップは時間的に不安定で正常に動作しなかったが，MITのグループの理論解析により安定に動作するための回路パラメータの十分条件が得られ[28]，現在はさまざまなビジョンチップが安定に動作している．

5.2 2次正則化問題チップ

UCLAのKobayashi，White，Abidiは2次正則化問題を解くチップを開発した[7],[18]．これは図4のように二つ隣りのノードに対しても負性抵抗の結合を持ち，その空間インパルス応答はGaussian分布に近い．図3の1次正則化問題を解くネットワークの空間インパルス応答は中心が尖ってしまうので，画像平滑化フィルタとしては2次正則化の方が優れていることが知られている．負性抵抗は図10のように電流インバータを用いたCMOS回路で実現している．

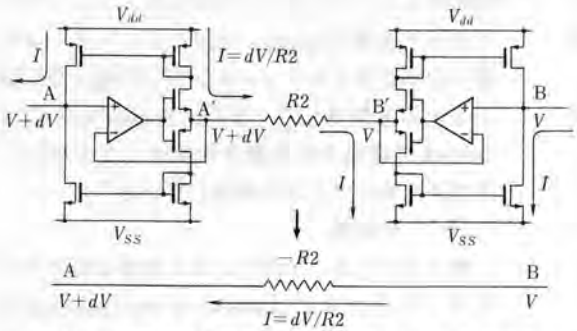


図10 負性抵抗の実現回路　電流の流れを反転させるCMOS回路により負性抵抗が実現できる．©1990 IEEE.

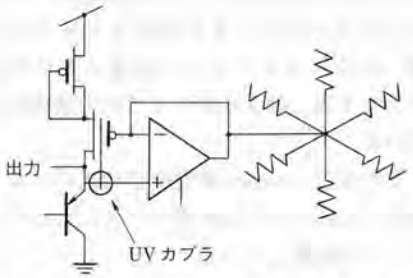


図11 Meadにより提案された適応網膜チップの回路　フローティングゲートを用い，その電荷量を紫外線で制御しスレッショルド電圧を調整することで精度を向上させる．©1990 IEEE.

また空間インパルス応答の幅を変えて平滑化の度合いを制御できるように，図4の抵抗$R_0=1/g_0$が可変抵抗になっている．この可変抵抗は図9（b）のように二つのFETを並列に並べ，そのゲート電圧を制御することで実現している．この回路は負性抵抗を含むが安定であることが示されている[17],[18]．Meadのチップと同様，このチップでも2次元ネットワークを六角形構造にしているが，これは四角形構造に比べ空間応答の対称性（Circular Symmetry）が良いためと，六角形構造は最も効率の良い2次元サンプリングであることが知られている[22]ためである．

5.3 動き検出チップ

動き検出は極めて難しい問題であって，生理学においてもそのメカニズムは完全には解明されておらず決定的なアルゴリズムはない．最大の困難の一つは，2次元画像入力の（空間／時間）微分をとらねばならない点にあると思われる．しかし，いくつかの重要な試みは行われている[21],[20]．時空間微分をとるチップ，$\nabla^2 G$フィルタをかけそのゼロ交差点の動きを検出するチップ等が開発されている．

5.4 適応網膜チップ

ビジョンチップではサブスレッショルド領域CMOS回路が多く用いられるが，この回路はオフセットが大きく精度があまり良くない．そこでMeadは図11に示すように回路の一部にフローティングゲートを用い，その電荷量を紫外線で制御しスレッショルド電圧を調整することでオフセットを減少させる方式を提案した．このようにした回路でより鮮明な画像を得ている．フローティングゲートを用いる方法はディジタルの分野ではEPROM，EEPROMに用いられている．但しこれらはやや特殊な（2層ポリシリコン）CMOSプロセスを必要とする．

5.5 抵抗ヒューズ

MITのHarrisは2.4（c）で述べたエッジ検出アルゴリズム（抵抗ヒューズ回路網）をチップとして実現し，エッジ検出の実験を行った(図12)[24]．ノイズを含んだ入力画像からきれいに

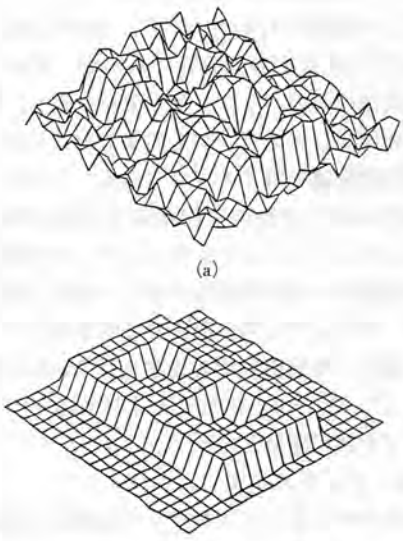


図12 抵抗ヒューズチップ（20×20画素）よるエッジ検出の測定結果（a）入力画像（b）出力画像　ノイズを含んだ入力画像からエッジ検出ができている．©1991 Harris.

エッジが検出されていることがわかる．Harrisは抵抗ヒューズをMOS FET 36個で実現したが，近年MITのYuらは少数のMOS FETで実現した[29]．図13（a）に示す構成は11個のEnhancement ModeのMOS FETで実現したものである．抵抗値，ヒューズの切れる電圧値が外部から制御できる．図13（b）に示す構成は4個のDepletion ModeのMOS FETによる実現である．Depletion ModeのNMOS（PMOS）はスレッショルド電圧が負（正）であり，通常のEnhancement ModeのNMOS（PMOS）に比べ極性が反転しているが，これを実現するためには，やや特殊なICプロセス工程を要する．

5.6 位置方向検出チップ

2次元画像のモーメントには多くの情報が含まれており，重心，オリエンテーション等を計算することができる．MITビジョンチッププロジェクトのStandleyはあるクラスの2次元並列抵抗回路網のキルヒホッフ電流則が（離散型）Greenの定理に対応している点に着目し，抵抗回路網の周辺電流（1次元）のみから0，1，2次モーメントの計算を行った（図14）[26],[27]．このモーメントを計算することで，物体の位置，方向を精度よく検出できる．

5.7 SCE（∇^2G-like）チップ

早稲田大学の松本らから提案された**2層回路網アーキテクチャ**（図15）は九州工大の八木による網膜神経回路の等価電気回路モデルからヒントを得たものである[32]．1層目で1次正則化問題が解かれ，2層目で2次正則化問題が解かれる．1層目から2層目のノード電圧を引くと画像のコントラスト強調だけでなく平滑化も

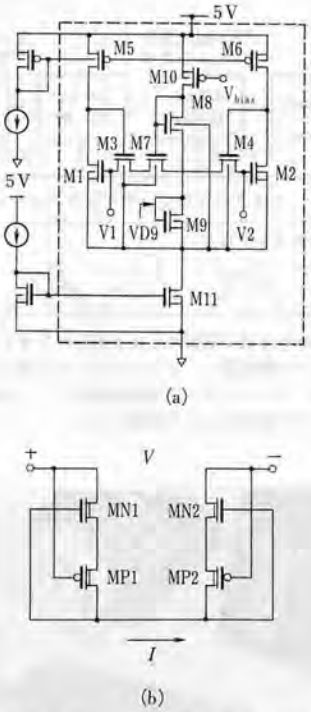


図13 MITでの抵抗ヒューズ実現（a）11トランジスタによる構成法（b）4トランジスタによる構成法　抵抗ヒューズのアナログCMOS回路での実現法がさまざま提案されている．(a)では抵抗値，ヒューズの切れる値がバイアス電圧で外部から制御できる．(b)ではDepletion Mode NMOSおよびPMOS（スレッショルド電圧が各々負および正）を用いている．©1990 IEEE.

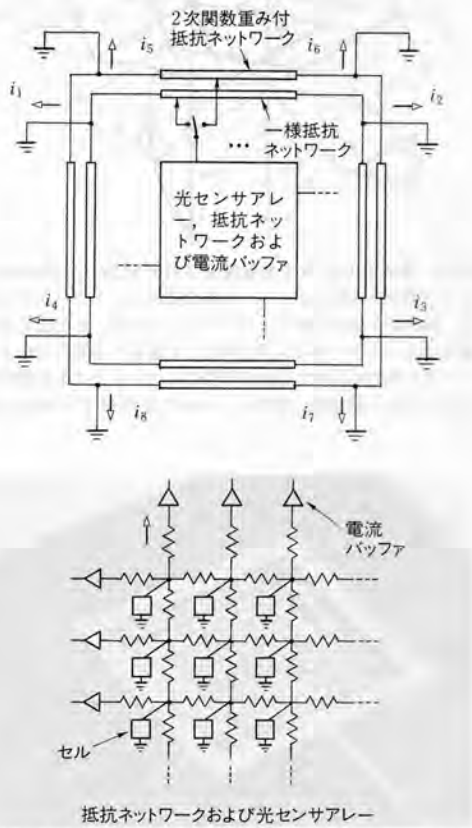


図14 MITで開発された位置・方向検出チップ　離散型Greenの定理により，抵抗ネットワークで物体の1，2次モーメントが得られ，位置・方向が計算される．©1990 IEEE.

同時に行われ，結果的にSCE（Smoothing-Contrast Enhancement）フィルタとなり，これは$\nabla^2 G$-likeな応答を示す．この構造は隣りのノードに対する正抵抗の結合だけのネットワークからなるので，**安定性**の問題は完全に解消され，**配線の複雑さ**は激減される[9]．図16にチップの写真，図17に実際に測定された入力画像(左)と出力画像（右）を示す[19],[20]．

5.8 小切手カスタマーナンバー認識システム

Synaptics社から発表されたこのシステムは，詳細は明らかにされていないのでどのような学習パラダイムが用いられているか不明であるが，既に数多くのテストを完了しており（1991 NIPS），確信を持っているとの印象を与えている．チップの半分は既に述べた網膜を模倣した

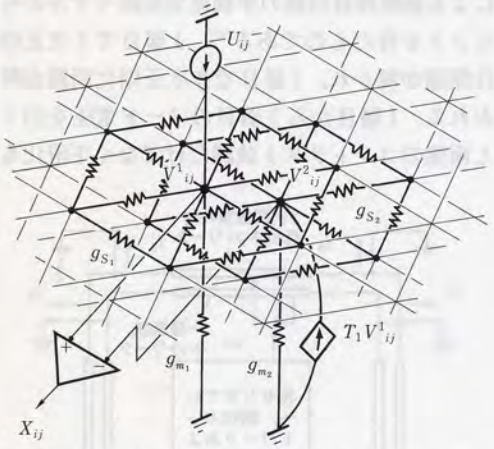


図15 早大とUCLAで共同開発されたSCE（$\nabla^2 G$-like）チップのアーキテクチャ　2層の抵抗ネットワークからなる．画像が1層目のネットワークに入力され，その出力が2層目のネットワークに入力される．1層目と2層目のネットワークの出力の差がとられて平滑化とコントラスト強調が同時になされ，結果的に空間インパルス応答が$\nabla^2 G$-likeとなる．

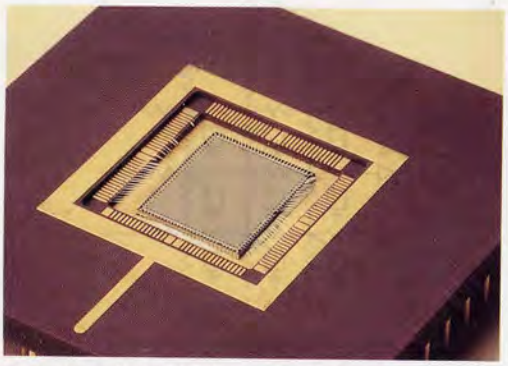


図16 SCE（$\nabla^2 G$-like）チップの写真　9.2 mm×7.8 mmのアナログCMOS VLSIに14万個のトランジスタが作りこまれ53×52画素が実現されている．

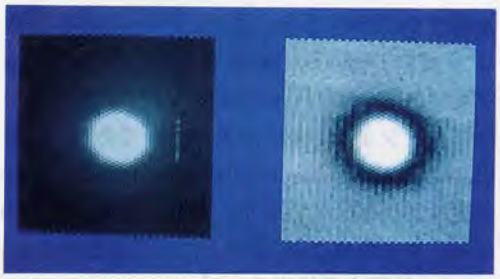


図17 SCE（$\nabla^2 G$-like）チップの測定結果（a）入力画像（b）出力画像　円形の入力画像（左）に対し，平滑化とコントラスト強調が同時にほどこされた出力画像（右）が得られている．

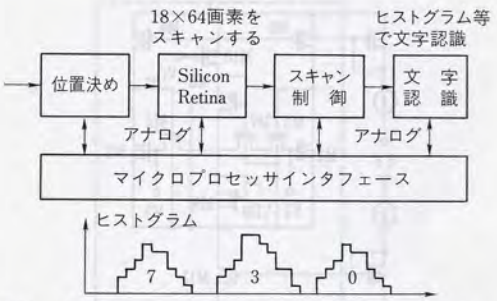


図18 Synaptics社で開発された小切手カスタマーナンバー認識システムの構成図　ビジョンチップを用いた最初の本格的な製品．ビジョンチップで前処理を行い，後段のアナログプロセッサで認識のための処理を行う．

図19 小切手カスタマーナンバー認識システムの写真
©Synaptics Inc.

入力部（18×64画素），残りの半分には学習機能が備わった認識部が組み込まれている（図18）．20,000画像／秒のスピードで画像を処理する．これはビジョンチップを用いた最初の本格的な製品といえよう（図19）．

6. ま と め

画像センサと画像処理機能を持つアナログCMOS VLSI，ビジョンチップが米国大学を中心に研究開発され，米国のベンチャー企業では本格的な製品を出すところも現れた．ビジョンチップはセンサがある程度の信号処理機能を有しているスマートセンサの一つと見ることもできる．従来のアナログICはA-D，D-A変換器のようにアナログとディジタルとのインタフェース部に多く使われており，信号処理部はディジタル回路が受け持つことが多かった．しかし，ビジョンチップの小規模ハードウェア，高速処理はアナログ信号処理への道を開くものと期待できる．また，従来のアナログICは小規模のものが多かったが，ビジョンチップはLSI技術の分野でアナログVLSIの領域を生みだし，さまざまな回路技術もそこで提案されている．国際固体回路会議（ISSCCやESSCIRC）でもここ数年ビジョンチップの論文が採択されている[8],[10],[12],[26],[20]．（なお，解説をするため文献を調べたが，100％網羅できたかどうかわからない．）ビジョンチップの開発過程で回路網理論上の新しい定理も導きだされており，また，生理学を参考にしたビジョンチップの新しいアーキテクチャが生まれている．網膜の情報処理，特に内網状層における情報処理機構を明らかにし，その知見に基づいたチップを実現することは高度にチャレンジングな，そして非常に面白い問題であろう．そのようなプロジェクトを遂行する研究者（群）は，当然生理学に精通していなければならず，また現在利用可能なテクノロジ（シリコン，GaAs等）の機能とその限界を知りつくしていなければならない．加えて現在手に入る“カナモノ”にのせ得るアルゴリズムを考え付く能力が要求される．このようなアナログVLSIはさまざまな分野とかかわりあいを持ちながら，今後一層普及されることが期待される．

文 献

(1) 安田稔，山口幸也，福島邦彦，長田昌次郎：“視覚系受容野の電子回路モデル”，信学論，54-C，6，pp.514-521（1971-06）．
(2) 石川正俊：“マトリクス状センサからの出力分布の中心の位置と総和の検出方法”，計測自動制御学会論文集，19-5，pp.381-386（1983）．
(3) 石川正俊，吉澤修治：“多層型並列処理回路を用いたn次モーメントの検出方法”，計測自動制御学会論文集，25-8，pp.904-906（1989）．
(4) 石川正俊：“並列処理を用いた能動的センサシステム”，計測自動制御学会論文集，24-8，pp.860-866（1988）．
(5) Mead C.：“Analog VLSI and Neural Systems”，Addison-Wesley Reading，MA（1989）．
(6) Vittoz E.A.：“Micropower Techniques”，Design of MOS VLSI Circuits for Telecommunications，Tsividis Y. and Antognetti P. ed.，Prentice-Hall，Englewood Cliffs，NJ（1985）．
(7) Kobayashi H.，White J.L. and Abidi A.A.：“An Active Resistor Network for Gaussian Filtering of Images”，IEEE J. Solid-State Circuits，26，5，pp.738-748（May 1991）．
(8) Kobayashi H.，White J.L. and Abidi A.A.：“An Analog CMOS Network for Gaussian Convolution with Embedded Image Sensing”，ISSCC Digest of Technical Papers，pp.216-217，San Francisco（Feb. 1990）．
(9) Kobayashi H.，Matsumoto T.，Yagi T.，and Shimmi T.：“Image Processing Regularization Filters on Layered Architecture”，Neural Networks（in press）．
(10) Wyatt J.L.Jr.，Standley D.L. and Yang W.：“The MIT Vision Chip Project：Analog VLSI Systems for Fast Image Acquisition and Ealy Vision Processing”，Proc. 1991 IEEE Int. Conf. Robotics Automation，pp.1330-1335（April 1991）．
(11) Chiang A.M. and LaFranchise J.R.：“A Programmable Image Processor”，ISSCC Digest of Technical Papers，pp.214-215，San Francisco（Feb. 1991）．
(12) Yang W. and Chiang A.M.：“A Full Fill-Factor CCD Imager with Integrated Signal Processors”，ISSCC Digest of Technical Papers，pp.218-219，San Francisco（Feb. 1990）．
(13) Chiang A.M. and Chuang M.L.：“A CCD Programmable Image Processor and its Neural Network Applications”，IEEE J. Solid-State Circuits，26，12，pp.1891-1894（Dec. 1991）．
(14) Chiang A.M.：“A CCD Programmable Signal Processor”，IEEE J. Solid-State Circuits，25，6，pp.1510-1517（June 1990）．
(15) Gruss A.，Carley L.R. and Kanade T.：“Integrated Sensor and Range-Finding Analog Signal Processor”，IEEE J. Solid-State Circuits，26，3，pp.184-191（March 1991）．
(16) Carley L.R.：“Trimming Analog Circuits Using Floating-Gate Analog MOS Memory”，IEEE J.

Solid-State Circuits, 24, 6, pp.1569-1575 (Dec. 1989).
(17) Matsumoto T., Kobayashi H. and Togawa Y. : "Spatial Versus Temporal Stability Issues in Image Processing Neuro Chips", IEEE Trans. Neural Networks, 3, 4, pp.540-569 (July 1992).
(18) White J.L. and Willson A.N. : "On the Equivalence of Spatial and Temporal Stability for Translation Invariant Linear Resistive Networks", IEEE Trans. Circuits and Systems-Ⅰ, 39, 9, pp.734-743 (Sept. 1992).
(19) Matsumoto T., Shimmi T., Kobayashi H., Abidi A.A., Yagi T. and Sawaji T. : "A Second Order Regularization Vision Chip for Smoothing-Contrast Enhancement", *Proc. of IJCNN 92*, Beijing (Nov. 1992).
(20) Shimmi T., Kobayashi H., Yagi T., Sawaji T., Matsumoto T. and Abidi A.A. : "A Parallel Analog CMOS Signal Processor for Image Contrast Enhancement", *Proc. of European Solid-State Circuits Conference* (Sept. 1992).
(21) Bair W., Koch C., Moore A., Horiuchi T., Bishofberger B. and Lazzaro J. : "Computiong Motion Using Analog VLSI Vision Chips : An Experimental Comparision Among Four Approaches" (submitted).
(22) Dudgeon D. and Mersereau R. : "Multidimensional Signal Processiong", Prentice Hall, Englewood Cliffs, NJ (1984).
(23) Mead C. and Mahowald M. : "A Silicon Model of Early Visual Processing", Neural Networks, 1.1, 1, pp.91-97 (1988).
(24) Harris, J. : "Analog Models for Early Vision", PhD Thesis, California Institute of Technology (1991).
(25) Tanner J.E. : "Integrated Optical Motion Detection", PhD Thesis, California Institute of Technology (1986).
(26) Standley D.L. and Horn B.K. : "An Object Position and Orientation IC with Embedded Imager", ISSCC Digest of Technical Papers, pp.38-39, San Francisco (Feb. 1991).
(27) Standley D.L. : "An Object Position and Orientation IC with Embedded Imager", IEEE J. Solid State Circuits, 26, 12, pp.1853-1859 (Dec. 1991).
(28) Standley D.L. and Wyatt J.L.Jr. : "Stability Criterion for Lateral Inhibition and Related Networks that is Robust in the Presence of Integrated Circuit Parasitics", IEEE Trans. Circuits and Systems, 36, pp.675-681 (May 1989).
(29) Yu P.C., Decker S.J., Lee H.-S., Sodini C.G., Wyatt J.L.Jr. : "CMOS Resisitve Fuses for Image Smoothing and Segmentation", IEEE J. Solid-State Circuits, 27, 4, pp.545-553 (April 1992).
(30) Hutchinson J., Koch C., Luo J. and Mead C. : "Computing Motion Using Analog and Binary Resistive Network", IEEE Computer, 21, pp.52-63 (March 1988).
(31) Mathur B., Liu S.C. and Wang H.T. : "Analog Neural Networks for Focal-Plane Image Processing", SPIE 1242, pp.141-151 (1990).
(32) Yagi T., Ariki F. and Funahashi Y. : "Dynamic Model of Dual Layer Neural Network for Vertebrate retina", Proc. IJCNN, Washington, 1, pp.787-789 (1989).

Credits

図10 : Reprinted with pemission from Kobayashi H., White J.L. and Abidi A.A. : "An Active Resistor Network for Gaussian Filtering of Images", IEEE J. Solid-State Circuits, 26, 5, pp.738-748 (May 1991). ©1991 IEEE.
図11 : Reprinted with pemission from C. Mead "Neuromorphic Electronic Systems", Proc. of the IEEE, vol.78, no.10, pp.1629-1636 (Oct. 1990). ©1990 IEEE.
図12 : Reprinted with pemission from Harris, J. : "Analog Models for Early Vision", PhD Thesis, California Institute of Technology (1991). ©1991 Harris.
図13 : Reprinted with pemission from Yu P.C., Decker S.J., Lee H.-S., Sodini C.G., Wyatt J.L.Jr : "CMOS Resisitve Fuses for Image Smoothing and Segmentation", IEEE J. Solid-State Circuits, 27, 4, pp.545-553 (April 1992). © 1992 IEEE.
図14 : Reprinted with pemission from Standley D.L. : "An Object Position and Orientation IC with Embedded Imager", IEEE J. Solid-State Circuits, 26, 12, pp.1853-1859 (Dec. 1991). ©1991 IEEE.
図19 : Reprinted with pemission from Synaptics Inc. © Synaptics Inc.

松本 隆（まつもと たかし）（正員）
昭41早大・理工・電気卒. 昭44ハーバード大大学院・応用数学修士. 昭47工博（早大）. 昭52～54カリフォルニア大バークレー・電気工学計算機科学研究員. 非線形回路の分岐とカオス, ニューラルネットワークの研究に従事. 現在, 早大・理工・電気工学科教授. 平2～3非線形問題研究専門委員会委員長. Proceedings of IEEE 編集委員. Circuits, Systems and Signal Processing 編集委員.

小林 春夫（こばやし はるお）（正員）
昭55東大・工・計数卒. 昭57同大学院修士課程了. 同年横河電機（株）入社. 以来, 計測器, ミニスーパコンピュータの研究開発に従事. 昭62から平元UCLA・電気・修士課程留学. アナログCMO SIC設計, ニューラルネットワークに関心を持つ.

八木 哲也（やぎ てつや）（正員）
昭54名大・理・物理卒. 昭60同大学院医学研究科了. 学術振興会特別研究員（生理学研究所）・名工大助手を経て, 平2九工大情報工学部助教授. 生体の視覚情報処理についての研究に従事. 医博. IEEE, 日本生理学会, 神経回路学会, 日本宇宙航空環境医学会各会員.

An Active Resistor Network for Gaussian Filtering of Images

Haruo Kobayashi, Joseph L. White, *Student Member, IEEE*, and Asad A. Abidi, *Member, IEEE*

Abstract—**The architecture of an active resistive mesh containing both positive and negative resistors to implement a Gaussian convolution in two dimensions is described. With an embedded array of photoreceptors, this may be used for image detection and smoothing. The convolution width is continuously variable by 2:1 under user control. Analog circuits implement a 45×40 mesh on a 2-μm CMOS IC, and perform an entire convolution in 20 μs on applied images.**

I. Introduction

HARDWARE capable of sensing an input in two dimensions and processing it in parallel to obtain results in real time is of great interest in applications such as low-power compact image recognition systems. In digital signal processors today, a 2D input from a sensor is first scanned and quantized, and subsequently processed using pipelined parallel algorithms to obtain a fast throughput rate [1]. The data at each grid point in the 2D input, corresponding to one pixel in the case of a sampled image, serially enter this signal processor and flow through it at some usually fast clock rate. A substantial increase in throughput may be obtained over this signal flow rate by using *simultaneous* processing *per pixel*, particularly if the signal fan-out is eliminated by not digitizing the input but retaining it as an analog quantity. This is how signal processing takes place in natural biological systems [2]–[4].

Much of signal processing consists of data reduction and the extraction of high-level content for purposes such as identification, classification, or storage. The hardware to accomplish this will very often implement an algorithm derived from a study of physical or biological systems, which naturally perform a similar task. In a programmable digital signal processor, an explicit algorithm is entered as a sequence of instructions, or as their hardwired equivalent in a dedicated processor. Analog hardware, on the other hand, cannot be programmed as digital operations may be, and is almost always hardwired: a circuit must be constructed in which Kirchhoff's laws and the terminal characteristics of the components together embody the desired algorithm. Insofar as this synthesis is guided by experience, ingenuity, and taste, the approach is ad hoc and limited in its generality; but when successfully executed, it may offer a savings in power and enhancement in speed by orders of magnitude over the digital approach [5]. The input to an analog signal processor is some current or voltage, the output some other voltage or current determined by the laws of physics governing the circuit. The early analog computers were built on this principle, but being composed of building blocks with quite general functions, they were not very efficient in hardware for massively parallel tasks. Translinear integrated circuits are one well-known example of an efficient use of hardware to embody complex nonlinear algorithms, although usually for scalar or one-dimensional array inputs. They achieve hardware efficiency by exploiting transistor device physics rather than from complex building blocks such as operational amplifiers; they are also hardwired to accomplish a specific task [6], [7]. Our work deals with a class of circuits suited to simultaneous signal processing in two dimensions also using processing at the transistor level.

Manuscript received September 27, 1990; revised January 25, 1991. This work was supported by the Office of Naval Research under Contract N00014-89-J-1282, Rockwell International, TRW, and the State of California MICRO Program. This paper was first presented at the 1990 International Solid-State Circuits Conference (ISSCC).

H. Kobayashi was with the Integrated Circuits and Systems Laboratory, Electrical Engineering Department, University of California, Los Angeles, CA 90024-1594 on leave from Yokogawa Electric Corporation, Tokyo, Japan.

J. L. White and A. A. Abidi are with the Integrated Circuits and Systems Laboratory, Electrical Engineering Department, University of California, Los Angeles, CA 90024-1594.

IEEE Log Number 9143314.

II. Image Smoothing using Simultaneous 2D Signal Processing

This section will discuss the algorithm and architecture of a particular image processing function we have implemented for potential use in compact machine vision systems [8].

A. Smoothing Images by a Gaussian Operation

Many electronic image recognition systems tend to replicate the hierarchy from low- to high-level processing found in biological organisms. A raw image is usually smoothed to suppress noisy features; its outline is then obtained with some form of edge-enhancement operation, and the outline after normalization and rotation is compared with stored templates. While the quantity of data might reduce along this chain, the complexity of the operations increases significantly. Our work relates to the lowest level of image processing, the smoothing of raw

0018-9200/91/0500-0738$01.00 ©1991 IEEE

image data with a Gaussian convolution function of variable width.

There is broad evidence suggesting that a noisy image is best smoothed by a Gaussian convolution kernel prior to edge enhancement. This corresponds to the defocusing action of a lens, and is inherent in many biological systems. The defocusing blurs the small sharp features characteristic of visual noise, which are extraneous to important objects in the field of view. Unless the image is properly smoothed beforehand, differentiating the intensity map of the image to enhance the edges will also accentuate the sharp noisy features. Theoretical work has proven that a noisy image is best smoothed by a Gaussian convolution kernel to obtain the largest signal-to-noise ratio after differentiation [9], [10].

The optimal width, or extent, of the convolution used to smooth a particular image depends on the spatial standard deviation of the noise, and also on the scale of the objects which is usually not known in advance. The width of the Gaussian smoothing must therefore be variable under the control of the user. Adaptive methods such as scale space filtering [11] rely on this capability. Our experiments suggest that a Gaussian with a width variable by a factor of 2 is adequate to smooth the noise in many simple images sampled at a resolution of 50 by 50 pixels.

We set about after these considerations to implement one analog integrated circuit capable of sampling an image at a resolution of 50 pixels on a side, smoothing it by a Gaussian in about 5 μs, and giving the user the flexibility of continuously varying the Gaussian width by a factor of 2:1. This speed of operation is orders of magnitude faster than digital implementations of this convolution function, which in addition to the requirements of image buffering also require the image to be circulated several times through a filter to obtain the property of variable width.

B. Computation in 2D Using Resistive Meshes

Resistor networks were used as analog computers in the past to solve complex boundary value problems in electromagnetics [12]–[15]. These were later replaced by numerical simulation on digital computers, primarily because of the ease of programmability. Digital computation, however, could neither surpass the low power dissipation nor the speed of analog computers, because when the latter solve complex 2D problems, the currents and voltages could attain their final values within a very short *RC* relaxation time. This high speed is the main attraction of analog computation for 2D real-time signal processing, in that the number of calculations unlike digital computation does not grow proportionally to the resolution, but more as the square root. The use of this concept for similar applications has also been noted elsewhere [16].

Unlike a resistive sheet subject to a potential difference between two edges, where the resulting lateral equipotential contours solve electrostatic or magnetostatic field problems, the contours in a sheet which also has a continuous *leakage* to ground will decay in a characteristic fashion in response to a voltage applied at a single point. The spatial rate of decay depends on the leakage conductivity to ground relative to the lateral conductivity. This decay function may be thought of as the spatial impulse response of the leaky resistive sheet, or, equivalently, its convolution kernel; the potential contours in response to multiple-point stimuli will then be determined by linear superposition. Consider, for example, a one-dimensional discrete version of the leaky resistive sheet composed of a uniform linear mesh of resistors R_1 with resistors R_0 from every node to ground (Fig. 1). In response to a current excitation at one node, the resulting voltage distribution on the mesh decays n nodes away from the excitation according to an exponential function $\exp(-nR_1/R_0)$ [16]. This convolution kernel differs from a Gaussian in two important ways: it has a slower decay at its tails, and the exponentials on either side of the excitation meet at the center to produce a cusp (Fig. 1). The discontinuity in derivative at this point would produce undesirable results when this function is applied to a noisy image and then followed by edge enhancement. The mesh must therefore be modified to produce a characteristic function which better resembles the flat-topped Gaussian at the point of excitation. Obtaining a practical realization of this mesh was one of the key contributions of our work.

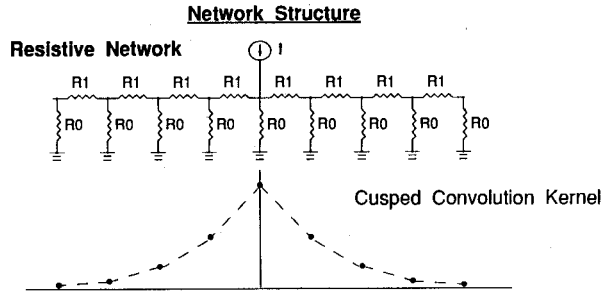


Fig. 1. 1D mesh with leakage resistors to ground, and its convolution kernel.

C. An Active Resistive Mesh Implementing Gaussian Convolution

We first qualitatively examine why the resistive mesh in the previous example produces a cusped convolution kernel, and how it must be modified. An indirect procedure for synthesizing the desired network is then described, followed by methods to extend it to two dimensions.

The spatial derivative of voltage at a point in a resistive sheet or discrete mesh specifies the potential gradient or the electric field there. According to the point form of Ohm's law, $J = \sigma E$, a current injected at a point (assuming the point has nonzero extent, so that the current density there is not infinite) on a resistive sheet with leakage to ground will produce some nonzero electric field (E) there, and therefore a nonzero potential gradient. A nonzero J may produce a zero E only if $\sigma \to \infty$,

which implies that the sheet must appear perfectly conductive at the point of injection. If a negative resistance is introduced to locally neutralize the dissipation in the sheet, while maintaining the dissipation across the large scale, a convolution function may be obtained with a flat top and decaying tails. It is plausible to achieve this in a discrete resistive mesh by introducing negative resistors not between every node, because that would simply modify the value of R_1, but between every other node, or perhaps even straddling several nodes. Investigating this numerically, we found that a mesh implementing a convolution of the desired shape could be obtained using negative resistors of a certain value connecting nodes with their *second nearest* neighbors. We also came upon an alternative procedure to synthesizing the same mesh, based on the theoretical work relating to the optimal smoothing of images. This is now described.

Poggio *et al.* [9] have analyzed how to smooth samples V_j, $-\infty < j < \infty$, of a noisy function to best estimate the derivative if the noise were not present. They seek a fitting function $U(x)$ with continuous first derivative which interpolates the sample points V_j with a least-mean-square difference, but with the constraint that the derivatives of $U(x)$ are not allowed to fluctuate excessively to obtain the least noisy estimate of the actual derivatives of the sampled function. This is expressed as the problem of minimizing an energy functional E, defined as the mean square difference between the interpolating function and the samples, subject to a penalty on excessively large second derivatives. The strength of the penalty is controlled by a parameter λ, called the regularization parameter:

$$E = \sum_j \left(U(x=j) - V_j\right)^2 + \lambda \int \left(\frac{d^2U}{dx^2}\right)^2 dx. \qquad (1)$$

It is shown that the $U(x)$ minimizing E in (1) is obtained by convolving V_j with an almost exactly Gaussian kernel, and the width of this kernel increases with λ. We may use this result by exploiting a fundamental connection between the minimum of an energy functional and the operating point of a circuit. It is known from circuit theory that Kirchhoff's laws and the constituent relations of the components drive a network to a state of minimum energy dissipation, so it is reasonable to construct a network whose energy dissipation is described by (1). The network equations may be obtained directly by setting the derivative of the right-hand side of (1) to zero.

Using a discrete estimate of the second derivative in (1), we get

$$E = \sum_j \left(U_j - V_j\right)^2 + \lambda \sum_j \left(U_{j+1} + U_{j-1} - 2U_j\right)^2 \qquad (2)$$

where $U_j = U(x=j)$. This is a quadratic form, and therefore has a unique minimum where $\partial E / \partial U_j = 0$ for all j, so

$$0 = 2\left(U_j - V_j\right) + \lambda \frac{\partial}{\partial U_j} \sum_i \left(U_{i+1} + U_{i-1} - 2U_i\right)^2 \qquad \text{for all } j. \qquad (3)$$

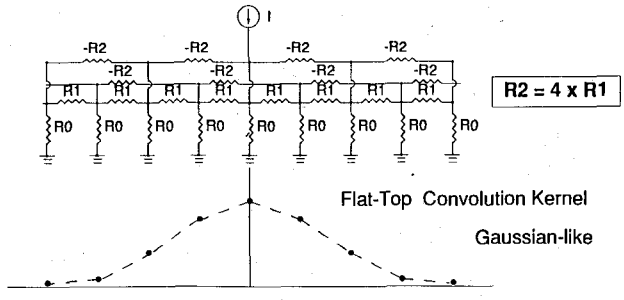


Fig. 2. 1D mesh with negative resistors between second nearest neighbors produces a convolution with a flat top.

Differentiating the terms in the sum and noting that $\partial U_i / \partial U_j = 0$ if $i \neq j$,

$$0 = \left(U_j - V_j\right) + \lambda\left(6U_j - 4\left(U_{j-1} + U_{j+1}\right) + \left(U_{j-2} + U_{j+2}\right)\right). \qquad (4)$$

This describes the node equations of a one-dimensional mesh [17] consisting of positive resistors (R_1) connecting nearest-neighbor nodes (i.e., $j-1, j$ and $j, j+1$), negative resistors ($-R_2 = -4R_1$) connecting second nearest neighbors, and resistors $R_0 = \lambda R_1$ to ground from every node, which are the leakage resistors described previously in the qualitative model (Fig. 2). The V_j correspond to voltage excitations in series with the leakage resistors. The network will produce as an array of node voltages (U_j) the convolution of the array of excitation voltages (V_j) with a Gaussian kernel whose width is controlled by λ. If $\{V_j\}$ were a set of photovoltages consisting of samples along a scan line through an image, the output set of voltages produced by the network would be the smoothed scan line.

The desired smoothing in an image, however, must take place across two dimensions. To obtain this, samples of a 2D image as a matrix of photovoltages should drive a *two-dimensional* mesh to obtain the desired result. The one-dimensional prototype of a Gaussian convolution mesh must then be extended to implement the kernel with circular symmetry in two dimensions. Noting, for instance, that a two-dimensional Gaussian function $G(x, y)$ is separable, that is, $G(x, y) = G(x) \cdot G(y)$, the desired 2D convolution may be obtained by driving an array of 1D meshes parallel to the y axis with the matrix of sampled photovoltages, and an identical array of 1D meshes along the x axis with the matrix of *buffered* outputs from the first array. This is not very efficient in hardware, because each mesh must have independent active circuits to produce the negative resistances, and an intermesh buffer must be used at every node.

Another possible implementation on a 2D rectangular grid is to connect every node to its *four* nearest neighbors oriented 90° apart with resistors R_1, and the *four* second nearest neighbors at the same orientations with resistors $-R_2$. The simulated spatial impulse response of this network decayed more rapidly along the diagonals than axially, producing an unacceptably large deviation from

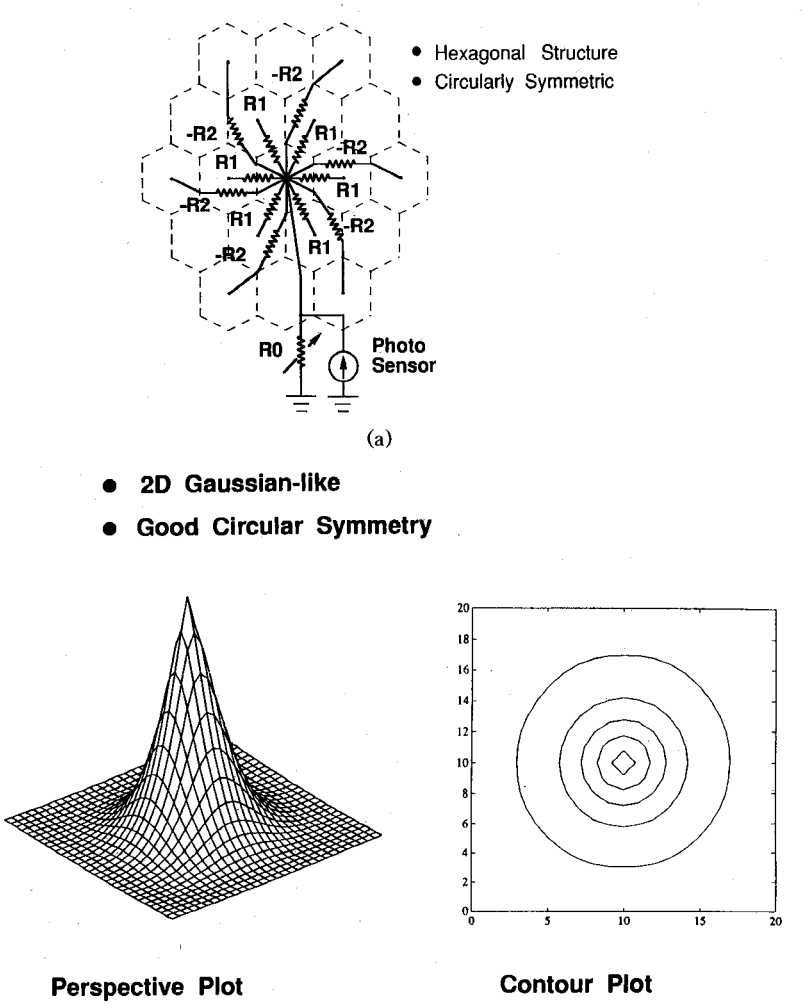


Fig. 3. (a) Extension of the mesh to 2D on a hexagonal grid produces (b) the best circular symmetry in the convolution kernel.

circular symmetry. A better circular symmetry was obtained by adding similar positive and negative resistive connections along the four diagonal directions, but weighted four times larger in magnitude. It became evident that a large number of components would be required to contrive circular symmetry on a rectangular grid, but not so on a hexagonal grid which inherently possesses a circular symmetry. The image must also be sampled on a hexagonal grid for compatibility with the mesh, which now consists of equal resistive connections 60° apart in orientation to nearest and second nearest neighbors. A hexagonal grid affords the greatest spatial sampling efficiency in the sense that the least photoreceptor sites will attain a desired coverage of the image [18], and the fewest network elements will yield the desired circular symmetry (Fig. 3(a)). The latter was verified in the simulated convolution kernel of this 2D network (Fig. 3(b)).

We required the kernel width to be variable by a factor of 2 under user control. That the convolution width depends on the ratio R_0/R_1 was known from the synthesis procedure, but the strength of this dependence was not. Simulations of the network showed a weak dependence (Fig. 4)

$$\text{Convolution width} \propto \left(\frac{R_0}{R_1}\right)^{1/4}. \tag{5}$$

It was simplest in terms of implementation to keep R_1 and R_2 fixed to preserve the Gaussian shape, and make R_0 alone variable by 16:1 to obtain the desired 2:1 variation in smoothing width.

Several aspects of this design procedure and simulated results invite analysis. Is there a systematic way to generalize a 1D mesh prototype with circular symmetry to 2D? Is the characteristic function of this combination of positive and negative resistors stable in space (i.e., does it decay rather than oscillate indefinitely)? Stable in time? Can the network be generalized to other convolution functions? What is the analytical relation between the width of the convolution function and the network elements? We have answered some of these questions elsewhere [19].

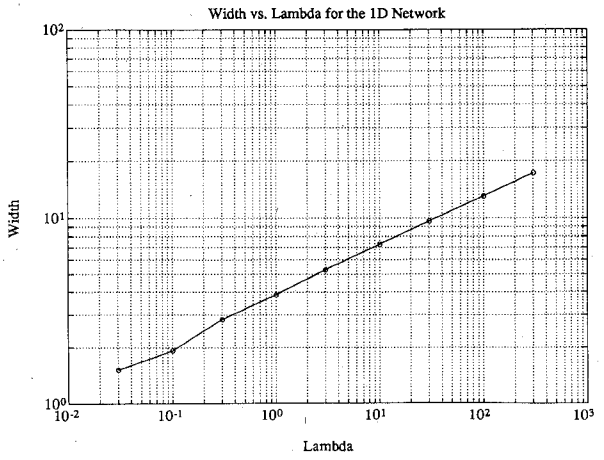


Fig. 4. The width of the convolution kernel increases as the 1/4th power of the grounded resistor.

III. Circuit Design

The practicality of implementing this signal processing technique depends greatly on whether it is realizable on a standard (digital) CMOS IC process. We discuss now the circuit design of the required components, including the photosensors, and the special considerations for layout of this highly interconnected 2D network as a monolithic integrated circuit.

A. Logarithmic Photoreceptor

An image focused on the chip surface may be sampled by a matrix of photoreceptors, one at every node of the network. The intensity across a simple image may vary by two to three orders of magnitude in a laboratory environment, more in natural backgrounds, so a linear photoreceptor, which converts the intensity to a proportional voltage or current, would drive the active circuits in the network into saturation. A logarithmic photoreceptor is therefore required, and as studies on image processing have shown, perfectly adequate for the task on hand [3]. Photosensing is most economically obtained using the parasitic vertical bipolar in a CMOS well as a phototransistor, whose collector current becomes proportional to the light intensity incident on the collector junction along the well boundary. This may be compressed into a logarithmic voltage by a diode-connected MOSFET biased in the subthreshold region by the small photocurrent density produced under room lighting conditions. A compact logarithmic photoreceptor is in this way obtained with a two-transistor circuit [20], [21] (Fig. 5).

Although the stimulus to the prototype network in the discussion above was a voltage source in series with the variable resistor R_0, the circuits for the photosensor output and R_0 (described below) are naturally grounded on one end, so the Norton transformation must be invoked to convert the stimulus into a parallel combination of a grounded current source and a shunt resistor. A transconductance photoreceptor buffer was used, consisting a

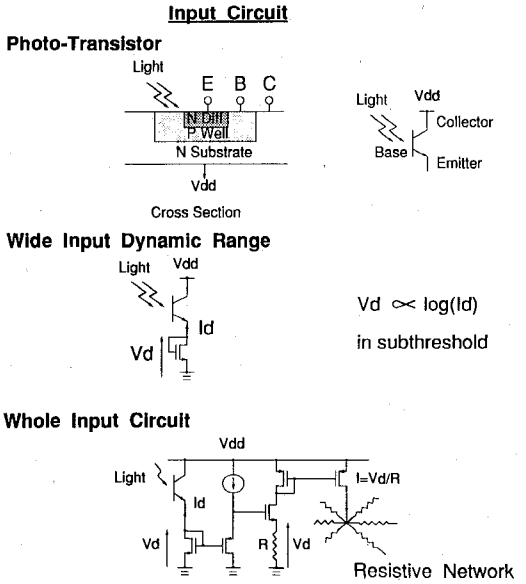


Fig. 5. The vertical bipolar transistor in a CMOS well produces logarithmic compression at the gate voltage by a MOSFET in subthreshold. A transconductance buffer drives the network.

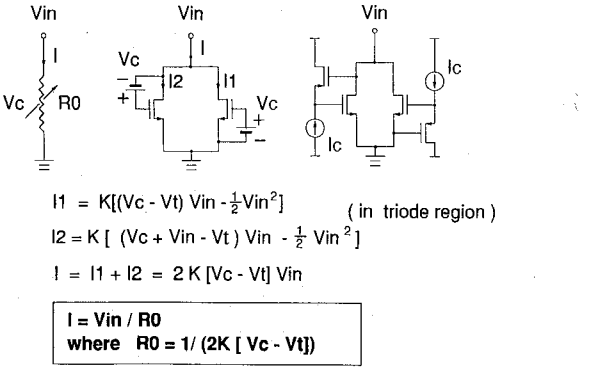


Fig. 6. The linearized variable resistor, with implementation of gate bias.

level-shift PMOS driving a resistively degenerated NMOSFET, which appears to the photoreceptor as a voltage-controlled current source (Fig. 5).

B. Variable Resistor

The width of the convolution kernel is set by a resistor R_0, whose value should ideally be continuously variable under user control. A single MOSFET operating in triode region used as a variable resistor would introduce an undesirable parabolic nonlinearity in the $I-V$ characteristics. Two MOSFET's in parallel obeying the simplified square law equations, however, can exactly cancel each other's parabolic nonlinearity in the triode region of operation if their gate biases are applied in a particular way, and the resulting linearized resistance is controlled by the bias. We used this as the variable resistor (Fig. 6). The floating-gate bias voltages were obtained as the V_{GS} of source-follower FET's carrying a control current.

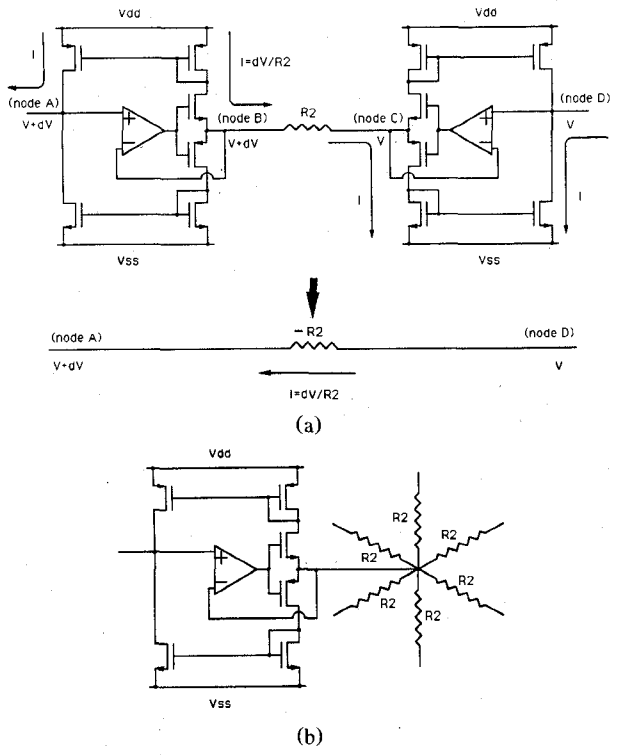


Fig. 7. (a) An NIC inverts the polarity of a resistor. (b) One NIC serves all resistors converging on a node.

The mean network voltage at a given level of photosensor illumination will change with the convolution width: for example, when the convolution width is decreased by making all R_0 large, the mean voltage will also increase because the buffered photocurrents will flow into larger resistors. This will impose the unnecessary demand of a large common-mode range of operation in active circuits such as R_0. We used a scheme to normalize the network inputs by slaving the buffer transconductance of the logarithmic photoreceptor proportionally to R_0, so as to maintain a constant mean network voltage at all illuminations.

C. Network Resistors

The 5-kΩ resistors for the nearest-neighbor internode connections in the network were implemented using p-well diffusions. A Gaussian convolution kernel would be obtained in spite of tolerances in the p-well resistivity as long as the relative magnitude of the positive and negative resistors remains 1:4. To make this ratio on the chip depend only on geometry, both R_1 and R_2 were implemented in the same material, p-well diffusion, and a negative impedance converter (NIC) was attached to R_2 to invert its polarity.

Our NIC implementation (Fig. 7) consists of the combination of a voltage follower and current inverter. The op-amp-based followers at each end of R_2 impose across it the potential difference at their inputs, and the resulting current flow, forced through the Class-B type output stages, is sourced from or sunk into the positive or negative power supply. Current mirrors in series then apply the same current at the input leads of the followers, inverting the sense of current flow as perceived at the network nodes. A negative resistance $-R_2$ is presented to the network.

Six negative resistors converge on every node in this hexagonal mesh. Six different NIC's are, however, not required at each node; instead, a single NIC placed at the node *after* the confluence of the resistors will simultaneously make them all negative (Fig. 7(b)). The dc gain in a simple five-FET op amp was large enough to obtain accurate inversion of the resistor $I-V$ characteristics and eliminate the crossover nonlinearity in the Class-B stage. The NIC at every node thus contained only 11 FET's.

D. Layout Considerations

A key concern in the implementation of this network as an IC is whether the usual two layers of metal and one of polysilicon can implement the starlike fan-out of interconnections emanating from every node. We proved to ourselves at the outset of this work that this was possible. A hexagonal grid was obtained by horizontally staggering successive rows of cells, and their interconnections implemented on a Manhattan geometry (Fig. 8(a)). All three available layers of interconnect were used to create abuttable cells. The power, ground, control, and output rails ran parallel to these rows from edge to edge of the chip.

A unit cell, including its portion of interconnect, measured 170×200 μm in 2-μm CMOS (Fig. 8(b)). The area of the photoreceptor collector–base junction, the blank rectangle in the cell layout at the lower left, measured 56×24 μm. No wires were allowed to traverse the photosensor because metal would absorb the incident light. Parasitic photocurrents generated in the source/drain junctions of other active circuits would have negligible effect on the voltages at the low-impedance nodes there. We observe finally that the active circuits occupied only 57% of the cell area, a measure of the toll exacted by the richness of interconnect in this circuit.

E. Output Means

This convolution network accepts a 2D input in the form of an incident image, does 2D signal processing across the resistive mesh, but on a standard IC is restricted to *1D output* at the pins along the periphery. The output therefore must be read at the pins (Fig. 9) by accessing one row of nodes at a time, and, at least in this implementation, becomes the bottleneck to the throughput rate. Addressable MOS switches were used to connect every node to output lines, and on-chip vertical bipolar transistors connected as emitter followers served as analog buffers at the pads. The speed of signal processing was determined by the relaxation time of this unclocked network, but a clock was introduced at the output to scan out the rows. To relieve this bottleneck, one can

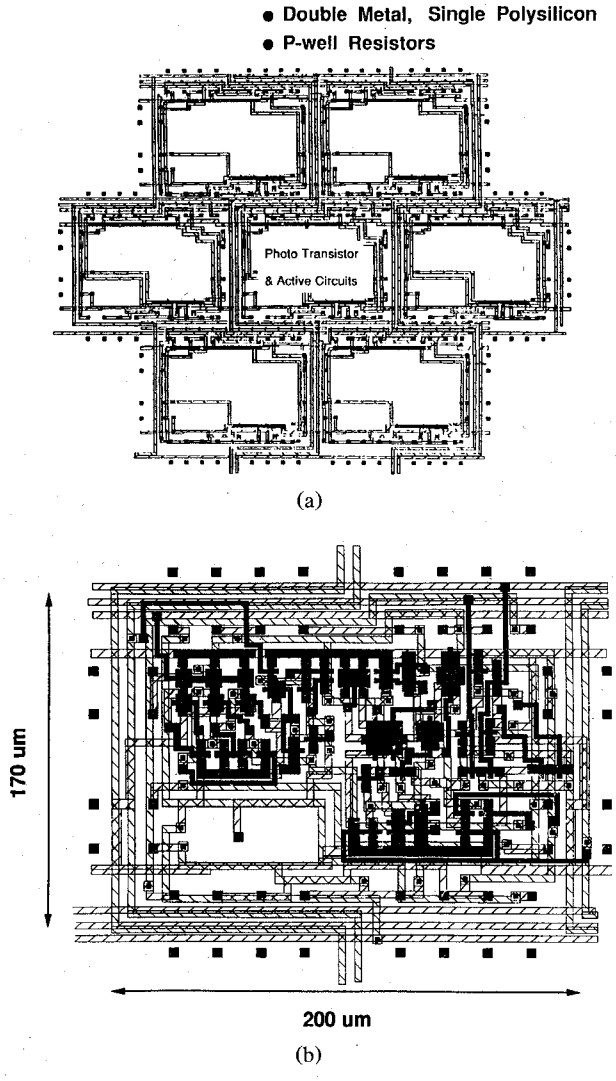


Fig. 8. (a) The layout of interconnects among a cluster of seven cells on a hexagonal grid; the blank areas contain the photoreceptor and associated active circuits in each cell. (b) Unit cell layout.

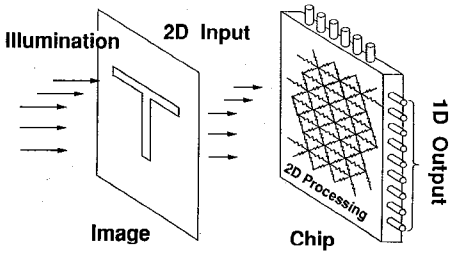


Fig. 9. Output mechanism. The network has 2D input, accomplishes 2D signal processing, but is forced to output results in 1D.

envisage connecting several 2D computational IC's performing a cascade of low-level vision tasks, with micro solder balls joining together matrices of pads on their surfaces, or through via holes on the back sides of the chips. This technique, originally developed for "flip-chip" mounting, is used at very high densities today to mate 2D focal plane array sensors to active substrates [22]. Once the desired data reduction has taken place at the output

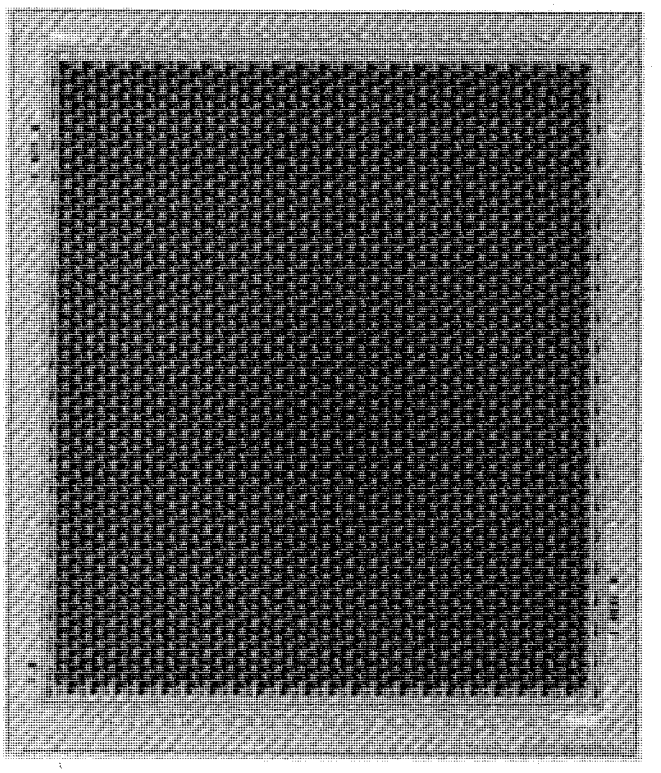


Fig. 10. Chip photograph.

of the such a cascade of chips, a few high-level outputs containing image features could be scanned out in parallel on pins with no loss in throughput speed.

IV. Experimental Results

We were able to fit a 45×40 array of unit cells on a 7.9×9.2-mm die, the largest die size available to us through the MOSIS foundry service. Power supplies of $+5$ and -5 V were used, mainly for convenience in circuit design; the circuits could be modified with a minor effort for operation on a single 5-V supply. The fabricated chip (Fig. 10) contained more than a 100 000 transistors and was fully functional.

The network response to optical input was measured by shining light on the exposed chip, and reading the outputs using a specially developed interface board under control of a personal computer. An array of analog column voltages along an addressed row were digitized and stored, and the smoothed output image reconstructed on the computer screen after all rows had been scanned.

A. Component Characteristics

Test circuits were included to independently verify operation of some of the key building blocks in the network. The log compression FET and the transconductance buffer following the photosensor gave the desired log-linear relationship across 2.5 decades of photocurrent (Fig. 11(a)). The variable resistor could be changed by the control current by a factor of 16:1 in magnitude, from 20 to 320 kΩ (Fig. 11(b)). The network simulations described

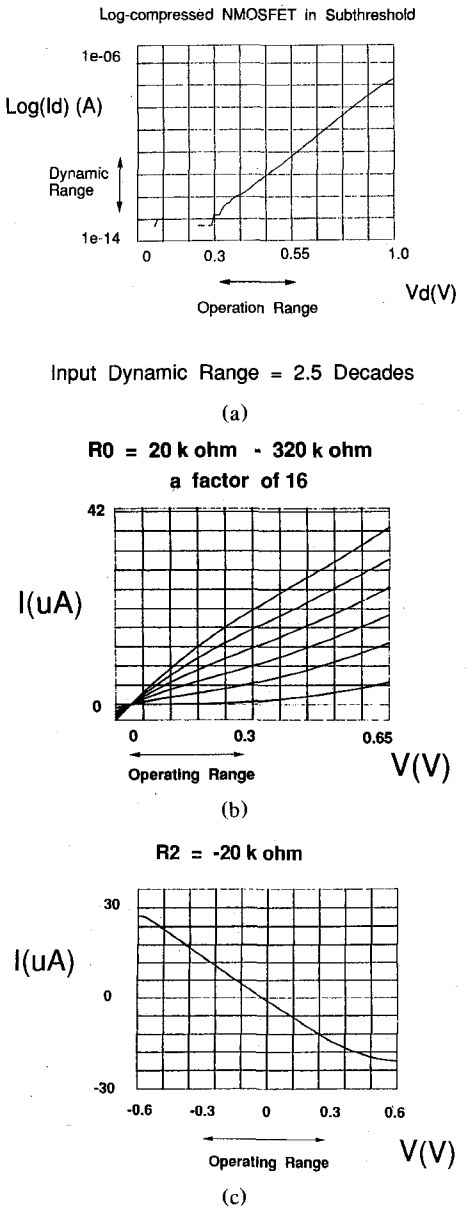


Fig. 11. Measured characteristics of the component circuits: (a) logarithmic compression at the photoreceptor output (V_d) versus photocurrent; (b) the variable resistor, which becomes nonlinear when one FET goes from triode to saturation; and (c) the negative resistor.

previously predict that this would yield the desired 2:1 variation in convolution width. A strong nonlinearity in the $I-V$ characteristics appeared for voltages larger than 0.3 V, but we had designed the range of the network voltages not to exceed this value under normal illumination. A negative resistor of the desired value was also obtained (Fig. 11(c)), with very little observable nonlinearity at applied voltages of 0.3 V of either polarity.

B. Response to Optical Inputs

The network function was characterized with two simple incident images, a pinhole excitation representing a spatial impulse, and the character "T." The images were produced on the chip surface by light transmitted through a mask used in place of the lid on the cavity of the ceramic PGA package. We had also made provision on the IC to measure the actual compressed signal driving the network, so that the true network function could be obtained by deconvolving it from the measured output.

The convolution kernel was thus deduced from measurements of the network input and output (Fig. 12). It was difficult at this sampling resolution to accurately

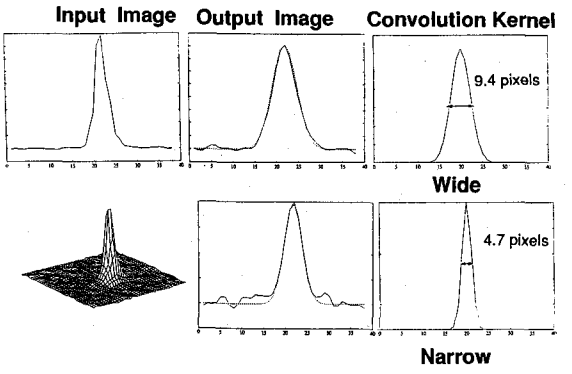


Fig. 12. Measured convolution kernel of the network. The measured network stimulus is deconvolved from the output. Dashed lines superimposed on output show the numerical smoothing used.

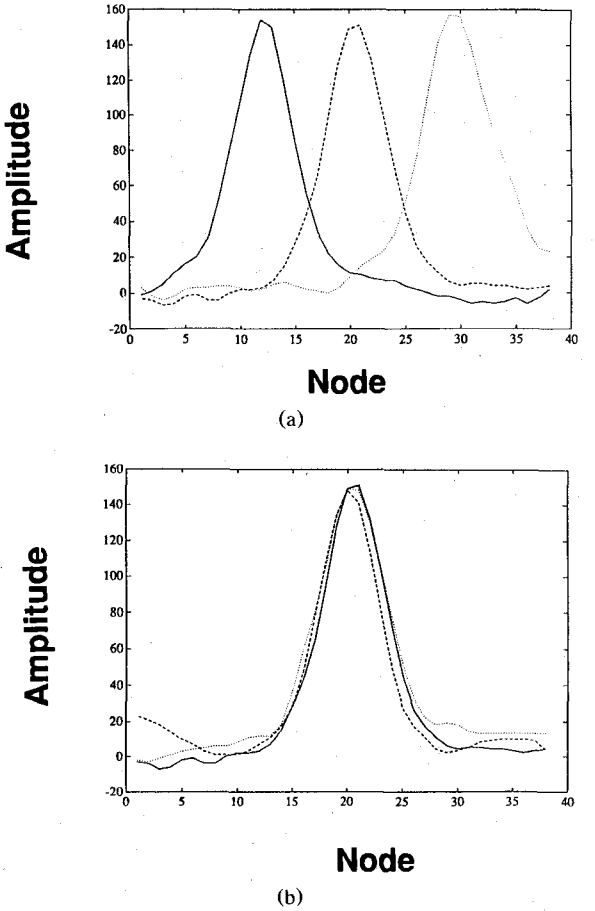


Fig. 13. The uniformity of output (a) across one chip, and (b) between three chips.

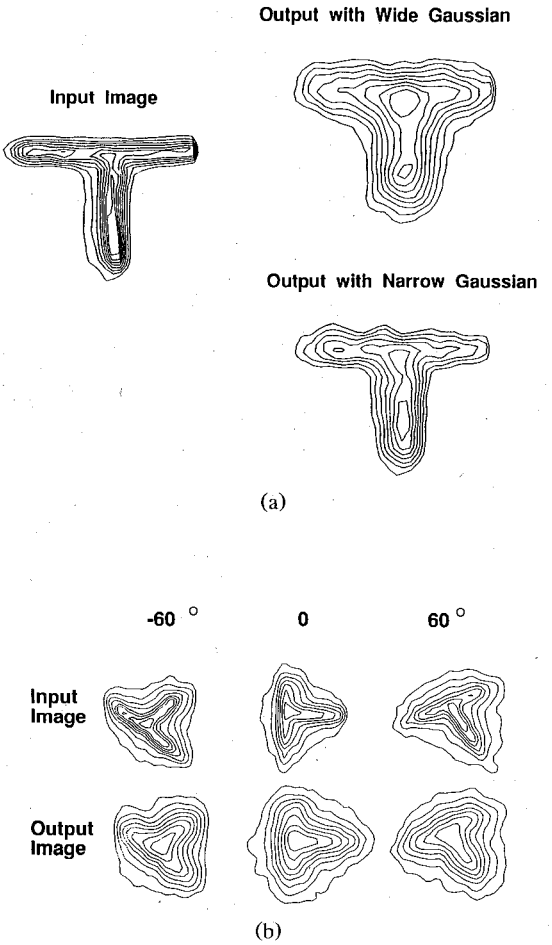


Fig. 14. (a) Measured outputs at two different smoothing widths on character "T." (b) Uniformity of network action versus rotation.

ascertain that it was a Gaussian function, but the characteristic inflection in the function as it approaches the peak value was evident. This would not appear unless the network contained negative resistors. We were able to change the full width at half maximum of the kernel by a factor of 2, from 4.7 to 9.4 pixels wide, by changing R_0 across its full span with the control current. The network output was most noisy at its tails at minimum R_0, and we had to use smoothing in the sense of a least-mean-square fit to deduce the kernel function. Light through the pinhole nominally sampled only a small neighborhood on the chip; we moved the pinhole to points on the chip either side of the center, and found an acceptable uniformity in the response (Fig. 13(a)), which is determined here by MOSFET matching across the extent of the chip surface [23]. The slight uptilt of the output at the ends of the measured response was caused by the edge effect when the network terminates at the chip boundary. The uniformity across three chips was also acceptable at this sampling resolution (Fig. 13(b)), except for one chip where a particularly large uptilt appears.

The smoothing effected by the network on a character "T" was also measured (Fig. 14(a)), and its symmetry after rotations relative to the chip axis verified (Fig. 14(b)). Both were satisfactory.

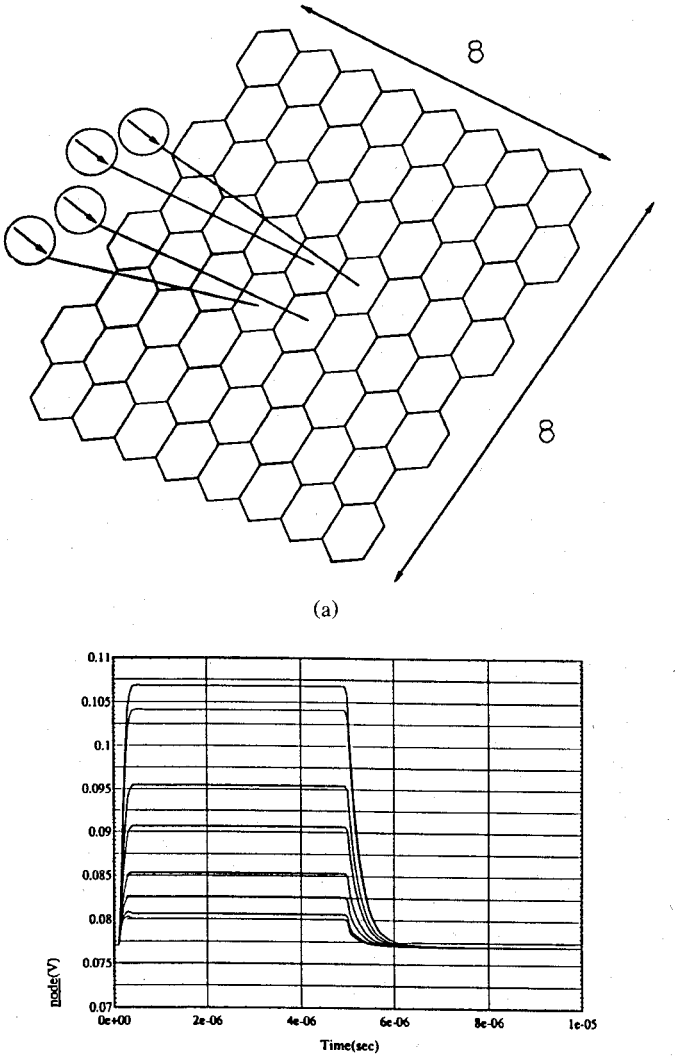


Fig. 15. (a) An 8×8 subnetwork simulated at the transistor level on SPICE, and (b) at various distances away from excitation, showing settling within 2 μs.

Precautions were required in making the measurement to compensate for the effects of the 2-W power dissipation when no heat sink was mounted on the package. This large power dissipation produced a thermal gradient across the IC, peaked at the center with circularly symmetric isotherms spreading out towards the chip boundary. We deduced this from a corresponding pattern in photoreceptor dark currents, which appeared as a stimulus to the network in the absence of an optical input. This had to be calibrated and subtracted from all measurements to obtain the true optical response. We emphasize that this relatively large power dissipation was not fundamental to the network; 75% of it was due to an unnecessarily large bias current in one building block, the control circuit for the variable resistor. A further reduction in quiescent power could be obtained by devising a voltage drive to the network nodes, because the current sources in the present implementation produce some steady power dissipation through R_0, even when the chip is not illumi-

TABLE I
ELECTRICAL CHARACTERISTICS

Photosensor sites	45×40
Sampling geometry	Hexagonal
Area per pixel	170×200 μm
Rise time of network (10–90%)	2 μs
Rise time of photosensors	20 μs
Width of convolution (FWHM)	4.7–9.4 pixels
Chip size	7.9×9.2 mm
Technology	2-μm CMOS, single poly, double metal
Power dissipation	2 W (75% in one function block)

nated. The power dissipation could be made even smaller by scaling down all the currents in the IC, but at a trade-off of longer relaxation times.

The settling time of the entire network in response to a step input from the photoreceptors determined the 2D computational speed. For all practical purposes, a step change in a photoreceptor has only to propagate a few nodes away before the decay in the convolution function will swamp it out, and the voltages at nodes farther away will remain relatively unchanged. We simulated an 8×8 subnetwork at the transistor level on SPICE, and the results indicated settling in less than 2 μs in response to a step in photocurrent (Fig. 15). However, a settling time of 20 μs was experimentally observed in response to illumination from a light chopper, which we surmise was dominated by the slow response of the phototransistors [20]. The graceful settling in the transient SPICE simulation verified the stability of the network response in time. A similar waveform of the settling of node voltages was also observed experimentally.

The electrical performance of the Gaussian convolution IC is summarized in Table I.

V. CONCLUSIONS

Parallel processing of images per pixel will offer the highest possible speed in functions related to low-level vision. This is indeed the present trend in real-time hardware for digital image processing. We have described a single-chip *analog* implementation of this concept to perform a Gaussian convolution with the use of an active mesh. Although it may be argued that a variable focus lens also effects this function, there are two significant differences: the active resistive mesh may be extended to many different convolution functions, including orientation selective ones [19], most of which cannot be simply implemented with geometric optics; furthermore, no mechanical system could attain the physical compactness and microsecond control of the convolution functions. The difference in output of two independent meshes on the same chip, for example, could implement the much sought after difference of Gaussian function in image processing [3]. In short, the notion of an active mesh opens many new opportunities for realizing application-specific analog signal processors. Digital signal processors have as advantages an immunity to component noise and mismatches, more ready programmability, and shorter development times, but tend to be considerably larger chips than their analog equivalents. On the other hand, inaccuracies in analog computation may not be limitations in low-level vision functions, but much more of a detriment in high-level classification tasks. This leads us to believe that compact hardware with the least power dissipation to implement real-time image recognition and classification may ultimately consist of a judicious mix of analog computation of the type described here, and conventional digital signal processing.

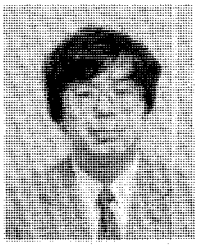
ACKNOWLEDGMENT

The formulation of the network was influenced in the early stages by B. Mathur and H. T. Wang of Rockwell International Science Center, and by our colleague R. L. Baker. A. Nahidipour designed and constructed the interface board used to measure the chip response. B. Furman contributed to simulations of the network action on complex images. Transient simulations of the network were carried out at the University of California at San Diego Supercomputer Center with support from the National Science Foundation.

REFERENCES

[1] P. A. Ruetz and R. W. Brodersen, "Architectures and design techniques for real time image processing ICs," *IEEE J. Solid-State Circuits*, vol. SC-22, pp. 233–250, Apr. 1987.
[2] J. Dowling, *The Retina: An Approachable Part of the Brain*. Cambridge, MA: Harvard University Press, 1987.
[3] D. Marr, *Vision*. San Francisco, CA: W. H. Freeman, 1982.
[4] C. A. Mead and M. A. Mahowald, "A silicon model of early visual processing," *Neural Networks*, vol. 1, pp. 91–97, 1988.
[5] E. A. Vittoz, "Future of analog in the VLSI environment," in *Proc. ISCAS* (New Orleans, LA), May 1990, pp. 1372–1375.
[6] B. Gilbert, "Translinear circuits: A proposed classification," *Electron. Lett.*, vol. 11, pp. 14–16, 1975.
[7] B. Gilbert, "A monolithic 16 channel analog array normalizer," *IEEE J. Solid-State Circuits*, vol. SC-19, pp. 954–963, Dec. 1984.
[8] H. Kobayashi, J. L. White, and A. A. Abidi, "An analog CMOS network for Gaussian convolution with embedded image sensing," in *ISSCC Dig. Tech. Papers* (San Francisco, CA), Feb. 1990, pp. 216–217.
[9] T. Poggio, H. Voorhees, and A. Yuille, "A regularized solution to edge detection," Mass. Inst. Technology, Cambridge, MA, AI Memo, May 1985.
[10] T. Poggio, V. Torre, and C. Koch, "Computational vision and regularization theory," *Nature*, vol. 317, pp. 314–319, Sept. 1985.
[11] J. Babaud, A. P. Witkin, M. Baudin, and R. O. Duda, "Uniqueness of the Gaussian kernel for scale-space filtering," *IEEE Trans. Pattern Anal. and Mach. Intell.*, vol. PAMI-8, pp. 26–33, Jan. 1986.
[12] T. K. Hogan, "A general experimental solution of Poisson's equation for two independent variables," *J. Inst. Eng. (Australia)*, vol. 15, pp. 89–92, Apr. 1943.
[13] G. Liebmann, "Solution of partial differential equations with a resistance network analogue," *Brit. J. Appl. Phys.*, vol. 1, pp. 92–103, Apr. 1950.
[14] G. W. Swenson, Jr. and T. J. Higgins, "A direct current network analyzer for solving wave equation boundary value problems," *J. Appl. Phys.*, vol. 23, pp. 126–131, Jan. 1952.
[15] J. R. Hechtel and J. A. Seeger, "Accuracy and limitations of the resistor network used for solving Laplace's and Poisson's equations," *Proc. IRE*, vol. 49, pp. 933–940, May 1961.
[16] C. A. Mead, *Analog VLSI and Neural Systems*. Reading, MA: Addison Wesley, 1989.
[17] T. Poggio and C. Koch, "Ill-posed problems in early vision: From computational theory to analogue networks," *Proc. Roy. Soc. London*, vol. B-226, pp. 303–323, 1985.

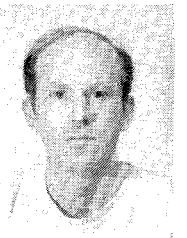
[18] D. Dudgeon and R. Mersereau, *Multidimensional Signal Processing*. Englewood Cliffs, NJ: Prentice Hall, 1984.
[19] J. L. White and A. A. Abidi, "Analysis and design of parallel analog computational networks," in *Proc. Int. Symp. Circuits Syst.* (Portland, OR), June 1989, pp. 70–73.
[20] S. G. Chamberlain and J. P. Y. Lee, "A novel wide dynamic range silicon photodetector and linear imaging array," *IEEE J. Solid-State Circuits*, vol. SC-19, pp. 41–48, Feb. 1984.
[21] C. Mead, "A sensitive electronic photoreceptor," in *Proc. 1985 Chapel Hill Conf. VLSI* (Chapel Hill, NC), 1985, pp. 463–471.
[22] S. B. Stetson, D. B. Reynolds, M. G. Stapelbroek, and R. L. Stermer, "Design and performance of blocked impurity band detector focal plane arrays," in *Proc. SPIE*, vol. 686 (San Diego, CA), Aug. 1986, pp. 48–65.
[23] M. J. M. Pelgrom, A. C. J. Duinmaijer, and A. P. G. Welbers, "Matching properties of MOS transistors," *IEEE J. Solid-State Circuits*, vol. 24, no. 5, pp. 1433–1440, Oct. 1989.



Haruo Kobayashi was born in Utsunomiya, Japan, in 1958. He received the B.S. and M.S. degrees in information physics and mathematical engineering from the University of Tokyo, Tokyo, Japan, in 1980 and 1982, respectively. From 1987 to 1989 he was at the University of California, Los Angeles, where he received the M.S. degree in electrical engineering in 1989.

He joined Yokogawa Electric Corporation, Tokyo, Japan, in 1982, where he has been engaged in the research and development of an FFT analyzer, a mini-supercomputer, and an LSI tester.

Mr. Kobayashi is a member of the Institute of Electronics, Information and Communication Engineers of Japan and the Society of Instrument and Control Engineers of Japan.



Joseph L. White (S'88) received the B.S. degree in applied physics from the California Institute of Technology, Pasadena, in 1982, and the M.S. and Ph.D. degrees in electrical engineering from the University of California, Los Angeles, in 1983 and 1991, respectively.

He has previously worked for the Hughes Aircraft Space and Communications Group and the Rand Corporation. His research interests include image processing and computer vision.

Asad A. Abidi (S'75–M'81) was born in 1956. He received the B.Sc.(Hon.) degree from Imperial College, London, in 1976 and the M.S. and Ph.D. degrees in electrical engineering from the University of California, Berkeley, in 1978 and 1981, respectively.

He was at Bell Laboratories, Murray Hill, NJ, from 1981 to 1984 as a Member of the Technical Staff in the Advanced LSI Development Laboratory. Since 1985 he has been at the Electrical Engineering Department of the University of California, Los Angeles, where he is an Associate Professor. He was a Visiting Faculty Researcher at Hewlett Packard Laboratories during 1989. His research interests are in high-speed analog integrated circuit design, parallel analog signal processing techniques, device modeling, and nonlinear circuit phenomena.

Dr. Abidi served as the Program Secretary for the International Solid-State Circuits Conference from 1984 to 1990, and is presently associated with the Symposium on VLSI Circuits, with the IEEE Solid-State Circuits Council, and as an Associate Editor with the IEEE JOURNAL OF SOLID-STATE CIRCUITS. He received the 1988 TRW Award for Innovative Teaching.

Spatial Versus Temporal Stability Issues in Image Processing Neuro Chips

Takashi Matsumoto, *Fellow, IEEE,* Haruo Kobayashi, *Member, IEEE,* and Yoshio Togawa

***Abstract*—A typical image processing neuro chip consists of a regular array of very simple cell circuits. When it is implemented by a CMOS process, two stability issues naturally arise:**

- **i) Parasitic capacitors of MOS transistors induce the *temporal* dynamics. Since a processed image is given as the stable limit point of the temporal dynamics, a temporally unstable chip is unusable.**
- **ii) Because of the array structure, the node voltage distribution induces the *spatial* dynamics, and it could behave in a wild manner, e.g., oscillatory, which is highly undesirable for image processing purposes, even if the trajectory of the temporal dynamics converges to a stable limit point.**

The main contributions of this paper are (i) a clarification of the spatial stability issue; (ii) explicit *if and only if* conditions for the temporal and the spatial stability in terms of circuit parameters; (iii) a rigorous explanation of the fact that even though the spatial stability is stronger than the temporal stability, the set of parameter values for which the two stability issues disagree is of (Lebesgue) *measure zero*; and (iv) theoretical estimates on the processing speed.

I. Introduction

A. Motivation

THIS study has been motivated by the temporal versus spatial stability issues of an image smoothing neuro chip [1]. The function of the chip is to smooth a two-dimensional image in an extremely fast manner. It consists of the 45×40 hexagonal array of very simple "cell" circuits, described by Fig. 1. An image is projected onto the chip through a lens (Fig. 2) and the photo sensor represented by the current source in Fig. 1 inputs the signal to the processing circuit. The output (smoothed) image is represented as the node voltage distribution of the array. With an appropriate choice of $g_0 > 0$, $g_1 > 0$, and $g_2 < 0$, the chip performs a *regularization* with *second-order* smoothness constraint and closely approximates the Gaussian convolver, which is known to have an optimal S/N as a preprocessor for edge detection [2], [3]. (APPENDIX IV explains why a regularization with second-order smoothness constraint demands negative conductance.) Conductance g_0 is designed to be variable in order to control the width of the Gaussian-like kernel. In engineering terms, this is a noncausal infinite impulse response (IIR) realization of a Gaussian-like convolver instead of a finite impulse response (FIR) realization, and this structure accomplishes high-speed processing while maintaining simplicity. The reader is referred to [1] for responses actually measured from the chip.

Since the *negative conductance* $g_2 < 0$ is involved, two stability issues naturally arise:

(i) Because the chip is fabricated by a CMOS process, parasitic capacitors induce the dynamics with respect to time. This raises the *temporal stability* issue of whether the network converges to a stable equilibrium point.

(ii) Because a processed (smoothed) image is given as the node voltage distribution of the array, the *spatial stability* issue also arises even if the temporal dynamics does converge to a stable equilibrium point. In other words, the node voltage distribution may behave in a wild manner, e.g., oscillatory.

In discussing relationships between the temporal and the spatial stability issues, several precautions need to be taken. In particular, it is important to realize that while the temporal dynamics is *causal*, i.e., $t \geq 0$, the spatial "dynamics" (a precise definition will be given later) is *noncausal*. Namely the spatial dynamics can go into the negative direction as well as the positive direction. Furthermore, the spatial dynamics is not an initial value problem but rather a *boundary value problem* which gives rise to several delicate issues.

Our earlier numerical experiments on these issues were rather intriguing. The results suggested that the network is temporally stable "if and only if" it is spatially stable. Fig. 3 shows spatial impulse responses at different sets of parameter values. For the sake of simplicity, the network is of a linear array instead of a two-dimensional array. The network has 61 nodes and the impulse is injected at the center node. Fig. 3(a) suggests that the network can be used for image smoothing because the response to an impulse is "bell-shaped." In fact, the Gaussian-like convolver chip [1] corresponds to Fig. 3(a) where g_0 is variable. Fig. 3(b) indicates that it can enhance contrast of an input image after smoothing because it inhibits the "surround" responses in addition to smoothing. Fig. 4 shows the corresponding temporal step responses of the center node. For simplicity, the only parasitic capacitors taken into account are those from each node to the ground. The responses shown in parts (a) and (b) of Fig. 4 are temporally stable while part (c) is not. Fig. 3(c) is spatially unstable because the response does not decay, which is highly undesirable for image

Manuscript received January 15, 1991; revised October 9, 1991. This work was supported by the Japanese Ministry of Education, the Ogasawara Foundation, the Casio Foundation, and the Science and Engineering Laboratory and Tokutei-Kadai of Waseda University.

T. Matsumoto is with the Department of Electrical Engineering, Waseda University, Tokyo 169 Japan.

H. Kobayashi is with Yokogawa Electric Corporation, Tokyo 180, Japan.

Y. Togawa is with the Department of Information Science, Science University of Tokyo, Tokyo, Japan.

IEEE Log Number 9105239.

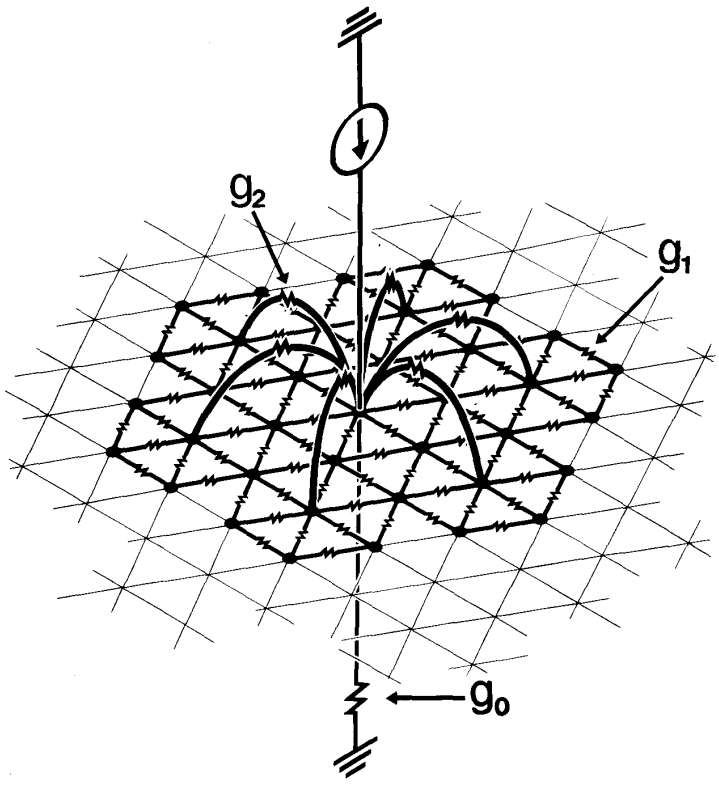


Fig. 1. The image smoothing neurochip. Only one "unit" is shown.

processing purposes. (A precise definition of spatial stability will be given later.) All of our earlier numerical experiments, including those shown in Fig. 3 and Fig. 4, suggested the equivalence of the two stability conditions. However there are no *a priori* reasons for them to be equivalent. As will be shown rigorously, the two stability conditions are *not* equivalent. The spatial stability condition is *stronger* than the temporal stability condition. Nevertheless, the set of parameter values (g_0, g_1, g_2) for which the two stability conditions disagree turns out to be a (Lebesgue) *measure zero* subset, which explains why our numerical experiments suggested equivalence between the two conditions. (A measure zero subset is difficult to "hit"). We will prove, in a very general setting, that the network is temporally stable if and only if it is *spatially regular,* a new concept which is weaker than the spatial stability, and it amounts to a decomposability of eigenvalues of a matrix describing the spatial dynamics. Explicit analytic conditions will be given for the temporal as well as the spatial stabilities in a general setting. Also given is an estimate on the speed of temporal responses of the networks.

Since our results are proved in a general setting, they can be applied to other neural networks of a similar nature, e.g. oriented receptive field filters [4] and Gabor filters [5], which we intend to pursue in our future projects. The results in this paper, however, are only for linear array cases. Extensions to two-dimensional array cases are nontrivial and are left for a future paper.

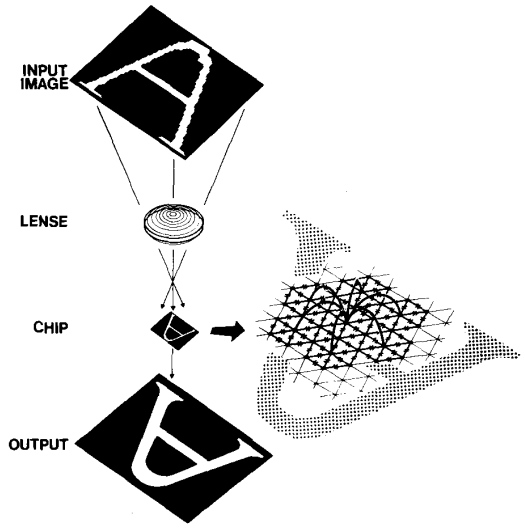


Fig. 2. A schematic diagram.

B. Related Works

A serious stability analysis is performed in [6] for lateral inhibition networks that are present, at least partly, in most

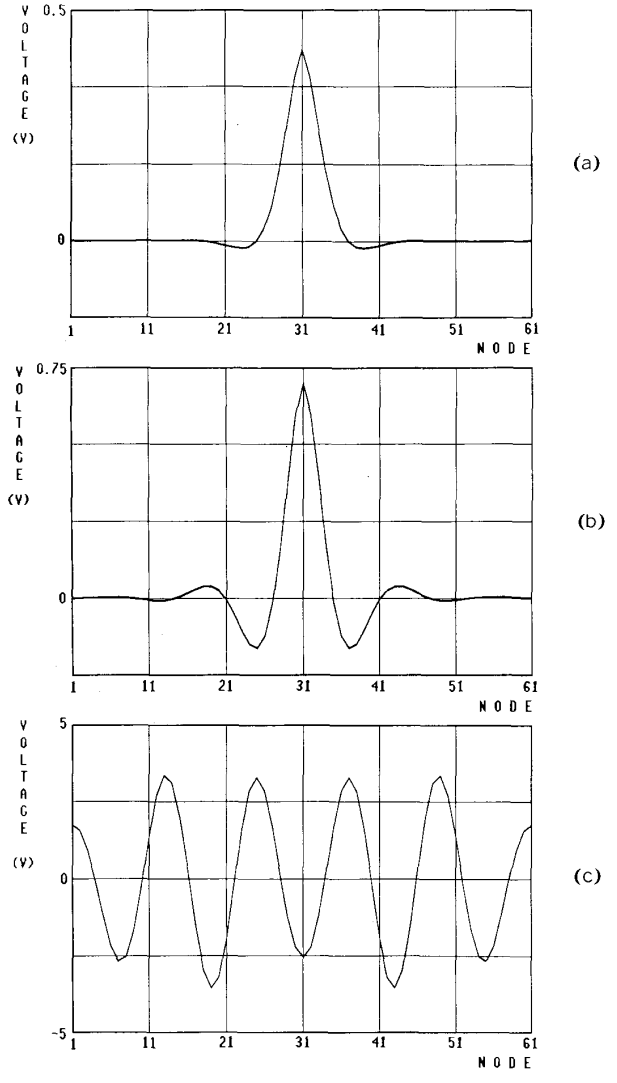


Fig. 3. Spatial impulse responses with $n = 61$, $m = 2$, $1/g_0 = 200$ kΩ, $1/g_1 = 5$ kΩ, $u_{31} = 10$ μA, $u_k = 0$ for $k \neq 31$. (a) $1/g_2 = -20$ kΩ; stable. (b) $1/g_2 = -18$ kΩ; stable. (c) $1/g_2 = -17$ kΩ; unstable.

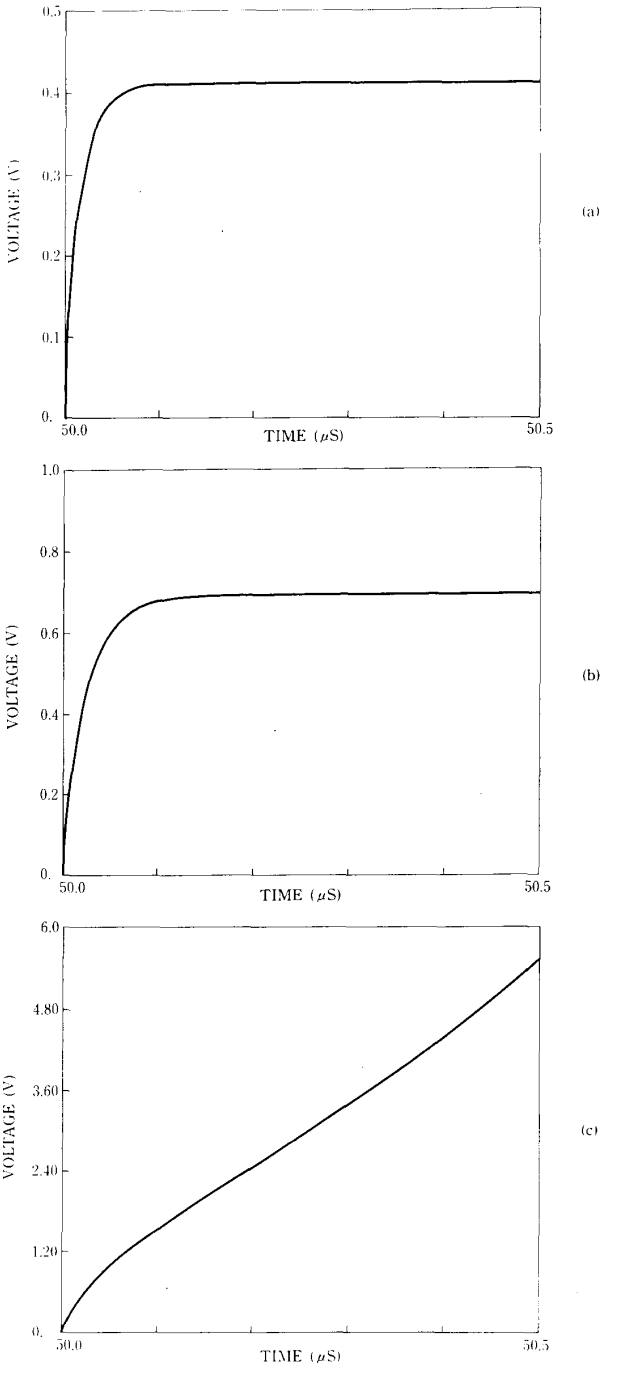


Fig. 4. Temporal step responses of the center node $v_{31}(t)$ with $n = 61$, $m = 2$, $1/g_0 = 200$ kΩ, $1/g_1 = 5$ kΩ, $c_0 = 0.1$ pF,

$$u_{31}(t) = \begin{cases} 0 & t < 50\ \mu s \\ 10\ \mu A & t \geq 50\ \mu s \end{cases}$$

$u_k(t) \equiv 0$ for $k \neq 31$. (a) $1/g_2 = -20$ kΩ; stable. (b) $1/g_2 = -18$ kΩ; stable. (c) $1/g_2 = -17$ kΩ; unstable.

of the early vision chips, e.g., [7]–[14] and the networks considered in the present paper. Each node has conductance connections only with immediate neighbors. However, the MOS capacitors, nonlinearities of MOS conductances, and amplifiers in the input circuit could cause, depending on the design, oscillations. On the one hand, the problem in [6] is more difficult than the one discussed in this paper because nonlinearities must be taken into account. On the other hand, it is simpler in the sense that each node has connections only with its immediate neighbors. In [6] several sufficient conditions are given for temporal stability using a rather interesting argument. We close this section by noting that the observation was made in [15] that active conductances can cause instability in early vision neural networks.

II. Stability–Regularity

Subsection A explains how the temporal and the spatial dynamics are described. It is pointed out that the boundary

conditions should be carefully examined for the spatial dynamics. Subsection B characterizes the spatial dynamics in terms of the eigenspaces of the matrix describing the dynamics. The first main result, Theorem 1, clarifies conditions under which spatial responses behave properly. In particular, it states that in addition to a condition on the eigenvalues of the matrix describing the dynamics, another condition on the boundary is necessary. In subsection C the second main result, Theorem 2, reveals a fundamental relationship between the temporal and the spatial dynamics by showing that a network is temporally stable if and only if it is spatially regular, a new concept to be defined. Propositions 2 and 3 give the stability as well as the regularity criteria in terms of the characteristic polynomial of the matrix describing the spatial dynamics.

A. Formulation

Consider a neural network consisting of a linear array of n nodes where each node is connected with its pth nearest neighborhoods, $p = 1, \cdots, m < n$ via a (possibly negative) conductance g_p and a capacitance c_p. Fig. 5 shows the case where $m = 3$. The network is described by

$$\sum_{p \in M} b_p \frac{dv_{i-p}}{dt} = \sum_{p \in M} a_p v_{i-p} + u_i, \qquad i = 1, \cdots, n, \quad (1)$$

where v_i and u_i are the voltage and the input current at the ith node, and

$$M = \{p \text{ integer } ||p| \leq m\} \quad (2)$$

$$a_0 = -\left(g_0 + 2\sum_{p=1}^{m} g_p\right) \quad (3)$$

$$a_{\pm p} = g_p, \qquad 1 \leq p \leq m$$

$$b_0 = c_0 + 2\sum_{p=1}^{m} c_p \quad (4)$$

$$b_{\pm p} = -c_p, \qquad 1 \leq p \leq m.$$

Equation (1) is obtained simply by writing down the Kirchhoff's current law (KCL) at each node. Letting $\boldsymbol{v} = (v_1, \cdots, v_n)^T$ and $\boldsymbol{u} = (u_1, \cdots, u_n)^T$ (T denoting transpose), one can recast (1) as

$$\boldsymbol{B}\frac{d\boldsymbol{v}}{dt} = \boldsymbol{A}\boldsymbol{v} + \boldsymbol{u} \quad (5)$$

where

$$\boldsymbol{A} = \begin{bmatrix} a_0 & a_1 & \cdots & a_m & 0 & \cdots & 0 \\ a_1 & & & & & & \vdots \\ \vdots & & & & & & 0 \\ a_m & & & & & & a_m \\ 0 & & & & & & \vdots \\ \vdots & & & & & & a_1 \\ 0 & \cdots & 0 & a_m & \cdots & a_1 & a_0 \end{bmatrix} \quad (6)$$

$$\boldsymbol{B} = \begin{bmatrix} b_0 & b_1 & \cdots & b_m & 0 & \cdots & 0 \\ b_1 & & & & & & \vdots \\ \vdots & & & & & & 0 \\ b_m & & & & & & b_m \\ 0 & & & & & & \vdots \\ \vdots & & & & & & b_1 \\ 0 & \cdots & 0 & b_m & \cdots & b_1 & b_0 \end{bmatrix} \quad (7)$$

Note that $\boldsymbol{A}$ as well as $\boldsymbol{B}$ is symmetric and has a uniform band structure, which, as will be seen, yields interesting properties. If $\boldsymbol{B}$ is nonsingular, an equilibrium point of (5) satisfies

$$-\sum_{p \in M} a_p v_{i-p} = u_i \quad (8)$$

which is a difference equation instead of a differential equation. Assuming that $a_m \neq 0$, one has

$$v_{i+m} = -\frac{1}{a_m}\left(\sum_{p \in M - \{m\}} a_p v_{i+p} + u_i\right). \quad (9)$$

Therefore, letting

$$\boldsymbol{F} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & 0 & \cdots & 0 & 1 \\ -1 & -\frac{a_{m-1}}{a_m} & \cdots & -\frac{a_1}{a_m} & -\frac{a_0}{a_m} \end{bmatrix} \quad (10)$$

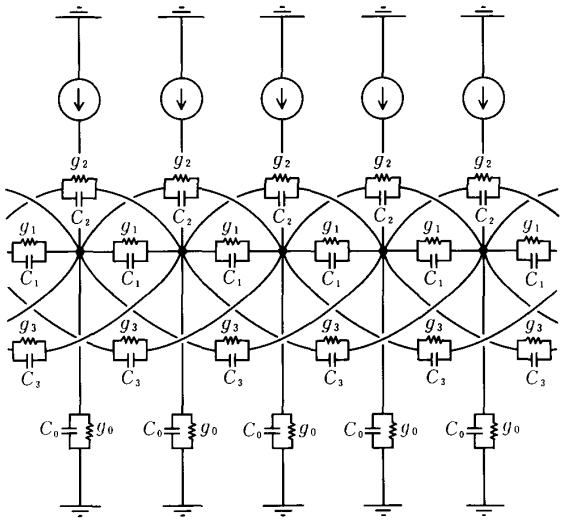


Fig. 5. Network described by (1) when $m = 3$.

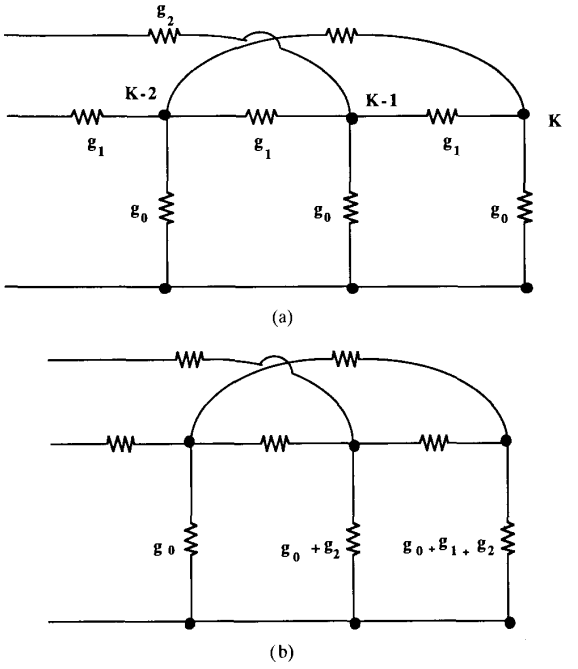


Fig. 6. Boundary conditions with $m = 2$ (a) The circuitry at the right end. (b) A modification of the boundary conditions establishes consistency.

with

$$\boldsymbol{x}_k = (v_{k-m}, v_{k-m+1}, \cdots, v_k, \cdots, v_{k+m-1})^T \in \mathbb{R}^{2m}$$
$$\boldsymbol{y}_k = (0, \cdots 0, -u_k/a_m)^T \in \mathbb{R}^{2m}$$

one can rewrite (9) as

$$\boldsymbol{x}_{k+1} = \boldsymbol{F}\boldsymbol{x}_k + \boldsymbol{y}_k \tag{11}$$

Observe that subscript k in (11) is not time. Equation (11) represents the *spatial dynamics* induced by the temporal dynamics (5). Note also that $\dim \boldsymbol{v} = n$, the number of nodes, while $\dim \boldsymbol{x}_k = 2m$, the size of the neighborhood, which is independent of n.

In image processing, input is $\boldsymbol{u}$ while output is $\boldsymbol{v}(\infty)$, the stable equilibrium point of (5). Equation (11) describes how the coordinates of $\boldsymbol{v}(\infty)$ are distributed with respect to k. There are several issues that need care.

First, the temporal dynamics given by (5) constitute an initial value problem while (8) or (11) is a boundary value problem. Namely, arbitrary $\boldsymbol{v}(0)$ and $\boldsymbol{u}(.)$ completely determine the solution to (5) while for (8) or (11), one *cannot* specify (for a given $\{\boldsymbol{y}_k\}$) an arbitrary $\boldsymbol{x}_0$ because a solution $\boldsymbol{x}_k$ must be consistent with the KCL's at the end points. Furthermore, the temporal dynamics given by (5) are *causal*; i.e., a solution at time t does not depend on the future. The spatial dynamics given by (11), however, are *noncausal*; i.e., a solution at node k depends on both the right-hand-side and left-hand-side neighbors. In order to be more specific, let us look at Fig. 6(a), where the right end point is shown with $m = 2$, $-K \le k \le K$, $n = 2K + 1$. Capacitors are omitted for the sake of simplicity. KCL's at the Kth and $(K-1)$th nodes are, respectively,

$$-(g_0 + g_1 + g_2)v_K + g_1 v_{K-1} + g_2 v_{K-2} = 0 \tag{12a}$$
$$-(g_0 + 2g_1 + g_2)v_{K-1} + g_1(v_K + v_{K-2}) + g_2 v_{K-3} = 0. \tag{12b}$$

The right-hand sides are nonzero when independent current sources are present. These equations define a two-dimensional linear subspace to which the boundary state $\boldsymbol{x}_K$ must belong. Another two-dimensional constraint is imposed at the left end. If these constraints are independent (generically they are), then a four-dimensional trajectory $\boldsymbol{x}_k \in \mathbb{R}^4$ is uniquely defined.

For a general m, there are m boundary conditions at the right end and there are another m conditions at the left end. An *impulse response* of (11), for instance, is determined in the following way. Let $\boldsymbol{y}_0 \neq \boldsymbol{0}$ whereas $\boldsymbol{y}_k = \boldsymbol{0}$ for $k \neq 0$ and consider $\boldsymbol{x}_0$, which is to be determined. Let $\mathbb{R}^{2m} \supset T_+$ (resp. T_-) be an m-dimensional linear subspace to which $\boldsymbol{x}_K$ (resp. $\boldsymbol{x}_{-K}$) must belong. Then

$$\boldsymbol{x}_K = F^K \boldsymbol{x}_0 + F^{K-1}\boldsymbol{y}_0 \in T_+ \tag{13a}$$

and

$$\boldsymbol{x}_{-K} = F^{-K}\boldsymbol{x}_0 \in T_-. \tag{13b}$$

determine $\boldsymbol{x}_0$ provided that T_+ and T_- are independent. Other $\boldsymbol{x}_k$'s are determined by

$$\boldsymbol{x}_k = \begin{cases} F^k \boldsymbol{x}_0 + F^{k-1}\boldsymbol{y}_0, & k \ge 1 \\ F^{-|k|}\boldsymbol{x}_0, & k \le -1 \end{cases}$$

Moving to the second issue, observe that the boundary conditions (12) are not consistent with the temporal dynamics (5) because the last two equations of an equilibrium are

$$-(g_0 + 2g_1 + 2g_2)v_n + g_1 v_{n-1} + g_2 v_{n-2} = 0 \tag{14a}$$
$$-(g_0 + 2g_1 + 2g_2)v_{n-1} + g_1(v_n + v_{n-2}) + g_2 v_{n-3} = 0. \tag{14b}$$

Here, we are slightly abusing our notations of K and n. There will be no confusion, however. The difference between (12) and (14) lies in the coefficients of the first terms. By a slight modification of circuit parameters, one can make (11) consistent with (5). That is, if one replaces the last two g_0's in Fig. 6(a) with $g_0 + g_1 + g_2$ and $g_0 + g_2$, respectively, as in Fig. 6(b), then it is consistent with (5). For a general m, one can maintain the consistency of (11) with (5) by replacing the last m g_0's by

$$g_0 + \sum_{p=1}^{m} g_p, \qquad g_0 + \sum_{p=2}^{m} g_p, \cdots, \qquad g_0 + g_m \tag{15}$$

respectively. We will assume, throughout, that this type of modification is always done.

The third issue is that the stability of the spatial dynamics (11) must be carefully defined. That "(11) is stable iff all the eigenvalues of $\boldsymbol{F}$ lie inside the unit circle" does not work because $\boldsymbol{F}$ has a special structure (see (42) below):

if λ is an eigenvalue, so is $1/\lambda$.

Therefore "$|\lambda| < 1$ for all λ" is never satisfied. Since $n = 2K + 1$ is *finite*, another standard definition of stability:

$$\sum_k \|\boldsymbol{y}_k\|^2 < \infty \text{ implies } \sum_k \|\boldsymbol{x}_k\|^2 < \infty \tag{16}$$

does not work either, because (16) is always satisfied. As was shown in Fig. 3(c), $\boldsymbol{x}_k$ can behave in a wild manner even if $n = 2K + 1$ is finite, which is highly undesirable for image processing purposes.

Finally, there is another problem concerning the finiteness of the network size n. Since $\boldsymbol{A}$ and $\boldsymbol{B}$ are symmetric, all eigenvalues are real. Thus, given a fixed n, while it is easy to *say* that (5) is asymptotically stable iff $\boldsymbol{B}^{-1}\boldsymbol{A}$ is negative definite, it is very hard to *derive* analytical (*a priori*) iff conditions for negative definiteness even with $m = 2$. One can derive, however, an interesting analytical condition if one looks for negative definiteness of $\boldsymbol{B}^{-1}\boldsymbol{A}$ *for all* n. Section III gives extremely simple analytical conditions for the temporal stability. With these conditions, a designer is guaranteed to have a stable network independent of the number of nodes. Without these conditions, a designer must compute all the eigenvalues of $\boldsymbol{B}^{-1}\boldsymbol{A}$. If one or more of the eigenvalues turn out to be nonnegative, one has to recompute the eigenvalues with a new trial set of parameter values. One also has to recompute eigenvalues when the network size is changed in response to certain design considerations.

Definition 1: A neural network described by (5) is said to be temporally stable if $\boldsymbol{B}^{-1}\boldsymbol{A}$ is negative definite for all n.

B. Spatial Dynamics

As was explained in subsection A, care needs to be exercised in studying the spatial dynamics (11). Let λ_{s_i}, λ_{c_i}, and λ_{u_i}, be the eigenvalues of $\boldsymbol{F}$ satisfying

$$|\lambda_{s_i}| < 1, \quad |\lambda_{c_i}| = 1 \quad \text{and} \quad |\lambda_{u_i}| > 1$$

respectively, and let E^s, E^c, and E^u be the (generalized) eigenspaces corresponding to λ_{s_i}, λ_{c_i}, and λ_{u_i}, respectively. They are called stable, center, and unstable eigenspaces, respectively. Let $E = \mathbb{R}^{2m}$. Then [16]

$$E = E^s \oplus E^c \oplus E^u \tag{17}$$

where $\oplus$ denotes a direct sum decomposition, and

$$\boldsymbol{F}(E^\alpha) = E^\alpha, \qquad \alpha = s, c, u, \tag{18}$$

i.e., E^s, E^c, and E^u are *invariant* under $\boldsymbol{F}$.

Our task here is to give an appropriate definition of spatial stability while maintaining consistency with (16) when $K \uparrow +\infty$.

Definition 2: A neural network described by (11) is said to be spatially stable if $\boldsymbol{F}$ is *hyperbolic*, i.e., if the center eigenspace E^c in (17) is empty.

Remark 1: Another way of saying this is that all the eigenvalues of $\boldsymbol{F}$ are off the unit circle. Of course, eigenvalues can be outside the unit circle. Note that this definition does not depend on the network size $n = 2K + 1$.

It is known that a noncausal linear system is stable in the sense of (16) iff its transfer function (in the frequency domain) has no poles on the unit circle. This, however, is when $K \uparrow +\infty$ and when there are no boundary conditions. One perhaps wants to argue (as, in fact, the authors did when they initiated the present study) that if the network size is sufficiently large, the behavior would be similar to that of the infinite case. This is simply wrong, as will be indicated by the following examples.

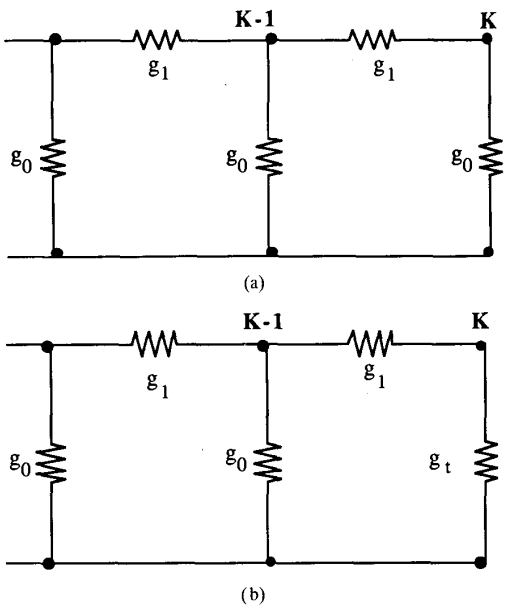


Fig. 7. A network with $m = 1$. (a) Original network. (b) Modified boundary condition, where the rightmost g_0 is replaced by g_t.

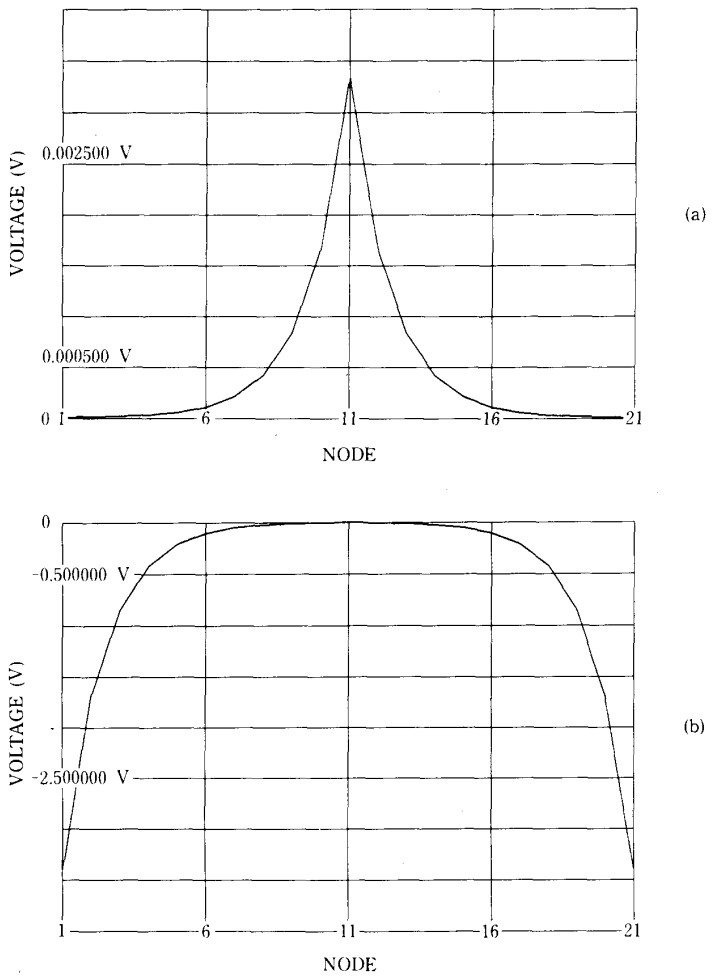


Fig. 8. Significance of boundary conditions. (a) Impulse response for Fig. 7(a) with $g_0 = g$, $g_1 = 2g$, $1/g = 50$ kΩ, $u_{31} = 0.1$ μA. (b) Impulse response for Fig. 7(b) with the same data except for $g_t = -g$, $u_{31} = 0.1$ μA.

Example 1: Consider the simplest case, $m = 1$ in (11) with $g_0 = g$, $g_1 = 2g$, $g > 0$ (Fig. 7(a)). Then

$$\boldsymbol{F} = \begin{bmatrix} 0 & 1 \\ -1 & \frac{5}{2} \end{bmatrix}$$

and $\boldsymbol{F}$ is hyperbolic because eigenvalues are $\lambda_1 = 1/2$ and $\lambda_2 = 2$. Fig. 8(a) shows the impulse response when $1/g = 50$ kΩ, where the impulse is injected at the center node. Let us now replace the rightmost g_0 and the leftmost g_0 with $g_t = -g$ as in Fig. 7(b). The impulse response is then given by Fig. 8(b), which "explodes" in the negative direction as $|k|$ increases. Note the difference of the voltage units. In both cases, the input current injected to the center node is the same and very small: 0.1 μA. It should be emphasized that the only difference is in the two g_t's, and the explosion happens in whichever way the network size is large. In fact, in our simulation with $n = 61$, an overflow occurred.

If the reader says that changing $g_t = g > 0$ to $-g < 0$ is unnatural, the following example shows the case in point.

Example 2: Consider Fig. 7(a) again with $g_0 = g > 0$ and $g_1 = -g/8$. Since eigenvalues of $\boldsymbol{F}$ are $-3 \pm 2\sqrt{2}$, $\boldsymbol{F}$ is hyperbolic, and Fig. 9(a) shows the impulse response with $1/g = 100$ kΩ. Next replace the rightmost and the leftmost g_0 with $g_t = g_0 - \left(g_0^2 + 4g_0g_1\right)^{1/2} = g\left(1 - 1/\sqrt{2}\right) > 0$ $(1/g_t \approx 341$ k$\Omega)$. The impulse response is given by Fig. 9(b), which again explodes. In both cases the input at the center node is 1 μA. Observe that since $g_1 < 0$ the stability issues are already nontrivial with $m = 1$. The stability issues for this example will be checked theoretically in Section III (see Example 5).

There is another story about spatial responses. Our simulation results indicate that spatial responses behave quite properly even if the g_t value is varied by a large amount. Namely, parts (a) and (b) of Fig. 3, Fig. 8(a), and Fig. 9(a) are very robust against variations of g_t from g_0.

Thus, two fundamental questions concerning the spatial dynamics must be answered:

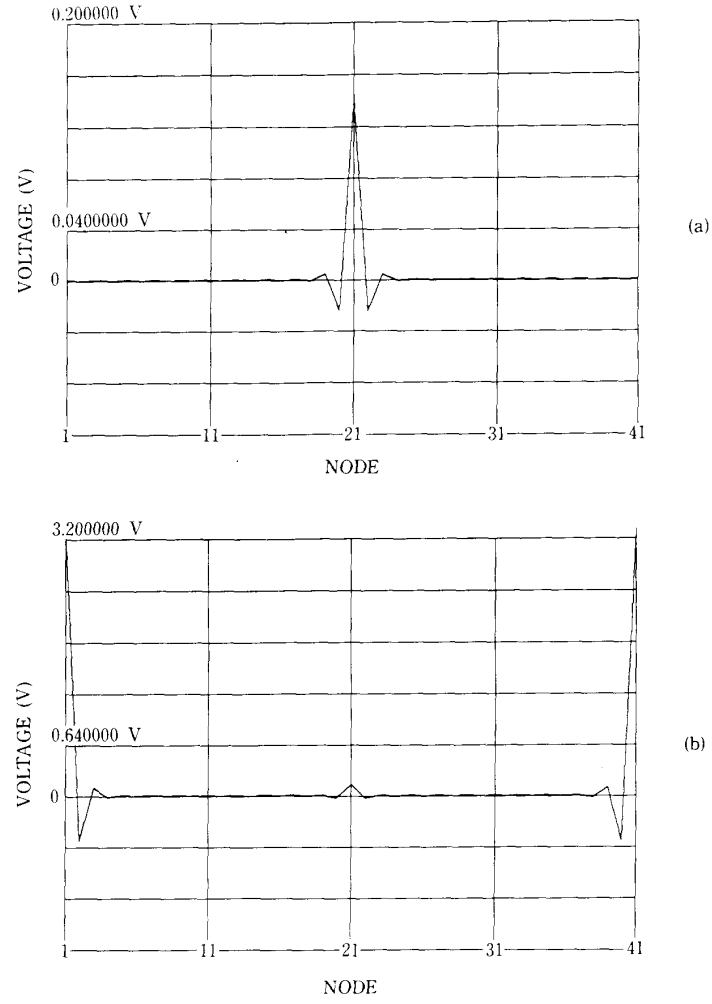


Fig. 9. Impulse response can explode even when $g_t > 0$. (a) Impulse response for Fig. 7(a) with $g_0 = g$, $g_1 = -g/8$, $1/g = 100$ kΩ, $u_{31} = 1$ μA. (b) Impulse response for Fig. 7(b) with the same data except for $g_t = g(1 - 1/\sqrt{2})$.

1) Why does a particular g_t value give rise to explosion of impulse responses even if the eigenvalues are off the unit circle?
2) Why do impulse responses behave properly over a wide range of g_t values?

One can answer the first question easily. Recall (13) and observe that a spatial response $\boldsymbol{x}_k$ depends not only on the input $\boldsymbol{y}_k$ but also on the boundary conditions T_+ and T_-. Therefore, if

$$T_+ = E^u \text{ (resp. } T_- = E^s) \tag{19}$$

then $\boldsymbol{x}_K$ (resp. $\boldsymbol{x}_{-K}$) is *forced* to lie in E^u (resp. E^s). Since E^u (resp. E^s) is invariant under $\boldsymbol{F}$, one has $\boldsymbol{x}_k \in E^u$ (resp. $\boldsymbol{x}_k \in E^s$) for *all* $k > 0$ (resp. $k < 0$); hence

$$\boldsymbol{x}_k = \lambda_2^k \boldsymbol{e}_2, \quad |\lambda_2| > 1, \quad \boldsymbol{e}_2 \in E^u, \quad k > 0$$
$$\left(\text{resp. } \boldsymbol{x}_k = \lambda_1^k \boldsymbol{e}_1, \quad |\lambda_1| < 1, \quad \boldsymbol{e}_1 \in E^s, \quad k < 0\right).$$

This means that $\boldsymbol{x}_k$ explodes as $|k|$ increases. For the network of Example 1 one can easily show that

$$E^u = \{(x_1, x_2) | 2x_1 - x_2 = 0\} \tag{20a}$$
$$E^s = \{(x_1, x_2) | x_1 - 2x_2 = 0\}. \tag{20b}$$

When $g_t = -g$ in Fig. 7(b), KCL at the Kth (resp. $-K$th) node reads $2gv_{K-1} - gv_K = 0$ (resp. $gv_{-K} - 2gv_{-K+1} = 0$), which implies (19). The situation is the same for Example 2. Another way of looking at Fig. 8(b) is to consider Fig. 10, where Fig. 10(a) is the original network and Fig. 10(b) shows that an equivalent conductance, $g_{\text{eq}}(K)$, as seen from node $K - 1$ is

$$g_{\text{eq}}(K) = (1/2g - 1/g)^{-1} = -2g.$$

Since $g + g_{\text{eq}}(K) = -g$, one sees that Fig. 10(b) is equivalent to Fig. 10(c); hence $g_{\text{eq}}(K - 1) = -2g$. It is clear that the equivalent conductance at any node k is $-2g$. This implies that KCL at every node $(k > 0)$ is $2gv_{k-1} - gv_k = 0$ so that

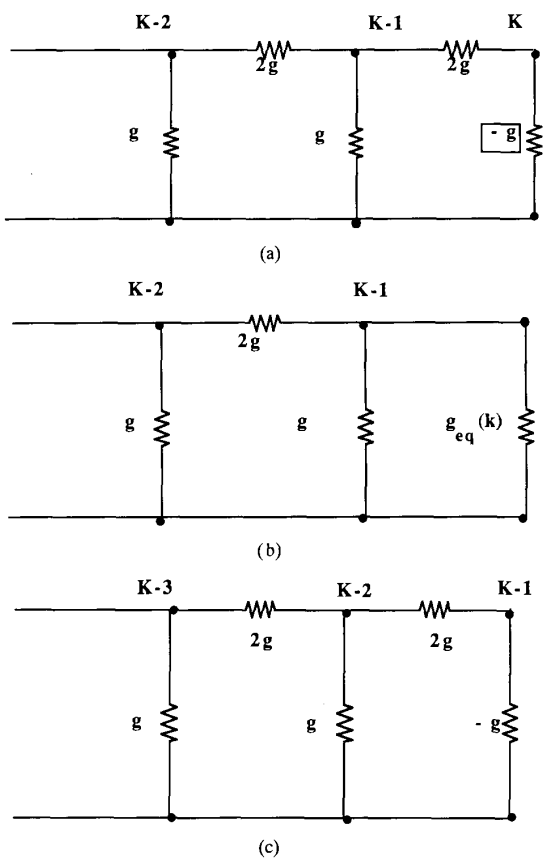


Fig. 10. An equivalent circuit of Fig. 7(b). (a) Original network. (b) The equivalent conductance $g_{eq}(K)$ seen from node $K-1$. (c) A circuit equivalent to Fig. 10(a).

$v_k = 2v_{k-1}$. Thus v_k explodes as $k > 0$ increases. A similar argument shows that v_k, $k < 0$, also explodes as k decreases. The situation in Example 2 is the same.

Answering the second question is much harder. The arguments used in answering the first question cannot be used here. Instead, it exemplifies the difficulty. Observe that KCL at the Kth node in Fig. 7(a) for Example 1 is

$$T_+ : -3gv_K + 2gv_{K-1} = 0$$

and hence

$$T_+ \neq E^u, \qquad T_+ \neq E^s \tag{21a}$$
$$T_- \neq E^u, \qquad T_- \neq E^s. \tag{21b}$$

These facts imply that the response $\boldsymbol{x}_k$ is of the form

$$\boldsymbol{x}_k = \lambda_1^k \boldsymbol{e}_1^+ + \lambda_2^k \boldsymbol{e}_2^+, \qquad k > 0 \tag{22a}$$
$$\boldsymbol{x}_k = \lambda_1^k \boldsymbol{e}_1^- + \lambda_2^k \boldsymbol{e}_2^-, \qquad k < 0 \tag{22b}$$

where $\boldsymbol{e}_1{}^{\pm}$ (resp. $\boldsymbol{e}_2^{\pm}$) are the eigenvectors associated with λ_1 (resp. λ_2) and all of them are *nonzero*. The situation given by (21) does not change for a wide range of g_t variations. This means that there is always an *expanding* term $\lambda_2^k \boldsymbol{e}_2^+$ (resp. $\lambda_1^k \boldsymbol{e}_1^-$) in (22a) (resp. (22b)), in addition to the decaying term $\lambda_1^k \boldsymbol{e}_1^+$ (resp. $\lambda_2^k \boldsymbol{e}_2^-$). This raises another serious question. Consider Example 1 again with $g_t = g > 0$. Since everything is passive, our intuition demands that there should be no stability problems. Nevertheless, (22) says that there are expanding terms.

Thus, another question arises: How can (22) involve expanding terms when everything is passive? In order to answer this, let us first consider the case where the network size is infinite and no boundary conditions are imposed. Let $\{\bar{\boldsymbol{x}}_k\}_{-\infty}^{+\infty}$ be the impulse response defined by

$$\bar{\boldsymbol{x}}_{k+1} = \boldsymbol{F}\bar{\boldsymbol{x}}_k, \qquad k \neq 0$$
$$\bar{\boldsymbol{x}}_1 = \boldsymbol{F}\bar{\boldsymbol{x}}_0 + \boldsymbol{y}_0.$$

Then the network is stable in the sense of (16) only if for every $\boldsymbol{y}_0$

$$\left\|\boldsymbol{F}^k \bar{\boldsymbol{x}}_1\right\| \to 0 \qquad \text{as } k \uparrow +\infty$$
$$\left\|\boldsymbol{F}^k \bar{\boldsymbol{x}}_0\right\| \to 0 \qquad \text{as } k \downarrow -\infty.$$

It will be shown later that this is possible only if E^c, the center eigenspace of $\boldsymbol{F}$, is empty. In order to see distinctions between solutions with and without boundary conditions more precisely, note that in image processing, the input $\{\boldsymbol{y}_k\}$ in (11) is not an impulse, but nonzero for $0 \leq k \leq d$.

Definition 3: Consider (11) and let $\{\boldsymbol{y}_k\}$ be nonzero only for $0 \leq k \leq d$. Then $\{\bar{\boldsymbol{x}}_k\}_{-\infty}^{+\infty}$ is said to be a *free-boundary solution* if

$$\bar{\boldsymbol{x}}_{k+1} = \boldsymbol{F}\bar{\boldsymbol{x}}_k, \qquad k < 0 \tag{23a}$$
$$\bar{\boldsymbol{x}}_d = \boldsymbol{F}^d \bar{\boldsymbol{x}}_0 + \sum_{k=0}^{d-1} \boldsymbol{F}^{d-k}\boldsymbol{y}_k \tag{23b}$$
$$\bar{\boldsymbol{x}}_{k+1} = \boldsymbol{F}\bar{\boldsymbol{x}}_k, \qquad k \geq d. \tag{23c}$$

Remark 2: If $d = 1$, then $\{\boldsymbol{y}_k\}$ is an impulse. If one redefines the summation term in (23b) as a new $\boldsymbol{y}_0$, then (23) can be replaced by

$$\bar{\boldsymbol{x}}_{k+1} = \boldsymbol{F}\bar{\boldsymbol{x}}_k, \qquad k \neq 0 \tag{24a}$$
$$\bar{\boldsymbol{x}}_1 = \boldsymbol{F}^d \bar{\boldsymbol{x}}_0 + \boldsymbol{y}_0. \tag{24b}$$

Since no boundary conditions are imposed, $\{\bar{\boldsymbol{x}}_k\}_{-\infty}^{+\infty}$ is not unique. The following proposition clarifies the uniqueness issue in terms of stability. Let

$$\lambda_{max} := \max\{|\lambda_{si}| \mid \lambda_{si} \text{ is a stable eigenvalue}\}$$
$$\lambda_{min} := \min\{|\lambda_{ui}| \mid \lambda_{ui} \text{ is an unstable eigenvalue}\}$$
$$\lambda_{\#} := \min\left(\lambda_{min}, \lambda_{max}^{-1}\right). \tag{25}$$

Proposition 1:

i) The $\boldsymbol{F}$ matrix of the spatial dynamics is hyperbolic if and only if for any $\boldsymbol{y}_0$ there is a unique free-boundary solution $\{\bar{\boldsymbol{x}}_k\}_{-\infty}^{+\infty}$ satisfying

$$\sum_{k=-\infty}^{+\infty} \|\bar{\boldsymbol{x}}_k\|^2 < \infty. \tag{26}$$

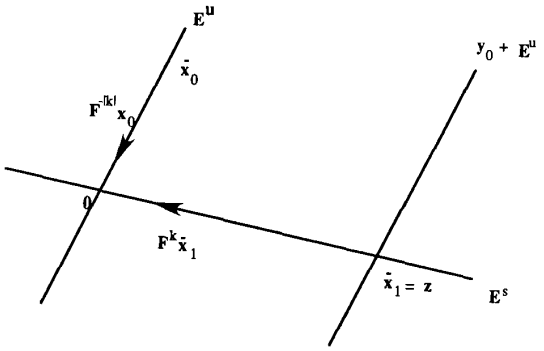


Fig. 11. Definition of z.

ii) The unique $\{\bar{\boldsymbol{x}}_k\}_{-\infty}^{+\infty}$ is determined by

$$\bar{\boldsymbol{x}}_1 \in E^s, \qquad \bar{\boldsymbol{x}}_0 \in E^u, \qquad \bar{\boldsymbol{x}}_1 = F^d \bar{\boldsymbol{x}}_0 + \boldsymbol{y}_0. \tag{27}$$

Proof:

i) $\Rightarrow$) Since E^c is empty (see (17)),

$$E = E^u \oplus E^s.$$

Since E^u is invariant under $\boldsymbol{F}$ (see (18)),

$$\left(\boldsymbol{F}^d E^u + \boldsymbol{y}_0\right) \cap E^s = (E^u + \boldsymbol{y}_0) \cap E^s \tag{28}$$

and this intersection is a singleton set, say $\{\boldsymbol{z}\}$ (Fig. 11). Define

$$\bar{\boldsymbol{x}}_0 := \boldsymbol{F}^{-d}(\boldsymbol{z} - \boldsymbol{y}_0), \qquad \bar{\boldsymbol{x}}_1 := \boldsymbol{z} \tag{29}$$

and let other $\bar{\boldsymbol{x}}_k$ be defined by (24a). Then

$$\begin{aligned} \sum_{k=-\infty}^{+\infty} \|\bar{\boldsymbol{x}}_k\|^2 &= \sum_{k=-\infty}^{0} \|\bar{\boldsymbol{x}}_k\|^2 + \sum_{k=1}^{\infty} \|\bar{\boldsymbol{x}}_k\|^2 \\ &\le \sum_{k=0}^{\infty} \lambda_{\#}^{-2k} \|\bar{\boldsymbol{x}}_0\|^2 + \sum_{k=1}^{\infty} \lambda_{\#}^{-2k} \|\bar{\boldsymbol{x}}_1\|^2 \\ &= \frac{\lambda_{\#}^2}{\lambda_{\#}^2 - 1} \|\bar{\boldsymbol{x}}_0\|^2 + \frac{1}{\lambda_{\#}^2 - 1} \|\bar{\boldsymbol{x}}_1\|^2 < \infty \end{aligned} \tag{30}$$

where $\lambda_{\#}$ is defined by (25). Note that (29) is equivalent to (27) and this is the only choice of $\bar{\boldsymbol{x}}_1$ and $\bar{\boldsymbol{x}}_0$ for which (26) holds, because if $\bar{\boldsymbol{x}}_1 \notin E^s$, for instance, then $\bar{\boldsymbol{x}}_1 = \bar{\boldsymbol{x}}_1^u + \bar{\boldsymbol{x}}_1^s$, with nonzero $\bar{\boldsymbol{x}}_1^u$. Hence $\left\|\boldsymbol{F}^k \bar{\boldsymbol{x}}_1^u\right\| \to \infty$ as $K \uparrow +\infty$. A similar argument holds for $\bar{\boldsymbol{x}}_0$.

$\Leftarrow$) If E^c is non-empty, then there is a $\boldsymbol{y}_0 \neq \boldsymbol{0}$ such that $(E^u + \boldsymbol{y}_0) \cap E^s = \phi$. It is clear that for such $\boldsymbol{y}_0$ there is no way of choosing $\bar{\boldsymbol{x}}_1$ and $\bar{\boldsymbol{x}}_0$ which satisfy (26).

ii) Clearly, (27) and (29) are equivalent. □

Definition 4: The unique $\{\bar{\boldsymbol{x}}_k\}_{-\infty}^{+\infty}$ given in Proposition 1 is said to be the *stable* free-boundary solution.

Remark 3:

i) Consider a free-boundary solution for Example 1, i.e., when $g_0 = g$ and $g_1 = 2g$ extending indefinitely. In spite of the fact that everything is passive, exploding solutions are mathematically legitimate. However, by demanding the finite total energy (26), one forces all exploding solutions to be illegitimate and makes only one solution legitimate, which is given by (27). Conversely, if a unique stable free-boundary solution exists, then the $\boldsymbol{F}$ matrix must satisfy hyperbolicity.

ii) The stable free-boundary solution in terms of (21) can be characterized as $\boldsymbol{e}_2^+ = \boldsymbol{e}_1^- = \boldsymbol{0}$.

Recall the boundary conditions T_+ and T_- in (13).

Definition 5: Let $\{\boldsymbol{y}_k\}$ be nonzero only for $0 \le k \le d$. Then $\{\boldsymbol{x}_k\}_{-K}^{+K}$ is said to be a *solution for* (T_+, T_-, K) if

$$\boldsymbol{x}_{k+1} = \boldsymbol{F}\boldsymbol{x}_k, \qquad -K \le k \le K, \qquad k \neq 0, \tag{31}$$

$$\boldsymbol{x}_1 = \boldsymbol{F}^d \boldsymbol{x}_0 + \boldsymbol{y}_0 \tag{32}$$

$$\boldsymbol{x}_{-K} \in T_-, \qquad \boldsymbol{x}_K \in T_+. \tag{33}$$

The following result thoroughly answers the second and third questions that arose in connection with spatial dynamics in a very general setting.

Theorem 1: Let a neural network described by (11) be spatially stable, i.e., let $\boldsymbol{F}$ be hyperbolic. If the boundary conditions T_+ and T_- satisfy

$$T_+ + E^u = E, \qquad T_- + E^s = E \tag{34}$$

then a solution $\{\boldsymbol{x}_k\}_{-K}^{+K}$ for (T_+, T_-, K) converges to the stable free-boundary solution $\{\bar{\boldsymbol{x}}_k\}_{-\infty}^{+\infty}$ as $K \uparrow +\infty$:

$$\lim_{K \to +\infty} \sum_{k=-K}^{+K} \|\bar{\boldsymbol{x}}_k - \boldsymbol{x}_k\|^2 = 0. \tag{35}$$

Proof: See Appendix I.

Remark 4:

i) In words, this theorem tells us that if the $\boldsymbol{F}$ matrix of the spatial dynamics satisfies the spatial stability condition (Definition 2) and, in addition, if the boundary conditions satisfy (34), then response $\boldsymbol{x}_k$ not only behaves properly but also converges to the stable free-boundary solution $\bar{\boldsymbol{x}}_k$ as $K \uparrow +\infty$.

ii) It will be shown in subsection III-B (see Example 3) that for parts (a) and (b) of Fig. 3, $\boldsymbol{F}$ is hyperbolic while for Fig. 3(c), it is nonhyperbolic. A simple computation shows that there are two distinct pairs of complex conjugate eigenvalues on the unit circle for Fig. 3(c).

iii) Since T_+, T_-, E^u, and E^s all have the same dimension m, the vector sum $+$ in (34) amounts to the same

as the direct sum $\oplus$. Therefore, dim T_+ + dim E^u = dim E and dim T_- + dim E^s = dim E; hence condition (34) is extremely mild. It is satisfied *unless* $\boldsymbol{x}_K$ (resp. $\boldsymbol{x}_{-K}$) is forced to lie in E^u (resp. E^s). This explains why all of our computer simulations look the same with various boundary conditions except for peculiar ones. What happens if T_+ (resp. T_-) is very close to E^u (resp. E^s)? This simply requires a very large K to observe a solution similar to the stable free-boundary solution.

iv) Since E^c is empty, $\boldsymbol{x}_1$defined by (32) can be written as

$$\boldsymbol{x}_1 = \boldsymbol{x}_1^u + \boldsymbol{x}_1^s, \qquad \boldsymbol{x}_1^u \in E^u, \qquad \boldsymbol{x}_1^s \in E^s. \tag{36}$$

A crucial step in the proof of Theorem 1 given in Appendix I is to obtain estimates on $\left\|\boldsymbol{F}^k\boldsymbol{x}_1^u\right\|$, $k \geq 0$, and $\left\|\boldsymbol{F}^k\boldsymbol{x}_0^s\right\|$, $k \leq 0$, because these terms are expanding instead of decaying. The following is roughly what is happening. Let $\{\boldsymbol{x}_k\}_{-K}^{+K}$ be a solution for (T_+, T_-, K), and let $K < K'$ while T_+ and T_- are fixed. In order for $\{\boldsymbol{x}_k\}_{-K'}^{+K'}$ to be a solution for (T_+, T_-, K'), it must make more iterations to reach T_+ from $\boldsymbol{x}_1'$ than that for $\{\boldsymbol{x}_k\}_{-K}^{+K}$. There are two ways to do this. In the first, $\boldsymbol{x}_1'$ locates itself farther away from the origin than $\boldsymbol{x}_1$. In a second, $\boldsymbol{x}_{K'}$ hits T_+ at a point closer to the origin than $\boldsymbol{x}_K$ does (Fig. 12). There is a limitation to the first method because $\boldsymbol{x}_1'$ must satisfy (32) while $\boldsymbol{y}_0$ and d are *fixed.* On the other hand, there is no such limitation to the second method because the dynamics can get as "slow" as it pleases as the origin is approached.[1] This allows one to give an appropriate estimate on $\left\|\boldsymbol{F}^k\boldsymbol{x}_1\right\|$, $k \geq 0$. A similar argument holds for T_-.

v) It is rather interesting to observe that the network given in Example 1 is exactly a D/A converter widely used in practice. See [17] for instance. The network is called the R–$2R$ ladder because $g_0 = g$ and $g_1 = 2g$. In order to convert an n-bit binary signal into an analog signal, one inputs a constant current source at the kth node if the kth bit is "1"; otherwise the current source is set to zero. In such a D/A converter, the rightmost g_0 is replaced with $g_t = 2g$ instead of[2] g so that KCL gives $v_{K-1} - 2v_K = 0$, which forces (see (20b))

$$\boldsymbol{x}_K \in E^s. \tag{37}$$

Since E^s is invariant and since the stable eigenvalue is $1/2$, one has $\boldsymbol{x}_K = \left(1/2^{(K-k)}\right)\boldsymbol{x}_k$. If the leftmost g_t is $2g$ also, then $\boldsymbol{x}_{-K} \in E^u$. Any response of a linear network is a superposition of impulse responses, hence the rightmost voltage v_K, which is the output, is given as

$$v_K = \text{constant} \times \sum_k 1/2^k \tag{38}$$

where k runs over those nodes where "1" is present. Note that if g_t were not chosen as $2g$, the D/A converter would give a wrong analog output.

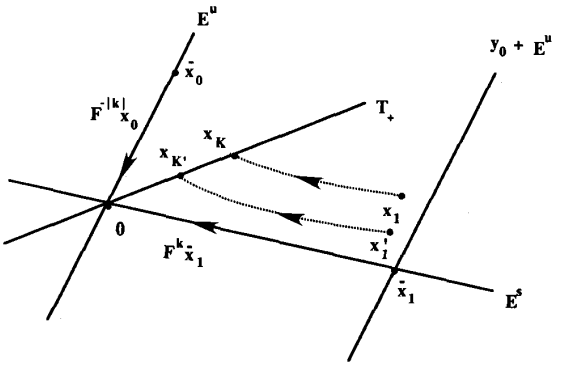


Fig. 12. An illustration of the proof of Theorem 1.

C. Temporal Stability—Spatial Regularity

Now we turn to the relationship between the temporal and the spatial dynamics for which a new concept is needed.

Definition 6: A neural network described by (11) is said to be *spatially regular* if there is a nonsingular $2m \times 2m$ matrix $\boldsymbol{T}$ such that

$$\boldsymbol{TFT}^{-1} = \begin{array}{c} \boldsymbol{E}^s \oplus \boldsymbol{E}^c \oplus \boldsymbol{E}^u \\ \left[\begin{array}{c|c|c} \boldsymbol{F}_s & & \\ \hline & \boldsymbol{F}_c \quad \boldsymbol{G} & \\ \hline & \boldsymbol{F}_c & \boldsymbol{F}_s^{-1} \end{array}\right] \end{array} \tag{39}$$

where a blank indicates a zero matrix, and elements of $\boldsymbol{G}$ consist of $+1$ or 0.

Remark 5: Spatial regularity demands several particular structures in the dynamics:

i) dim E^s = dim E^u and

$$\boldsymbol{F}|E^u = (\boldsymbol{F}|E^s)^{-1} \tag{40}$$

where $\boldsymbol{F}|E^u$ (resp. $\boldsymbol{F}|E^s$) denotes the restriction of $\boldsymbol{F}$ to E^u (resp. E^s). Namely, the dynamics on the unstable eigenspace E^u are exactly the same as the inverse dynamics on the stable eigenspace E^s.

ii) The center eigenspace E^c is decomposed as $E^{c1} \oplus E^{c2}$, dim E^{c1} = dim E^{c2}, and $\boldsymbol{F}|E^{c1}$ and $\boldsymbol{F}|E^{c2}$ have essentially the same structure.

iii) If a neural network described by (11) is spatially stable, E^c is empty. It will be shown later (see (43)) that (40) is satisfied for (10). Therefore, spatial stability *implies* spatial regularity, *but not conversely.*

The following standing assumptions are made throughout the paper unless stated otherwise.

Standing Assumptions: In (5),

(i) $a_0 < 0$, $a_m \neq 0$;

[1] The dynamics $\boldsymbol{x}_{k+1} = \boldsymbol{F}\boldsymbol{x}_k$ have "zero" speed at the origin because $\boldsymbol{F0} = \boldsymbol{0}$; i.e., it does not move. Since a solution depends continuously on its initial condition, one sees that the dynamics gets slower without limit as it approaches the origin.

[2] Recall that in Fig. 8(a) $g_t = g$, while in Fig. 8(b) $g_t = -g$.

(ii) $\boldsymbol{B}$ is positive definite for all n.

Since we are looking for conditions under which $\boldsymbol{B}^{-1}\boldsymbol{A}$ is negative definite for all n, the diagonal element a_0 of $\boldsymbol{A}$ must be negative (provided that $\boldsymbol{B}$ is positive definite), which is the inequality in (i). If $a_m = 0$, then the neighborhood M is of a smaller size. No restrictions will be imposed on the sign of a_p, $p \neq 0$. In image processing neuro chips, c_p in (4) are parasitic capacitors of MOS processes, and positive definiteness of $\boldsymbol{B}$ is a mild condition. The following result establishes a fundamental relationship between the temporal and spatial dynamics.

Theorem 2: A neural network described above is temporally stable *if and only if* it is spatially regular.

Proof: Consider the characteristic polynomial of $\boldsymbol{F}$:

$$P_F(\lambda) := \det(\lambda \mathbf{1} - \boldsymbol{F}) = \lambda^m \left[\frac{a_0}{a_m} + \sum_{p=1}^{m} \frac{a_p}{a_m} \left(\lambda^p + \lambda^{-p} \right) \right] \tag{41}$$

which satisfies

$$P_F(\lambda) = \lambda^{2m} P_F\left(\frac{1}{\lambda} \right). \tag{42}$$

This implies that if λ_s (resp. λ_u) is a stable (resp. unstable) eigenvalue, i.e., $|\lambda_s| < 1$ (resp. $|\lambda_u| > 1$), then λ_s^{-1} (resp. λ_u^{-1}) is also an eigenvalue and unstable (resp. stable). $\boldsymbol{F}$ is nonsingular, for $\det \boldsymbol{F} = 1$; hence there are no zero eigenvalues. This implies that $\dim E^s = \dim E^u$ and

$$\boldsymbol{F}|E^u = (\boldsymbol{F}|E^s)^{-1}. \tag{43}$$

In order to discuss $\boldsymbol{F}|E^c$, let

$$\omega = \lambda + \lambda^{-1} \quad \text{or} \quad \lambda = \frac{1}{2}\left(\omega \pm \sqrt{\omega^2 - 4} \right). \tag{44}$$

By a repeated use of the binomial formula:

$$\lambda^{2p} + \lambda^{-2p} = \omega^{2p} - \sum_{i=1}^{p-1} {}_{2p}C_i \left[\lambda^{2(p-i)} + \lambda^{-2(p-i)} \right] - {}_{2p}C_p$$

$$\lambda^{2p+1} + \lambda^{-(2p+1)} = \omega^{2p+1} - \sum_{i=1}^{p} {}_{2p+1}C_i \left[\lambda^{2(p-i)+1} + \lambda^{-2(p-i)-1} \right]$$

one sees that

$$\frac{a_0}{a_m} + \sum_{p=1}^{m} \frac{a_p}{a_m} \left(\lambda^p + \lambda^{-p} \right) = \sum_{p=0}^{m} \alpha_p \omega^p := Q(\omega) \tag{45}$$

for real α_p's. Since $\boldsymbol{F}$ has no zero eigenvalues,

$$P_F(\lambda) = 0 \quad \text{iff } Q(\omega) = 0 \tag{46}$$

where λ and ω are related via (44). Hence if λ_c is *real* and $|\lambda_c| = 1$, then (44) forces λ_c to be a *double* eigenvalue $\{\lambda_c, \lambda_c\}$ or its *multiple*. We next claim that

$$\dim \ker(\lambda \mathbf{1} - \boldsymbol{F}) = 1 \tag{47}$$

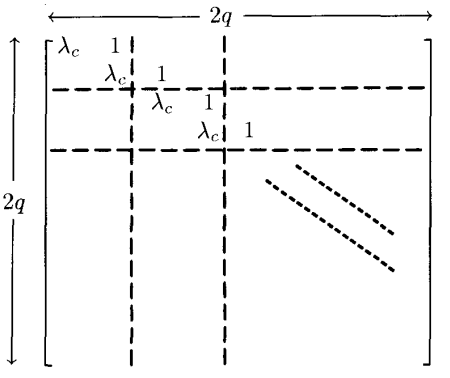
for any eigenvalue λ, where "ker" denotes the kernel of a matrix. In order to see this, note first that λ being an eigenvalue implies

$$\det(\lambda \mathbf{1} - \boldsymbol{F}) = 0.$$

The determinant of the $(2m-1) \times (2m-1)$ principal minor of $\lambda \mathbf{1} - \boldsymbol{F}$ is given by

$$\det \begin{bmatrix} \lambda & 1 & & & \\ & \lambda & 1 & & \\ & & \ddots & \ddots & \\ & & & \lambda & 1 \\ & & & & \lambda \end{bmatrix} = \lambda^{(2m-1)} \neq 0$$

because $\boldsymbol{F}$ has no zero eigenvalues. This shows (47). Thus, for each eigenvalue λ of $\boldsymbol{F}$, there is *only one elementary Jordan block* [16]. Therefore the *real canonical form* of $\boldsymbol{F}|E^{\lambda c}$ restriction of $\boldsymbol{F}$ to the eigenspace corresponding to λ_c, is given by



(48)

where $2q$ is the multiplicity. This is clearly of the form (39).

So far, no use has been made of the negative definiteness of $\boldsymbol{B}^{-1}\boldsymbol{A}$ and yet we are already close to (39), the regularity. The situation, however, is slightly subtle when it comes to a *nonreal* λ_c with $|\lambda_c| = 1$, because (42) tells us nothing except for the fact that λ_c^*, the complex conjugate, is also an eigenvalue. This last is of no use since $\boldsymbol{F}$ is a real matrix and λ_c^* also being an eigenvalue is automatic. We now assume that $\boldsymbol{B}^{-1}\boldsymbol{A}$ is negative definite for all n. Since $\boldsymbol{B}$ is positive definite for all n, $\boldsymbol{A}$ is negative definite for all n. It is known [18], then, that there are $z_p \in \mathbb{R}$, $p = 0, \cdots, m$, such that the elements of $\boldsymbol{A}$ satisfy

$$-a_p = \sum_{i=0}^{m-p} z_i z_{i+p}, \qquad p = 0, \cdots, m \tag{49}$$

i.e., a_p's can be decomposed as in (49). Substitution of (49) into (41) yields

$$P_F(\lambda) = -\frac{\lambda^m}{z_0 z_m} \left[\sum_{i=0}^{m} z_i^2 + \sum_{p=1}^{m} \sum_{i=0}^{m-p} z_i z_{i+p} \left(\lambda^i + \lambda^{-i} \right) \right]$$

$$= -\frac{\lambda^m}{z_0 z_m} \left(\sum_{i=0}^{m} z_i \lambda^{-i} \right) \left(\sum_{i=0}^{m} z_i \lambda^{-i} \right). \tag{50}$$

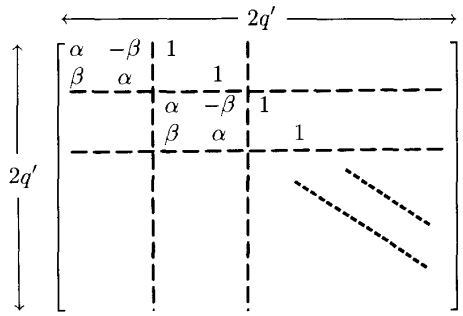
Since $0 \neq a_m = -z_0 z_m$ and since $\boldsymbol{F}$ has no zero eigenvalues, one sees that

$$P_F(\lambda) = 0 \quad \text{iff } R(\lambda) R\left(\frac{1}{\lambda} \right) = 0 \tag{51}$$

where

$$R(\lambda) = \sum_{i=0}^{m} z_i \lambda^i. \tag{52}$$

Therefore if λ is a nonreal eigenvalue with $|\lambda_c| = 1$, (51) forces the eigenvalue configuration to be of the form $\{\lambda_c, \lambda_c^*, \lambda_c, \lambda_c^*\}$ or its *multiple*. It follows from (47) that the real canonical form of $\boldsymbol{F}$ on this eigenspace is given by



(53)

where

$$\alpha^2 + \beta^2 = 1 \tag{54}$$

and $2q'$ is the multiplicity. This, again, is of the form (39).

If a neural network is spatially regular, the real canonical form of the spatial dynamics $\boldsymbol{F}$ is equivalent to (39). The characteristic polynomial of $\boldsymbol{F}$, then, admits a decomposition of the form given by (50). Comparing (50) with (45), one sees that (49) holds. This condition is known [18] to be not only a necessary but also a sufficient condition for $\boldsymbol{A}$ to be negative definite for all n. Since $\boldsymbol{B}$ is positive definite and symmetric for all n, it follows from [19] that

$$\text{max. eigenvalue of } \boldsymbol{B}^{-1}\boldsymbol{A} = \max_{\boldsymbol{v}\neq 0} \frac{\boldsymbol{v}^T\boldsymbol{A}\boldsymbol{v}}{\boldsymbol{v}^T\boldsymbol{B}\boldsymbol{v}} < 0 \tag{55}$$

for any n which implies temporal stability. □

Remark 6: Suppose that a neural network is temporally stable. Although its spatial dynamics can be unstable, it has a sort of symmetry in that the spatial dynamics cannot have a component which is essentially different from the rest; i.e., every component has its partner.

Remark 7:

i) Consider (1) and let

$$W := \sum_{i=1}^{n} v_i u_i$$

which is the *power* injected into the network. It follows from (1) that

$$\begin{aligned} W &= -\sum_i \sum_p v_i a_p v_{i-p} + \sum_i \sum_p v_i b_p \frac{dv_{i-p}}{dt} \\ &= -\boldsymbol{v}^T\boldsymbol{A}\boldsymbol{v} + \boldsymbol{v}^T\boldsymbol{B}\frac{d\boldsymbol{v}}{dt} \\ &:= W_R + W_C. \end{aligned}$$

Thus the first term

$W_R = -\boldsymbol{v}^T\boldsymbol{A}\boldsymbol{v}$ = power dissipated by the resistive part of the network. Therefore a neural network is temporally stable iff its resistive part is *strictly passive,* i.e.,

$$W_R > 0, \qquad \boldsymbol{v} \neq 0 \qquad \text{for all } n.$$

ii) It follows from the previous remark that spatial stability demands more than strict passivity of the resistive part.

iii) Observe that

$\boldsymbol{v}^T\boldsymbol{B}\boldsymbol{v}/2$ = energy stored in the capacitors.

Therefore (55) says that

$$\begin{aligned} &\text{max. eigenvalue of } \boldsymbol{B}^{-1}\boldsymbol{A} \\ &= \max\left(\frac{-\text{power dissipated by resistors}}{2 \cdot \text{ energy stored in capacitors}}\right) \\ &= -\min\left(\frac{\text{power dissipated by resistors}}{2 \cdot \text{ energy stored in capacitators}}\right). \end{aligned}$$

Remark 8: Since the capacitance matrix $\boldsymbol{B}$ has exactly the same structure as that of $\boldsymbol{A}$, one can derive an iff condition for its positive definiteness. If all c_p's are positive, however, then the positive definiteness is straightforward because

$$b_0 = c_0 + 2\sum_{p=1}^{m} c_p > 2\sum_{p=1}^{m} c_p = 2\sum_{p=1}^{m} |b_p| \tag{56}$$

i.e., the diagonal element is larger than the sum of the row elements. Since $\boldsymbol{B}$ is symmetric, this implies positive definiteness.

Remark 9: Since an actual chip is made up of MOS transistors, the formulation given by (1)–(4) is naturally a model. For example, in [1] both the variable conductance g_0 and the negative conductance g_2 are composite CMOS circuits. As one of the reviewers correctly points out, a reasonable justification of the model should be given. Appendix VII supplies a justification.

Now the question naturally arises as to how one checks temporal stability or spatial regularity. Since temporal stability is equivalent to spatial regularity, we will say, hereafter, that the *stability–regularity* condition is satisfied if a network is temporally stable or spatially regular. Recall $Q(\omega)$ defined by (45).

Proposition 2: The following are equivalent:

i) Stability–regularity.

ii) Every nonreal eigenvalue λ_c of $\boldsymbol{F}$ with $|\lambda_c| = 1$ has an even multiplicity.

iii) Every real zero ω_R of Q with $|\omega_R| < 2$ has an even multiplicity.

Proof: Equivalence between (i) and (ii) was demonstrated in the proof of Theorem 2. To show that (ii) and (iii), suppose that $\lambda_c = e^{j\theta}$, $\theta \neq k\pi$, is an eigenvalue of $\boldsymbol{F}$. Then (44) implies that the corresponding ω is real and $|\omega| < 2$. Conversely, if ω is real and $|\omega| < 2$, then (44) says that $\lambda_c = e^{\pm j\theta}$, $\theta \neq k\pi$. □

For the sake of the completeness, we will state the following:

Proposition 3: The following are equivalent:

i) Spatial stability.

ii) Eigenvalues of $\boldsymbol{F}$ are off the unit circle.
iii) Q has no real zero on $[-2, 2]$.

III. Explicit Stability Criteria

Even though both conditions (ii) and (iii) of Proposition 2 give a specific way of checking the stability–regularity, explicit analytical conditions in terms of the circuit parameters greatly help in designing circuits. The same is true for the spatial stability. In subsection III-A two stability indicator functions will be given for a general m, with which one can easily check the stability–regularity or the spatial stability in terms of circuit parameters. In subsections B through D, the stability indicator functions will be specialized to $m \leq 3$. In particular, it will be shown that the conductance values of the neuro chip which motivated the present study satisfy the temporal as well as the spatial stability conditions. Furthermore, it will be rigorously shown why our numerical experiments indicated the "equivalence" between the temporal and the spatial stability.

A. Stability Indicator Functions

The following functions play a crucial role throughout the rest of the paper and will be called the stability indicator functions:

$$\begin{aligned} \sigma_+(a_0, a_1, \cdots, a_m) &:= \max_{\omega \in [-2,2]} a_m Q(\omega) \\ \sigma_-(a_0, a_1, \cdots, a_m) &:= \min_{\omega \in [-2,2]} a_m Q(\omega) \end{aligned} \tag{57}$$

where Q is defined by (45).

Proposition 4: A neural network described by (5) and (11) satisfies the stability–regularity condition *if and only if*

$$\sigma_+(a_0, a_1, \cdots, a_m) \leq 0. \tag{58}$$

Proof: It follows from Proposition 2 that the stability–regularity holds iff every real zero of Q on $(-2, 2)$ has an *even* multiplicity. This means that, if Q has a zero on $(-2, 2)$, it must be an *extremum*. Since any zero at ± 2 is necessarily even (see (44)), one sees that the stability—regularity is equivalent to

$$\max_{\omega \in [-2,2]} Q(\omega) \leq 0 \quad \text{or} \quad \min_{\omega \in [-2,2]} Q(\omega) \geq 0. \tag{59}$$

One can easily show that (59) is equivalent to

$$\max_{\omega \in [-2,2]} a_m Q(\omega) \leq 0 \quad \text{or} \quad \min_{\omega \in [-2,2]} a_m Q(\omega) \geq 0. \tag{60}$$

We claim that the second inequality in (60) is always violated under our standing assumptions: $a_0 < 0$, $a_m \neq 0$. In order to show this, consider (see (45))

$$a_m Q(\omega) = a_0 + \sum_{p=1}^{m} a_p \left(\lambda^p + \lambda^{-p}\right) \tag{61}$$

where ω and λ are related via (44). Since we are interested in ω on $[-2, 2]$, λ is represented as

$$\lambda = e^{j\theta}, \qquad \theta \in [0, \pi].$$

Hence

$$\begin{aligned} \sum_{p=1}^{m} a_p \left(\lambda^p + \lambda^{-p}\right) &= \sum_{p=1}^{m} a_p \left(e^{jp\theta} + e^{-jp\theta}\right) \\ &= 2 \sum_{p=1}^{m} a_p \cos(p\theta). \end{aligned} \tag{62}$$

It follows from

$$\int_0^{\pi} 2 \sum_{p=1}^{m} a_p \cos(p\theta)\, d\theta = 0 \tag{63}$$

that (62) is either identically zero or changes sign on $[0, \pi]$. Since $a_m \neq 0$, the first possibility is excluded. Therefore, if $a_0 < 0$, then (61) cannot be always positive on $[-2, 2]$; hence the second inequality in (60) is always violated. □

Proposition 5: A neural network described by (11) is spatially stable *if and only if*

$$\sigma_+(a_0, a_1, \cdots, a_m) < 0. \tag{64}$$

Proof: It follows from Proposition 3 that the spatial stability is equivalent to the fact that Q has no real zero on $[-2, 2]$, which, in turn, is equivalent to

$$\max_{\omega \in [-2,2]} a_m Q(\omega) < 0 \quad \text{or} \quad \min_{\omega \in [-2,2]} a_m Q(\omega) > 0. \tag{65}$$

By using the argument used in the proof of Proposition 4, one sees that the second inequality in (65) is always violated. □

The following fact gives upper and lower bounds for eigenvalues of the temporal dynamics $\boldsymbol{A}$.

Proposition 6:

i) Any eigenvalue μ of the temporal dynamics $\boldsymbol{A}$ for any n satisfies the following bounds:

$$\sigma_-(a_0, a_1, \cdots, a_m) < \mu < \sigma_+(a_0, a_1, \cdots, a_m). \tag{66}$$

ii) The bounds (66) are *optimal* in the sense that if σ_+^* (resp. σ_-^*) is any number which satisfies

$$\begin{aligned} &\sigma_+^* < \sigma_+(a_0, a_1, \cdots, a_m) \\ &\left(\text{resp. } \sigma_-(a_0, a_1, \cdots, a_m) < \sigma_-^*\right) \end{aligned}$$

then there is an eigenvalue μ of $\boldsymbol{A}$ for some n such that

$$\sigma_+^* < \mu \qquad \left(\text{resp. } \mu < \sigma_-^*\right).$$

Proof: See Appendix II.

Remark 10: Note that (58) is a weak inequality, i.e., equality is allowed, while (66) does not allow the equality. This is exactly what it should be. If, for instance $\sigma_+(a_0, a_1, \cdots, a_m) = 0$, then Proposition 4 tells us that the network is temporally

stable and hence that all the eigenvalues of $\boldsymbol{A}$ are strictly negative, which is what (66) says.

We would like to emphasize the *if and only if* nature of Proposition 4 as well as Proposition 5 and the *optimality* of Proposition 6, which indicate that $\sigma_+(a_0, a_1, \cdots, a_m)$ and $\sigma_-(a_0, a_1, \cdots, a_m)$ are crucial to the stability issues of our interest. The above propositions, however, would not be very useful unless one could compute explicit formulas for $\sigma_+(a_0, a_1, \cdots, a_m)$ and $\sigma_-(a_0, a_1, \cdots, a_m)$. In the following, we will compute these functions for $m \leq 3$.

B. $m = 2$

We begin with $m = 2$, which motivated the present study.

Proposition 7: When $m = 2$, the stability indicator functions are given by

$$\sigma_+(g_0, g_1, g_2) = \begin{cases} -g_0 - 2g_1 + 2|g_1| & \text{when } g_2 > 0 \text{ or } g_2 < 0 \text{ and } |g_1/g_2| \geq 4 \\ -g_0 - 2g_1 - 4g_2 - g_1^2/4g_2 & \text{when } g_2 < 0 \text{ and } |g_1/g_2| \leq 4 \end{cases}$$

$$\sigma_-(g_0, g_1, g_2) = \begin{cases} -g_0 - 2g_1 - 2|g_1| & \text{when } g_2 < 0 \text{ or } g_2 > 0 \text{ and } |g_1/g_2| \geq 4 \\ -g_0 - 2g_1 - 4g_2 - g_1^2/4g_2 & \text{when } g_2 > 0 \text{ and } |g_1/g_2| \leq 4. \end{cases} \tag{67}$$

Proof: See Appendix III.

Example 3: With Propositions 4–7 at hand, we can now check Fig. 3 and Fig. 4 theoretically. In Figs. 3 and 4, $1/g_0 = 200\ \text{k}\Omega$ and $1/g_1 = 5\ \text{k}\Omega$ are fixed while g_2 is varied: (a) $1/g_2 = -20\ \text{k}\Omega$; (b) $1/g_2 = -18\ \text{k}\Omega$; and (c) $1/g_2 = -17\ \text{k}\Omega$. In order to check (a), note that $|g_1/g_2| = 4$; hence (67) gives

$$\sigma_+(g_0, g_1, g_2) = -g_0 < 0.$$

Propositions 4 and 5 guarantee the temporal as well as the spatial stability. For (b), $|g_1/g_2| = 18/5 < 4$ and (67) reads

$$\begin{aligned} \sigma_+(g_0, g_1, g_2) &= -g_0 - 2g_1 - 4g_2 - g_1^2/4g_2 \\ &= (-1/200 - 2/5 + 4/18 + 18/100) \\ &\quad \times 10^{-3} < 0 \end{aligned}$$

which checks Fig. 3(b) and Fig. 4(b). Finally, for (c),

$$\begin{aligned} \sigma_+(g_0, g_1, g_2) &= (-1/200 - 2/5 + 4/17 + 17/100) \\ &\quad \times 10^{-3} > 0 \end{aligned}$$

and hence the network is temporally and spatially unstable, which checks Fig. 3(c) and Fig. 4(c).

Example 4: For the Gaussian-like convolver [1]

$$g_1 > 0, \qquad g_2 < 0, \qquad g_1 = 4|g_2|. \tag{68}$$

(Appendix IV gives a simple explanation for this choice of conductance values.) Propositions 4 and 7 tell us that the stability–regularity is equivalent to

$$\sigma_+(g_0, g_1, g_2) = -g_0 \leq 0,$$

i.e., passivity of g_0. Furthermore, Proposition 5 says that the network is spatially stable iff

$$\sigma_+(g_0, g_1, g_2) = -g_0 < 0,$$

i.e., iff g_0 is strictly passive. Thus g_0 can be safely varied over any range as long as it is positive.

Remark 11:

i) Even when g_1 as well as g_2 is negative, a network can satisfy the stability–regularity or/and the spatial stability condition provided that g_0 is "sufficiently" passive because

$$\sigma_+(g_0, g_1, g_2) = \begin{cases} -g_0 + 4|g_1| & \text{when } |g_1/g_2| \geq 4 \\ -g_0 + 2|g_1| + 4|g_2| + g_1^2/4|g_2| & \text{when } |g_1/g_2| \leq 4. \end{cases}$$

ii) If $g_2 > 0$, then

$$\sigma_+(g_0, g_1, g_2) = \begin{cases} -g_0 & \text{when } g_1 \geq 0 \\ -g_0 + 4|g_1| & \text{when } g_1 \leq 0. \end{cases}$$

iii) Since Q is quadratic, conditions (ii) and (iii) of Proposition 2 are sharpened, respectively to the following:

(ii)′ $\boldsymbol{F}$ has no simple nonreal eigenvalue on the unit circle.

(iii)′ Q has no real zero on $(-2, 2)$.

It follows from Proposition 4 (resp. Proposition 6) that the set of parameter values (g_0, g_1, g_2) for which stability–regularity and the spatial stability hold are given, respectively, by

$$\text{SR} = \{(g_0, g_1, g_2) | \sigma_+(g_0, g_1, g_2) \leq 0,\ g_0 + 2g_1 + 2g_2 > 0\} \tag{69}$$

$$\text{SS} = \{(g_0, g_1, g_2) | \sigma_+(g_0, g_1, g_2) < 0,\ g_0 + 2g_1 + 2g_2 > 0\}. \tag{70}$$

We will now give a fact which, as its by-product, explains why our numerical experiments suggested SR = SS, which is untrue. Let

$$G = \{(g_0, g_1, g_2) | g_2 < 0\}$$

on which our numerical experiments were performed.

Proposition 8:

i) $\text{meas}[\text{SS} \cap G] > 0$

ii) $\text{meas}[(\text{SR} - \text{SS}) \cap G] = 0$

where meas[·] denotes the Lebesgue measure on $\mathbb{R}^3$.

Proof: It follows from (67) that SS $\cap$ G contains an open set of $\mathbb{R}^3$ and hence it is of positive Lebesgue measure. Since SR $\supset$ SS, the set difference SR − SS makes sense, and

meas$[\mathrm{SR} \cap G] > 0$. The set $(\mathrm{SR} - \mathrm{SS}) \cap G$ is a subset of $\mathrm{SR} \cap G$ such that

$$\sigma_+(g_0, g_1, g_2) = 0. \tag{71}$$

Since the gradient of $\sigma_+(g_0, g_1, g_2)$ on $(\mathrm{SR} - \mathrm{SS}) \cap G$ is given by

$$D\sigma_+(g_0, g_1, g_2) = \begin{cases} (-1, 0, 0) & \text{when } g_1 > 0,\ |g_1/g_2| \geq 4 \\ (-1, -4, 0) & \text{when } g_1 < 0,\ |g_1/g_2| \geq 4 \\ (-1, -2 - g_1/2g_2, -4 + g_1^2/4g_2^2) & \text{when } |g_1/g_2| \geq 4 \end{cases}$$

and since this is nonvanishing, (71) forces (g_0, g_1, g_2) to lie in a Lebesgue measure zero subset [20, lemma 4]. □

Remark 12: This proposition explains why our experiments suggested SR = SS for a Lebesgue measure zero subset is "hard to hit."

C. $m = 1$

Neural networks with $m = 1$ are used in an extensive manner [6]–[8]. Although those networks contain only positive conductances $(g_0, g_1 > 0)$, it would be worth clarifying the temporal as well as the spatial stability issues when $g_1 < 0$. We will state the result without proof because the proof is much simpler than in the $m = 2$ case.

Proposition 9: When $m = 1$, the stability indicators are given by

$$\sigma_+(g_0, g_1) = -g_0 - 2g_1 + 2|g_1|$$
$$\sigma_-(g_0, g_1) = -g_0 - 2g_1 - 2|g_1|.$$

Example 5: When $g_0 > 0$ but $g_1 < 0$, the stability issues are nontrivial. The network is temporally (resp. spatially) stable iff

$$-g_0 + 4|g_1| \leq 0 \qquad (\text{resp.} - g_0 + 4|g_1| < 0).$$

In Example 2, $1/g_0 = 100\ \mathrm{k\Omega}$, $1/g_1 = -800\ \mathrm{k\Omega}$, and $\sigma_+(g_0, g_1) = (-1/100 + 4/800) \times 10^{-3} < 0$ and the network is temporally as well as spatially stable which checks Fig. 9(a).

Remark 13: One can show for this case also that the set of (g_0, g_1) values on which the temporal stability holds, and yet the spatial stability fails, is of measure zero,

D. $m = 3$

As was remarked earlier, neurochips with $m \leq 2$ have already been designed and fabricated. Although no result has been reported on chips with $m = 3$, we conjecture that this architecture might be suitable for noncausual IIR implementations of interesting image processing filters.

We saw in subsection III-B and Appendix III that the case $m = 2$ is already sufficiently complicated to require a careful analysis. Naturally, the case $m = 3$ is even more involved and we need to prepare several notations. First note that, when $m = 3$,

$$Q(\omega) = \left(\frac{a_0}{a_3} - 2\frac{a_2}{a_3}\right) + \left(\frac{a_1}{a_3} - 3\right)\omega + \frac{a_2}{a_3}\omega^2 + \omega^3 \tag{72}$$

and that

$$a_0 = -(g_0 + 2g_1 + 2g_2 + 2g_3), \qquad a_1 = g_1, \qquad a_2 = g_2, \qquad a_3 = g_3. \tag{73}$$

The zeros of the derivative $dQ/d\omega$ are

$$\xi_\pm = \left(-\frac{a_2}{a_3} \pm \sqrt{D}\right)/3 = \left(-\frac{g_2}{g_3} \pm \sqrt{D}\right)/3 \tag{74}$$

where

$$D = \left(\frac{a_2}{a_3}\right)^2 - 3\left(\frac{a_1}{a_3}\right) + 9 = \left(\frac{g_2}{g_3}\right)^2 - 3\left(\frac{g_1}{g_3}\right) + 9. \tag{75}$$

Using (74), one has

$$\begin{aligned} Q(\xi_\pm) &= 3\xi_\pm\left[-\frac{2}{9}\left(\frac{a_2}{a_3}\right)^2 + \frac{2}{3}\left(\frac{a_1}{a_3}\right) - 2\right] \\ &\quad - \frac{1}{9}\frac{a_2}{a_3}\frac{a_1}{a_3} - \frac{5}{3}\frac{a_2}{a_3} + \frac{a_0}{a_3} \\ &= 3\xi_\pm\left[-\frac{2}{9}\left(\frac{g_2}{g_3}\right)^2 + \frac{2}{3}\left(\frac{g_1}{g_3}\right) - 2\right] - \frac{1}{9}\frac{g_2 g_1}{g_3^2} \\ &\quad - \frac{g_0}{g_3} - \frac{2g_1}{g_3} - \frac{11}{3}\frac{g_2}{g_3} - 2. \end{aligned} \tag{76}$$

Note that

$$Q(2) = \frac{a_0}{a_3} + 2\frac{a_1}{a_3} + 2\frac{a_2}{a_3} + 2 = -\frac{g_0}{g_3}$$
$$Q(-2) = \frac{a_0}{a_3} - 2\frac{a_1}{a_3} + 2\frac{a_2}{a_3} - 2 = -\frac{g_0}{g_3} - 4\frac{g_1}{g_3} - 4$$

and define

$$f_+ := a_3 Q(2) = -g_0 \tag{77}$$
$$f_- := a_3 Q(-2) = -g_0 - 4g_1 - 4g_3 \tag{78}$$

$$\begin{aligned} h_\pm &:= a_3 Q(\xi_\pm) \\ &= 3\xi_\pm\left[-\frac{2}{9}\frac{g_2^2}{g_3} + \frac{2}{3}g_1 - 2g_3\right] - \frac{1}{9}\frac{g_2 g_1}{g_3} \\ &\quad - g_0 - 2g_1 - \frac{11}{3}g_2 - 2g_3. \end{aligned} \tag{79}$$

Proposition 10: When $m = 3$, the stability indicator functions are given by

$$\sigma_+(g_0,g_1,g_2,g_3)=\begin{cases} f_+ & \text{when } g_3>0,\ D\le 0 \quad \text{or} \quad g_3>0,\ D>0,\ \xi_-<\xi_+\le -2 \\ & \text{or} \quad g_3>0,\ D>0,\ 2\le\xi_-<\xi_+ \\ & \text{or} \quad g_3<0,\ D>0,\ \xi_-\le -2,\ 2\le\xi_+ \\ f_- & \text{when } g_3>0,\ D>0,\ \xi_-\le -2,\ 2\le\xi_+ \\ & \text{or} \quad g_3<0,\ D\le 0 \quad \text{or} \quad g_3<0,\ D>0,\ \xi_-<\xi_+\le -2 \\ & \text{or} \quad g_3<0,\ D>0,\ 2\le\xi_-<\xi_+ \\ \max[f_+,f_-] & \text{when } g_3>0,\ \xi_-\le -2\le\xi_+\le 2 \\ & \text{or} \quad g_3<0,\ -2\le\xi_-\le 2\le\xi_+ \\ h_+ & \text{when } g_3>0,\ \xi_-\le -2\le\xi_+\le 2 \\ h_- & \text{when } g_3>0,\ -2\le\xi_-\le 2\le\xi_+ \\ \max[f_+,h_-] & \text{when } g_3>0,\ -2\le\xi_-<\xi_+\le 2 \\ \max[f_-,h_+] & \text{when } g_3<0,\ -2\le\xi_-\le\xi_+\le 2 \end{cases}$$

$$\sigma_-(g_0,g_1,g_2,g_3)=\begin{cases} f_+ & \text{when } g_3>0,\ D>0,\ \xi_-\le -2,\ 2\le\xi_+ \quad \text{or} \quad g_3<0,\ D\le 0 \\ & \text{or} \quad g_3<0,\ D>0,\ \xi_-<\xi_+\le -2 \\ & \text{or} \quad g_3<0,\ D>0,\ 2\le\xi_-<\xi_+ \\ f_- & \text{when } g_3>0,\ D\le 0, \quad \text{or} \quad g_3>0,\ D>0,\ \xi_-<\xi_+\le -2 \\ & \text{or} \quad g_3>0,\ D>0,\ 2\le\xi_-<\xi_+ \\ & \text{or} \quad g_3<0, D>0, \xi_-<-2, 2\le\xi_+ \\ \min[f_+,f_-] & \text{when } g_3<0,\ -2\le\xi_-\le 2\le\xi_+ \\ & \text{or} \quad g_3>0,\ \xi_-\le -2\le\xi_+\le 2 \\ h_+ & \text{when } g_3>0,\ -2\le\xi_-\le 2\le\xi_+ \\ h_- & \text{when } g_3<0,\ \xi_-\le -2\le\xi_+\le 2 \\ \min[f_+,h_-] & \text{when } g_3<0,\ -2\le\xi_-\le\xi_+\le 2 \\ \min[f_-,h_+] & \text{when } g_3>0,\ -2\le\xi_-<\xi_+\le 2. \end{cases}$$

Proof: See Appendix V.

IV. Transients

This section analyzes the capacitance matrix $\boldsymbol{B}$ in (4) using the method used for analyzing $\boldsymbol{A}$. As a by-product, an estimate will be obtained of the "processing speed" of neuro chips.

It follows from (4) that the capacitance matrix $\boldsymbol{B}$ has exactly the same structure as that of $\boldsymbol{A}$. Therefore, one can derive conditions under which $\boldsymbol{B}$ is *positive* definite and bounds on its eigenvalues. Let

$$P_B(\lambda) := \lambda^m\left[\frac{b_0}{b_m}+\sum_{p=1}^{m}\frac{b_p}{b_m}\left(\lambda^p+\lambda^{-p}\right)\right]$$

$$Q_B(\omega) := \frac{b_0}{b_m}+\sum_{p=1}^{m}\frac{b_p}{b_m}\left(\lambda^p+\lambda^{-p}\right)$$

where $b_0,\cdots,b_m$ are as in (4) while ω and λ are as in (44). Define

$$\eta_+(b_0,b_1,\cdots,b_m) := \max_{\omega\in[-2,2]} b_m Q_B(\omega) \tag{80}$$

$$\eta_-(b_0,b_1,\cdots,b_m) := \min_{\omega\in[-2,2]} b_m Q_B(\omega) \tag{81}$$

The following fact can be proved by an argument similar to that used for the negative definiteness of $\boldsymbol{A}$.

Proposition 11: Replace (ii) of the standing assumptions (subsection II-C) by

$$b_0>0, \qquad b_m\neq 0. \tag{82}$$

i) The following conditions are equivalent:

a) $\boldsymbol{B}$ is positive definite for all n.

b) Every nonreal zero λ_c of $P_B(\lambda)$ with $|\lambda_c|=1$ has an even multiplicity.

c) Every real zero ω_R of $Q_B(\omega)$ with $|\omega_R|<2$ has an even multiplicity.

d)

$$\eta_-(b_0,b_1,\cdots,b_m)\ge 0 \tag{83}$$

ii) Any eigenvalue ν of $\boldsymbol{B}$ for any n, satisfies

$$\eta_-(b_0,b_1,\cdots,b_m)<\nu<\eta_+(b_0,b_1,\cdots,b_m) \tag{84}$$

and the bounds are optimal.

Corollary 1: Assume (82) and consider the temporal dynamics (5) with $\boldsymbol{v}(0)=\boldsymbol{0}$. If (58) and (83) are satisfied, then the solution $\boldsymbol{v}(t)$ of (5) satisfies the following bounds:

$$\frac{\eta_-}{\sigma_-}\left[\exp\left(\frac{\sigma_-}{\eta_-}t\right)-1\right]\left\|\boldsymbol{B}^{-1}\boldsymbol{u}\right\| \le \|\boldsymbol{v}(t)\| \le \frac{\eta_+}{\sigma_+}\left[\exp\left(\frac{\sigma_+}{\eta_+}t\right)-1\right]\left\|\boldsymbol{B}^{-1}\boldsymbol{u}\right\| \tag{85}$$

Proof: See Appendix VI.

Remark 14:

i) The result tells us how fast/slow a step response of (5) grows. Although there is no precise concept of the

time constant RC for (5) ($\dim \boldsymbol{v} \gg 1$), (85) can be interpreted as

$$-\frac{\eta_-}{\sigma_-} \leq \text{“time constant”} \leq -\frac{\eta_+}{\sigma_+}. \tag{86}$$

ii) Let us compute the upper bound in (86) for $m = 2$. It is not difficult to show that

$$\eta_+(c_0, c_1, c_2) = \begin{cases} c_0 + 2c_1 + 2|c_1| & \text{when } c_2 < 0 \text{ or } c_2 > 0 \text{ and } |c_1/c_2| \geq 4 \\ c_0 + 2c_1 + 4c_2 + c_1^2/4c_2 & \text{when } c_2 > 0 \text{ and } |c_1/c_2| \leq 4. \end{cases}$$

If $g_0, g_1, c_0, c_1, c_2 > 0$, then it follows from (67) and the above formula that

$$-\frac{\eta_+}{\sigma_+} = -\frac{\eta_+}{g_0} = \begin{cases} (c_0 + 4c_1)/g_0 & \text{when } |c_1/c_2| \geq 4 \\ (c_0 + 2c_1 + 4c_2 + c_1^2/4c_2)/g_0 & \text{when } |c_1/c_2| \leq 4. \end{cases}$$

Since it is difficult to estimate parasitic capacitances accurately, this is as much as one can tell from the corollary.

V. Concluding Remarks

(i) We would like to call the reader's attention to the fact that the spatial dynamics of the class of neural networks discussed here are *zero phase* and yet IIR. More specifically, consider the transfer function of the spatial dynamics in the frequency domain:

$$1/H(z) = 1/\left[a_m z^{-m} P_F(z)\right]$$

where P_F is the characteristic polynomial defined by (41). Then

$$H\left(e^{j\omega}\right) = a_0 + \sum_{p=1}^{m} 2a_p \cos(p\omega)$$

which is *real*. Obviously, a zero-phase filter is ideal in signal processing, for if the phase does not behave properly, the signal would be distorted. It is known [21] that a stable linear-phase IIR *cannot* be realized by a *causal* system (linear-phase meaning here that the phase is linear in ω). Thus the spatial dynamics (11) are a zero-phase noncausal IIR filter. The results reported here establish conditions under which those noncausal IIR filters are temporally and/or spatially stable.

(ii) Using an argument used in the proof of Lemma A2, one can show that if $a_0 > 0$, i.e., if the diagonal element of $\boldsymbol{A}$ is positive, then $\boldsymbol{A}$ is *positive* definite iff the spatial dynamics is regular. Since the definition of spatial stability (Definition 2) is the hyperbolicity of $\boldsymbol{F}$, the spatial dynamics can be stable even when $a_0 > 0$. Thus, the spatial regularity or stability can be satisfied even when $\boldsymbol{A}$ is positive definite, while temporal stability is certainly violated if $\boldsymbol{A}$ is positive definite. This *asymmetry* is due to the fact that the spatial dynamics is noncausal whereas the temporal dynamics is causal.

(iii) Recall Proposition 8, which states that, for $m = 2$, the temporal stability coincides with the spatial stability except for a measure zero subset of $\mathbb{R}^3$.

Conjecture: Proposition 8 will be true for a general m.

(iv) The following is a list of possible future research projects:

a) Generalizations to nonlinear cases, e.g., the chip reported in [11]. While the temporal stability results can be established under reasonable conditions, the spatial stability results may not be easy to obtain because the spatial dynamics are not only nonlinear but also nonautonomous with respect to node number k. More specifically, let

$$\boldsymbol{B}\frac{d\boldsymbol{v}}{dt} = \boldsymbol{G}(\boldsymbol{v}) + \boldsymbol{u} \tag{87}$$

be the temporal dynamics where $\boldsymbol{G} : \mathbb{R}^n \to \mathbb{R}^n$. Let $\boldsymbol{v}$ be an equilibrium of (87) and suppose that

$$\boldsymbol{x}_{k+1} = \boldsymbol{F}(\boldsymbol{x}_k) + \boldsymbol{y}_k \tag{88}$$

represents the spatial dynamics in the sense of (11), where $\boldsymbol{F} : \mathbb{R}^{2m} \to \mathbb{R}^{2m}$. Therefore, the spatial stability means the stability of the trajectory (88), which is not necessarily a fixed point of $\boldsymbol{F}$. Furthermore, if the conductances are nonlinear, the temporal dynamics are not necessarily of the popular form

$$\boldsymbol{B}\frac{d\boldsymbol{v}}{dt} = -\frac{1}{R}\boldsymbol{v} + \boldsymbol{T}\boldsymbol{G}(v) + \boldsymbol{u}$$

where $\boldsymbol{T}$ is symmetric, $\boldsymbol{G} = \left(G^1, \cdots, G^n\right)$, and G^i, $i = 1, \cdots, n$, is sigmoidal.

b) Generalization to two-dimensional array cases.

c) IIR implementations and associated stability of other interesting filters, e.g., oriented receptive field filters [4] and Gabor filters [5].

d) It could be interesting to investigate the relationship, if any, with the stability results for neural field equations [22], [23].

Appendix I
Proof of Theorem 1

Throughout this appendix, the center eigenspace E^c is empty. Hence any vector $\boldsymbol{x} \in E$ can be written as $\boldsymbol{x} = \boldsymbol{x}^u + \boldsymbol{x}^s$, $\boldsymbol{x}^u \in E^u$, $\boldsymbol{x}^s \in E^s$. Proposition A1 says that a slight enlargement of E^u does not destroy the property $E^u \cap T_+ = \{\boldsymbol{0}\}$; i.e., the intersection between E^u and T_+ is the singleton set $\{\boldsymbol{0}\}$ and that the same is true for E^s and T_-. Proposition A2 says that $\boldsymbol{x}_1$ (resp. $\boldsymbol{x}_0$) approaches E^u (resp. E^s) as $K \to +\infty$. Lemma A1 tells us that $\boldsymbol{x}_1$ (resp. $\boldsymbol{x}_0$) approaches $\bar{\boldsymbol{x}}_1$ (resp. $\bar{\boldsymbol{x}}_0$) as $K \to +\infty$.

Proposition A1: There are positive numbers α_+ and α_- such that

$$\begin{aligned} \Lambda_{\alpha+}(E^u) \cap T_+ &= \{\boldsymbol{0}\} \\ \Lambda_{\alpha-}(E^s) \cap T_- &= \{\boldsymbol{0}\} \end{aligned} \tag{A1}$$

where $\Lambda_{\alpha+}(E^u)$ and $\Lambda_{\alpha-}(E^s)$ are the α_+ sector of E^u and the α_- sector of E^s, respectively (Fig. 13):

$$\begin{aligned} \Lambda_{\alpha+}(E^u) &:= \{(\boldsymbol{z}^u, \boldsymbol{z}^s) \in E^u \oplus E^s \mid \|\boldsymbol{z}^s\| < \alpha_+ \|\boldsymbol{z}^u\|\} \\ \Lambda_{\alpha-}(E^s) &:= \{(\boldsymbol{z}^u, \boldsymbol{z}^s) \in E^u \oplus E^s \mid \|\boldsymbol{z}^u\| < \alpha_- \|\boldsymbol{z}^s\|\}. \end{aligned}$$

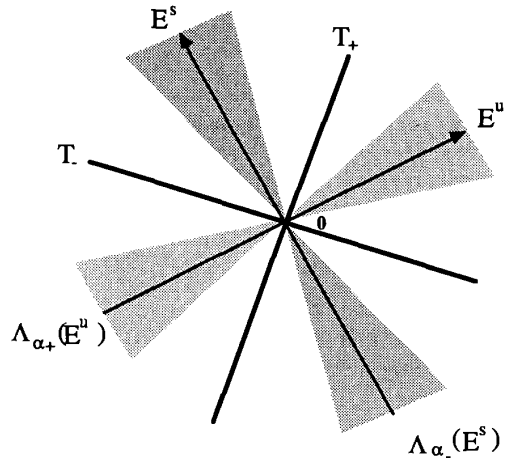


Fig. 13. The α_+ section of E^u and the α_- section of E^s.

Proof: Since $T_+ \oplus E^u = E^s \oplus E^u$, there is a unique linear map $\tau_+^u : E^u \to E^s$ such that

$$T_+ = \{(z^u, \tau_+^u z^u) \in E^u \oplus E^s \mid z^u \in E^u\}.$$

Since $T_+ \cap E^u = \{0\}$, the map τ_+^u is nonsingular; hence

$$\alpha_+ := \inf\{\|\tau_+^u z^u\| \mid z^u \in E^u, \|z^u\| = 1\}$$

is positive. Clearly (A1) is satisfied. A similar argument is valid for $\Lambda_{\alpha-}(E^s)$.

Proposition A2: If $\{x_k\}_{-K}^{+K}$ is a solution for (T_+, T_-, K), then

$$\|x_1^u\| \le \alpha_+^{-1}\lambda_\#^{-2(K-1)}\|x_1^s\|$$
$$\|x_0^s\| \le \alpha_-^{-1}\lambda_\#^{-2K}\|x_0^u\|$$

where $\lambda_\#$ is defined by (25).

Proof: Since $x_K \in T_+$, one has $x_K \notin \Lambda_{\alpha+}(E^u)/\{0\}$ so that $\|x_K^s\| \ge \alpha_+\|x_K^u\|$. It follows from $\|x_{k+1}^s\| = \|Fx_k^s\| \le \lambda_\#^{-1}\|x_k^s\|$ that $\|x_K^s\| \le \lambda_\#^{-(K-1)}\|x_1^s\|$. Therefore

$$\|x_K^u\| \le \alpha_+^{-1}\|x_K^s\| \le \alpha_+^{-1}\lambda_\#^{-(K-1)}\|x_1^s\|,$$

hence

$$\|x_1^u\| \le \alpha_+^{-1}\lambda_\#^{-2(K-1)}\|x_1^s\|. \tag{96}$$

The other inequality can be derived in a similar manner. □

Now let $\overline{\Lambda}_\varepsilon(E^u)$ and $\overline{\Lambda}_\varepsilon(E^s)$ denote the *closed* ε sectors of E^u and E^s respectively:

$$\overline{\Lambda}_\varepsilon(E^u) := \{(z^u, z^s) \in E^u \oplus E^s \mid \|z^s\| \le \varepsilon\|z^u\|\}$$
$$\overline{\Lambda}_\varepsilon(E^s) := \{(z^u, z^s) \in E^u \oplus E^s \mid \|z^u\| \le \varepsilon\|z^s\|\}.$$

Then Proposition A2 says that a solution $\{x_k\}_{-K}^{+K}$ for (T_+, T_-, K) satisfies

$$x_0 \in \overline{\Lambda}\varepsilon_1(E^u), \qquad x_1 \in \overline{\Lambda}\varepsilon_2(E^s), \qquad x_1 = F^d x_0 + y_0 \tag{A2}$$

where

$$\varepsilon_1 = \alpha_-^{-1}\lambda_\#^{-2K} \qquad \varepsilon_2 = \alpha_+^{-1}\lambda_\#^{-2(K-1)}. \tag{A3}$$

Lemma A1: Let $\varepsilon_1, \varepsilon_2 > 0$ satisfy

$$\max\left[\varepsilon_1(1-\varepsilon_1)^{-1}, \varepsilon_2(1-\varepsilon_2)^{-1}, \varepsilon_1, \varepsilon_2\right] \le 1/4 \tag{A4}$$

and let

$$r = 2\left(\|\overline{x}_1\| + \left\|F^d\overline{x}_0\right\|\right). \tag{A5}$$

If

$$z \in \left(\overline{\Lambda}\varepsilon_1(E^u) + y_0\right) \cap \overline{\Lambda}\varepsilon_2(E^s), \tag{A6}$$

then

$$z = w + y_0 \qquad w \in \overline{\Lambda}\varepsilon_1(E^u) \tag{A7}$$

and

$$\|z\| + \|w\| \le r \tag{A8}$$
$$\|z - \overline{x}_1\| \le \varepsilon_1(1-\varepsilon_1)^{-1}\|w\| + \varepsilon_2(1-\varepsilon_2)^{-1}\|z\| \tag{A9}$$
$$\left\|w - F^d\overline{x}_0\right\| \le \varepsilon_1(1-\varepsilon_1)^{-1}\|w\| + \varepsilon_2(1-\varepsilon_2)^{-1}\|z\|. \tag{A10}$$

Proof: If (A6) holds, $z \in \overline{\Lambda}\varepsilon_2(E^s)$ implies

$$z = z^u + z^s, \qquad z^u \in E^u, \qquad z^s \in E^s,$$
$$\text{and} \quad \|z^u\| \le \varepsilon_2\|z^s\|$$

which, in turn, implies

$$\|z^s\| = \|z - z^u\| \le \|z\| + \|z^u\| \le \|z\| + \varepsilon_2\|z_s\|.$$

Therefore

$$\|z^s\| \le (1-\varepsilon_2)^{-1}\|z\| \tag{A11}$$

and hence

$$\|z^u\| \le \varepsilon_2(1-\varepsilon_2)^{-1}\|z\|. \tag{A12}$$

Since $z \in \overline{\Lambda}\varepsilon_1(E^u) + y_0$ ((A6)), one has

$$z - y_0 = w^u + w^s, \qquad w^u \in E^u, \qquad w^s \in E^s$$
$$\|w^s\| \le \varepsilon_1\|w^u\| \tag{A13}$$

from which it follows that

$$\|w^s\| \le \varepsilon_1(1-\varepsilon_1)^{-1}\|w\|. \tag{A14}$$

Let us rewrite the equality (see (A13))

$$z^u + z^s = w^u + w^s + y_0$$

and

$$z^s - w^s = w^u - z^u + y_0.$$

Then the uniqueness of the stable free-boundary solution (Proposition 1) implies that there is a unique pair $(\overline{x}_1, \overline{x}_0)$ satisfying (see (24))

$$\overline{x}_1 = z^s - w^s \qquad F^d\overline{x}_0 = w^u - z^u. \tag{A15}$$

It follows from (A11), (A12), (A14), and (A15) that

$$\begin{aligned}\|z-\bar{x}_1\| &= \|z^u+z^s-(z^s-w^s)\| \\ &= \|z^u+w^s\| \le \|z^u\|+\|w^s\| \\ &\le \varepsilon_1(1-\varepsilon_1)^{-1}\|w\|+\varepsilon_2(1-\varepsilon_2)^{-1}\|z\|\end{aligned}$$

which proves (A9). Similarly

$$\begin{aligned}\left\|w-F^d\bar{x}_0\right\| &= \|w^u+w^s-(w^u-z^u)\| \\ &\le \|w^s\|+\|z^u\| \\ &= \varepsilon_1(1-\varepsilon_1)^{-1}\|w\|+\varepsilon_2(1-\varepsilon_2)^{-1}\|z\|\end{aligned}$$

which proves (A10). In order to show (A8), observe that

$$\|z\|-\|\bar{x}_1\| \le \|z-\bar{x}_1\|$$

and

$$\|w\|-\left\|F^d\bar{x}_0\right\| \le \left\|w-F^d\bar{x}_0\right\|$$

imply

$$\begin{aligned}\|z\| &\le \|\bar{x}_1\|+\|z-\bar{x}_1\| \\ &\le \|\bar{x}_1\|+\varepsilon_1(1-\varepsilon_1)^{-1}\|w\|+\varepsilon_2(1-\varepsilon_2)^{-1}\|z\|\end{aligned} \tag{A16}$$

and

$$\begin{aligned}\|w\| &\le \left\|F^d\bar{x}_0\right\|+\left\|w-F^d\bar{x}_0\right\| \\ &\le \left\|F^d\bar{x}_0\right\|+\varepsilon_1(1-\varepsilon_1)^{-1}\|w\|+\varepsilon_2(1-\varepsilon_2)^{-1}\|z\|.\end{aligned} \tag{A17}$$

Adding (A16) and (A17), one has

$$\begin{aligned}\|z\|+\|w\| &\le \|\bar{x}_1\|+\left\|F^d\bar{x}_0\right\|+2\varepsilon_1(1-\varepsilon_1)^{-1}\|w\| \\ &\quad +2\varepsilon_2(1-\varepsilon_2)^{-1}\|z\| \\ &\le \|\bar{x}_1\|+\left\|F^d\bar{x}_0\right\| \\ &\quad +2\max\left[\varepsilon_1(1-\varepsilon_1)^{-1},\varepsilon_2(1-\varepsilon_2)^{-1}\right] \\ &\quad \cdot(\|w\|+\|z\|) \\ &\le \|\bar{x}_1\|+\left\|F^d\bar{x}_0\right\|+1/2(\|w\|+\|z\|)\end{aligned}$$

where (A4) was used. This inequality together with (A5) implies

$$\|z\|+\|w\| \le 2\left(\|\bar{x}_1\|+\left\|F^d\bar{x}_0\right\|\right)=r. \qquad \square$$

Completion of the Proof: It follows from Proposition A2 that (A4) is satisfied for K sufficiently large. Since F^d expands the vectors in E^u while it contracts the vectors in E^s, one sees that

$$x_0 \in \bar{\Lambda}\varepsilon_1(E^u) \text{ implies } F^d x_0 \in \bar{\Lambda}\varepsilon_1(E^u).$$

Therefore, one can take

$$x_1 = F^d x_0 + y_0$$

as the z in (A7) of Lemma A1, and (A8) reads

$$\|x_1\|+\left\|F^d x_0\right\| \le r$$

and

$$\|x_1^s\| \le \|x_1\| \le r. \tag{A18}$$

Since $x_K = F^{K-1}x_1$, (A18) implies

$$\|x_K\| \le \lambda_\#^{-(K-1)} r. \tag{A19}$$

In order to estimate $\|x_k^u\|$, let

$$\tau_+^s : E^s \to E^u$$

be the linear map such that

$$T_+ = \left\{\left(\tau_+^s z^s, z^s\right) \in E^u \oplus E^s | z^s \in E^s\right\}. \tag{A20}$$

This map is well defined and is unique because of (34). It follows from (A20) that

$$\begin{aligned}\|x_K\| = \left\|\left(\tau_+^s x_K^s, x_K^s\right)\right\| &\le \left(\left\|\tau_+^s\right\|+1\right)\|x_K^s\| \\ &\le \lambda_\#^{-(K-1)}\left(\left\|\tau_+^s\right\|+1\right)r.\end{aligned}$$

Note that for the stable free-boundary solution, (24) implies $\bar{x}_k \in E^s$, $k \ge 1$; hence

$$\bar{x}_k^u = 0,$$

which, in turn, implies

$$\|x_K^u - \bar{x}_K^u\| = \|x_K^u\| \le \lambda_\#^{-(K-1)}\left\|\tau_+^s\right\| r.$$

It follows from this that

$$\begin{aligned}\|x_k^u-\bar{x}_k^u\| = \|x_k^u\| &= \left\|F^{-(K-k)}x_K^u\right\| \\ &\le \lambda_\#^{-(K-k)}\|x_K^u\| \le \lambda_\#^{-(K-k)}\lambda_\#^{-(K-1)}\left\|\tau_+^s\right\| r \\ &= \lambda_\#^{-2K+k+1}\left\|\tau_+^s\right\| r.\end{aligned} \tag{A21}$$

On the other hand,

$$\begin{aligned}\|x_1^s-\bar{x}_1^s\| &\le \varepsilon_1(1-\varepsilon_1)^{-1}\left\|F^d x_0\right\|+\varepsilon_2(1-\varepsilon_2)^{-1}\|x_1\| \\ &\le \max\left[\varepsilon_1(1-\varepsilon_1)^{-1},\varepsilon_2(1-\varepsilon_2)^{-1}\right]r \\ &\le 2\max(\varepsilon_1,\varepsilon_2)r \\ &\le 2\lambda_\#^{-2K}\max\left(\alpha_-^{-1},\alpha_+^{-1}\lambda_\#^2\right)r\end{aligned}$$

where (A3) was used. Therefore

$$\begin{aligned}\|x_k^s-\bar{x}_k^s\| &= \left\|F^{k-1}(x_1^s-\bar{x}_1^s)\right\| \\ &\le \lambda_\#^{-(k-1)}\|x_1^s-\bar{x}_1^s\| \\ &\le \lambda_\#^{-(k-1)}2\lambda_\#^{-2K}\max\left(\alpha_-^{-1},\alpha_+^{-1}\lambda_\#^2\right)r \\ &= \lambda_\#^{-2K-k+1}2\max\left(\alpha_-^{-1},\alpha_+^{-1}\lambda_\#^2\right)r.\end{aligned} \tag{A22}$$

It follows from (A21) and (A22) that

$$\begin{aligned}\sum_{k=1}^{K}\|\boldsymbol{x}_k-\bar{\boldsymbol{x}}_k\|^2 &\le \sum_{k=1}^{K}\|\boldsymbol{x}_k^u-\bar{\boldsymbol{x}}_k^u\|^2+\sum_{k=1}^{K}\|\boldsymbol{x}_k^s-\bar{\boldsymbol{x}}_k^s\|^2\\ &\le \sum_{k=1}^{K}\lambda_{\#}^{-4K+2k+2}\|\boldsymbol{\tau}_+^s\|^2r^2\\ &\quad+\sum_{k=1}^{K}\lambda_{\#}^{-4K-2k+2}\\ &\quad\cdot 4\left\{\max\left(\alpha_-^{-1},\alpha_+^{-1}\lambda_{\#}^2\right)\right\}^2r^2\\ &=\left(\sum_{k=1}^{K}\lambda_{\#}^{2k}+\sum_{k=1}^{K}\lambda_{\#}^{-2k}\right)\lambda_{\#}^{-4K+2}r^2\\ &\quad\cdot\max\left(\|\boldsymbol{\tau}_+^s\|^2,4\left\{\max\left(\alpha_-^{-1},\alpha_+^{-1}\lambda_{\#}^2\right)\right\}^2\right)\\ &\quad\to 0\quad\text{as}\quad K\to+\infty.\end{aligned}$$

Using a similar argument, one can show that

$$\sum_{k=-K}^{0}\|\boldsymbol{x}_k-\bar{\boldsymbol{x}}_k\|^2\to 0\quad\text{as}\quad K\to+\infty.\qquad\square$$

Appendix II
Proof of Proposition 6

(i) If μ is an eigenvalue of $\boldsymbol{A}$,

$$\boldsymbol{A}_\mu:=\boldsymbol{A}-\mu\mathbf{1}\tag{A23}$$

is singular, and hence it is not negative definite. Therefore $\boldsymbol{A}_\mu$ cannot be temporally stable. If

$$a_0-\mu<0\tag{A24}$$

then Proposition 2 says that $Q_\mu(\omega)$ defined by (45) for $\boldsymbol{A}_\mu$ has a real zero on $(-2,2)$ with odd multiplicity. Since the multiplicity is odd, the zero on $(-2,2)$ cannot be an extremum of Q_μ so that

$$\min_{\omega\in[-2,2]}Q_\mu(\omega)<0\quad\text{and}\quad\max_{\omega\in[-2,2]}Q_\mu(\omega)>0\tag{A25}$$

which is equivalent to

$$\min_{\omega\in[-2,2]}a_mQ_\mu(\omega)<0<\max_{\omega\in[-2,2]}a_mQ_\mu(\omega).\tag{A26}$$

Since

$$a_mQ_\mu(\omega)=a_0-\mu+\sum_{p=1}^{m}a_p\left(\lambda^p+\lambda^{-p}\right)\tag{A27}$$

where ω and λ are related via (44), one has

$$a_mQ_\mu(\omega)=a_mQ(\omega)-\mu.\tag{A28}$$

Equations (A26) and (A28) imply (66). In order to consider the case

$$a_0-\mu>0\tag{A29}$$

one needs the following:

Lemma A2: If $a_0-\mu>0$, the following are equivalent:

i) $\boldsymbol{A}_\mu$ is *positive* definite for all n.

ii) The corresponding spatial dynamics are regular.

iii) Every real zero of Q_μ on $(-2,2)$ has an even multiplicity.

Proof: Recall (49) and replace $-a_p$ by a_p:

$$a_p=\sum_{p=1}^{m-p}z_iz_{i+p},\qquad p=1,\cdots,m$$

$$a_0-\mu=\sum_{p=0}^{m}z_i^2.$$

One can use an argument similar to that used in the proof of Theorem 2 and Proposition 2 to show the result. $\square$

In order to complete the proof of (i) of Proposition 6, observe the fact that $\boldsymbol{A}_\mu$ being singular violates (i)–(iii) in Lemma A2. Since (iii) of Lemma A2 is the same as (iii) of Proposition 2, one can use the same argument as for $a_0-\mu<0$. It is clear from the form of $\boldsymbol{A}$ and $a_0\neq 0$, $a_m\neq 0$ that $a_0-\mu=0$ is impossible.

(ii) In order to prove the optimality of the upper bound, note that for any $\gamma\in\mathbb{R}$

$$\sigma_+(a_0+\gamma,a_1,\cdots,a_m)=\sigma_+(a_0,a_1,\cdots,a_m)+\gamma.\tag{A30}$$

Now fix $a_1,\cdots,a_m$ and consider

$$P_n(a_0-\mu,a_1,\cdots,a_m):=\det(\boldsymbol{A}-\mu\mathbf{1})\tag{A31}$$

where n denotes the size of $\boldsymbol{A}$. It follows from (A31) that if $\{\mu_{n,i}(a_0^n)\}_{i=1}$ and $\{\mu_{n,j}(a_0^n)\}_{j=1}$ are the eigenvalues of $\boldsymbol{A}$ when the diagonal is a_0 and a_0', respectively, then by an appropriate relabeling,

$$\mu_{n,i}(a_0)-a_0=\mu_{n,i}(a_0')-a_0',\qquad i=1,\cdots,n.\tag{A32}$$

In order to demonstrate the optimality of the upper bound, we first consider the case

$$\sigma_+(a_0,a_1,\cdots,a_m)=0.$$

If this is not optimal, there is a $\delta>0$ such that

$$\mu_{n,i}(a_0,a_1,\cdots,a_m)<-\delta<\sigma_+(a_0,a_1,\cdots,a_m)\tag{A33}$$

for all n and $1\le i\le n$. It follows from (A32) that

$$\begin{aligned}&\mu_{n,i}(a_0+\delta/2,\ a_1,\cdots,a_m)-(a_0+\delta/2)\\ &\quad=\mu_{n,i}(a_0,a_1,\cdots,a_m)-a_0\end{aligned}$$

whence

$$\mu_{n,i}(a_0,a_1,\cdots,a_m)=\mu_{n,i}(a_0+\delta/2,a_1,\cdots,a_m)-\delta/2\tag{A34}$$

for all n and $1\le i\le n$. Equation (A33) and (A34) imply

$$\mu_{n,i}(a_0+\delta/2,a_1,\cdots,a_m)<-\delta/2<0\tag{A35}$$

for all n and $1\le i\le n$. This means that $(a_0+\delta/2,a_1,\cdots,a_m)$ results in the temporal stability. On the other hand (A30) implies

$$\begin{aligned}\sigma_+(a_0+\delta/2,a_1,\cdots,a_m)&=\sigma_+(a_0,a_1,\cdots,a_m)\\ &\quad+\delta/2=\delta/2>0.\end{aligned}$$

This contradicts (A35) because Proposition 4 says that the temporal stability is equivalent to $\sigma_+<0$. In order to show

the general case

$$\sigma_+(a_0, a_1, \cdots, a_m) = \sigma^*$$

suppose that σ^* is not optimal. Then, there is a $\delta > 0$ such that

$$\mu_{n,i}(a_0, a_1, \cdots, a_m) < \sigma^* - \delta \tag{A36}$$

for all n and $1 \leq i \leq n$. It follows from (A32) that

$$\mu_{n,i}(a_0, a_1, \cdots, a_m) - a_0 = \mu_{n,i}(a_0 - \sigma^*, a_1, \cdots, a_m) - (a_0 - \sigma^*) \tag{A37}$$

for all n and $1 \leq i \leq n$. Equations (A36) and (A37) imply

$$\mu_{n,i}(a_0 - \sigma^*, a_1, \cdots, a_m) < -\delta \tag{A38}$$

for all n and $1 \leq i \leq n$. It follows from (A30) that

$$\sigma_+(a_0 - \sigma^*, a_1, \cdots, a_m) = 0.$$

It was shown earlier than when $\sigma_+ = 0$, it is the optimal upper bound. Therefore, (A38) contradicts the optimality. In order to show the optimality of the lower bound $\sigma_-(a_0, a_1, \cdots, a_m)$, note that

$$\sigma_+(-a_0, -a_1, \cdots, -a_m) = -\sigma_-(a_0, a_1, \cdots, a_m).$$

Furthermore, if μ is an eigenvalue of $\boldsymbol{A}$, then $-\mu$ is an eigenvalue of $-\boldsymbol{A}$. Since $-\sigma_-(a_0, a_1, \cdots, a_m)$ is the optimal upper bound for $-\boldsymbol{A}$, i.e.,

$$-\mu_{n,i}(a_0, a_1, \cdots, a_m) < -\sigma_-(a_0, a_1, \cdots, a_m)$$

one sees that

$$\sigma_-(a_0, a_1, \cdots, a_m) < \mu_{n,i}(a_0, a_1, \cdots, a_m)$$

and σ_- is optimal. □

APPENDIX III
PROOF OF PROPOSITION 7

We will give all the details for the sake of completeness. Since $m = 2$, $\boldsymbol{F}$ is 4×4 and is given by

$$\begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -1 & -\frac{a_1}{a_2} & -\frac{a_0}{a_2} & -\frac{a_1}{a_2} \end{bmatrix}$$

The characteristic polynomial is

$$P_F(\lambda) = \lambda^2 \left[\frac{a_0}{a_2} + \frac{a_1}{a_2}(\lambda + \lambda^{-1}) + (\lambda^2 + \lambda^{-2}) \right]$$

and

$$Q(\omega) = \omega^2 + \frac{a_1}{a_2}\omega + \frac{a_0}{a_2} - 2 = \left(\omega + \frac{1}{2}\frac{a_1}{a_2}\right)^2 - \frac{1}{4}\left(\frac{a_1}{a_2}\right)^2 + \frac{a_0}{a_2} - 2.$$

Case 1:

$$-\frac{1}{2}\frac{a_1}{a_2} \leq -2.$$

Since

$$\max_{\omega \in [-2,2]} Q(\omega) = Q(2), \qquad \min_{\omega \in [-2,2]} Q(\omega) = Q(-2)$$

one has

$$\sigma_+(a_0, a_1, a_2) = \begin{cases} a_2 Q(2) & \text{when } a_2 > 0 \\ a_2 Q(-2) & \text{when } a_2 < 0 \end{cases}$$

$$\sigma_-(a_0, a_1, a_2) = \begin{cases} a_2 Q(-2) & \text{when } a_2 < 0 \\ a_2 Q(2) & \text{when } a_2 > 0. \end{cases}$$

Case 2:

$$-\frac{1}{2}\frac{a_1}{a_2} \geq 2$$

$$\max_{\omega \in [-2,2]} Q(\omega) = Q(-2), \qquad \min_{\omega \in [-2,2]} Q(\omega) = Q(2)$$

$$\sigma_+(a_0, a_1, a_2) = \begin{cases} a_2 Q(-2) & \text{when } a_2 > 0 \\ a_2 Q(2) & \text{when } a_2 < 0 \end{cases}$$

$$\sigma_-(a_0, a_1, a_2) = \begin{cases} a_2 Q(2) & \text{when } a_2 > 0 \\ a_2 Q(-2) & \text{when } a_2 < 0. \end{cases}$$

Case 3:

$$-2 \leq -\frac{1}{2}\frac{a_1}{a_2} \leq 0$$

$$\max_{\omega \in [-2,2]} Q(\omega) = Q(2), \qquad \min_{\omega \in [-2,2]} Q(\omega) = Q(-a_1/2a_2)$$

$$\sigma_+(a_0, a_1, a_2) = \begin{cases} a_2 Q(2) & \text{when } a_2 > 0 \\ a_2 Q(-a_1/2a_2) & \text{when } a_2 < 0 \end{cases}$$

$$\sigma_-(a_0, a_1, a_2) = \begin{cases} a_2 Q(-a_1/2a_2) & \text{when } a_2 > 0 \\ a_2 Q(2) & \text{when } a_2 < 0. \end{cases}$$

Case 4:

$$0 < -\frac{1}{2}\frac{a_1}{a_2} < 2$$

$$\max_{\omega \in [-2,2]} Q(\omega) = Q(-2), \qquad \min_{\omega \in [-2,2]} Q(\omega) = Q(-a_1/2a_2)$$

$$\sigma_+(a_0, a_1, a_2) = \begin{cases} a_2 Q(-2) & \text{when } a_2 > 0 \\ a_2 Q(-a_1/2a_2) & \text{when } a_2 < 0 \end{cases}$$

$$\sigma_-(a_0, a_1, a_2) = \begin{cases} a_2 Q(-a_1/2a_2) & \text{when } a_2 > 0 \\ a_2 Q(-2) & \text{when } a_2 < 0. \end{cases}$$

Now note that

$$Q(2) = 2 + 2\frac{a_1}{a_2} + \frac{a_0}{a_2} \tag{A39}$$

$$Q(-2) = 2 - 2\frac{a_1}{a_2} + \frac{a_0}{a_2} \tag{A40}$$

$$Q\left(-\frac{1}{2}\frac{a_1}{a_2}\right) = -\frac{1}{4}\left(\frac{a_1}{a_2}\right)^2 + \frac{a_0}{a_2} - 2. \tag{A41}$$

In order to obtain the desired final form, we need to check the following cases:

i) $a_2 > 0$ and *case 1* $\leftrightarrow a_1 \geq 4a_2 > 0$
ii) $a_2 > 0$ and *case 2* $\leftrightarrow a_1 \leq -4a_2 < 0$
iii) $a_2 > 0$ and *case 3* $\leftrightarrow 0 \leq a_1 < 4a_2$
iv) $a_2 > 0$ and *case 4* $\leftrightarrow -4a_2 < a_1 < 0$
v) $a_2 < 0$ and *case 1* $\leftrightarrow a_1 \leq 4a_2 < 0$

vi) $a_2 < 0$ and *case 2* $\leftrightarrow a_1 \geq -4a_2 > 0$
vii) $a_2 < 0$ and *case 3* $\leftrightarrow 4a_2 < a_1 \leq 0$
viii) $a_2 < 0$ and *case 4* $\leftrightarrow -4a_2 > a_1 > 0$.

It follows from (A39)–(A41) that

$$\sigma_+(a_0, a_1, a_2) = \begin{cases} a_2 Q(2) & \text{when (i) or (iii) or (vi)} \\ a_2 Q(-2) & \text{when (ii) or (iv) or (v)} \\ a_2 Q(-a_1/2a_2) & \text{when (vii) or (viii)} \end{cases} \tag{A42}$$

$$\sigma_-(a_0, a_1, a_2) = \begin{cases} a_2 Q(2) & \text{when (ii) or (v) or (vii)} \\ a_2 Q(-2) & \text{when (i) or (iv) or (viii)} \\ a_2 Q(-a_1/2a_2) & \text{when (iii) or (iv).} \end{cases} \tag{A43}$$

It follows from (3) that

$$a_0 = -(g_0 + 2g_1 + 2g_2), \qquad a_1 = g_1, \qquad a_2 = g_2$$

so that (A39)–(A41) give

$$a_2 Q(2) = g_2\left[2 + 2\frac{g_1}{g_2} - \frac{1}{g_2}(g_0 + 2g_1 + 2g_2)\right] = -g_0 \tag{A44}$$

$$a_2 Q(-2) = g_2\left[2 + 2\frac{g_1}{g_2} - \frac{1}{g_2}(g_0 + 2g_1 + 2g_2)\right] = -g_0 - 4g_1 \tag{A45}$$

$$a_2 Q\left(-\frac{g_1}{2g_2}\right) = g_2\left[-\frac{1}{4}\left(\frac{g_1}{g_2}\right)^2 - \frac{1}{g_2}(g_0 + 2g_1 + 2g_2) - 2\right] = -g_0 - 2g_1 - 4g_2 - g_1^2/4g_2. \tag{A46}$$

Substituting (A44)–(A46) into (A42) and (A43), one has the relations shown at the bottom of the page. It is easy to see that these are the ones given by (67). □

Appendix IV

In [1], $g_0, g_1 > 0$ while $g_2 < 0$, and $g_1 = 4|g_2|$. We will give a simple explanation for the reader who is unfamiliar with the regularization theory [2], [3].

Let a set of noisy data $x_1, \cdots, x_n$ be given. Suppose one wants to interpolate the data with appropriate smoothness. A reasonable way of accomplishing this is to minimize

$$G(v) = \sum_{k=1}^{n} (x_k - v_k)^2 + \lambda \sum_{k=1}^{n} (2v_k - v_{k-1} - v_{k+1})^2 \tag{A47}$$

with respect to $\boldsymbol{v} = (v_1, \cdots, v_n)$. The first term is called the data term while the second term is called the penalty term and it represents the penalty on the "second-order derivative," i.e.,

$$\approx \lambda \int \left(\frac{d^2 v(x)}{dx^2}\right)^2 dx$$

and $\lambda > 0$ is the weight on the penalty. Since this is a straightforward quadratic minimization problem, the solution is obtained by differentiating (A47) with respect to v_k and setting it to zero:

$$x_k - v_k + \lambda[-6v_k + 4(v_{k-1} + v_{k+1}) - (v_{k-2} + v_{k-2})] = 0$$

and hence

$$-(1/\lambda + 6)v_k + 4(v_{k-1} + v_{k+1}) - (v_{k-2} + v_{k+2}) + (1/\lambda)x_k = 0. \tag{A48}$$

If $g_0, g_1 > 0$, $g_2 < 0$, and $g_1/|g_2| = 4$, $g_0/|g_2| = 1/\lambda$, then (A48) reads

$$-(g_0 + 2g_1 + 2g_2)v_k + g_1(v_{k-1} + v_{k+1}) + g_2(v_{k-2} + v_{k+2}) + u_k = 0$$

which is exactly (8) with $m = 2$, where $u_k = (1/\lambda)x_k$. Thus, by varying g_0 while g_1 and g_2 are fixed, one can control the weight λ which corresponds to varying the width of the Gaussian-like kernel. It should be noticed, however, that the architecture shown in Fig. 1 is a rather crude approximation for the two-dimensional problem.

Conversely, given a circuit, one can recover $G(\boldsymbol{v})$ as the *total cocontent:*

$$G(v) = \frac{1}{2}\boldsymbol{v}^T \boldsymbol{A}\boldsymbol{v} + \boldsymbol{v}^T \boldsymbol{u}$$

and the dynamics of the circuit minimizes $-G(\boldsymbol{v})$ by

$$\frac{d}{dt}[-G(\boldsymbol{v}(t))] = -(\boldsymbol{A}\boldsymbol{v} + \boldsymbol{u})^T \frac{d\boldsymbol{v}(t)}{dt} = -\frac{d\boldsymbol{v}(t)}{dt}^T \boldsymbol{B}^{-1} \frac{d\boldsymbol{v}(t)}{dt} < 0.$$

Note, however, that if $\boldsymbol{A}$ were not symmetric, the total cocontent would be undefined even if the circuit were linear.

Appendix V
Proof of Proposition 10

Recall D defined by (75).

$$\sigma_+(g_0, g_1, g_2) = \begin{cases} -g_0 & \text{when } g_1, g_2 > 0 \quad \text{or} \quad g_1 > 0, g_2 < 0, |g_1/g_2| \geq 4 \\ -g_0 - 4g_1 & \text{when } g_1 < 0, g_2 > 0 \quad \text{or} \quad g_1 < 0, g_2 < 0, |g_1/g_2| \geq 4 \\ -g_0 - 2g_1 - 4g_2 - g_1^2/4g_2 & \text{when } g_2 < 0, |g_1/g_2| \leq 4 \end{cases}$$

$$\sigma_-(g_0, g_1, g_2) = \begin{cases} -g_0 & \text{when } g_1 < 0, g_2 < 0 \quad \text{or} \quad g_1 < 0, g_2 > 0, |g_1/g_2| \geq 4 \\ -g_0 - 4g_1 & \text{when } g_1 > 0, g_2 < 0 \quad \text{or} \quad g_1 > 0, g_2 > 0, |g_1/g_2| \geq 4 \\ -g_0 - 2g_1 - 4g_2 - g_1^2/4g_2 & \text{when } g_2 > 0, |g_1/g_2| \leq 4. \end{cases}$$

Case 1: $D < 0$. It follows from (72) and (74) that $d\theta/d\omega$ has no real zero and hence Q is monotonically increasing. Therefore

$$\sigma_+ = \begin{cases} a_3 Q(2) & \text{when } a_3 > 0 \\ a_3 Q(-2) & \text{when } a_3 < 0 \end{cases} \tag{A49}$$

$$\sigma_- = \begin{cases} a_3 Q(-2) & \text{when } a_3 > 0 \\ a_3 Q(2) & \text{when } a_3 < 0. \end{cases} \tag{A50}$$

Case 2: $D = 0$. In this case $d\theta/d\omega$ has a double zero. But since Q is cubic, it is monotonically increasing and (A49) and (A50) are true.

Case 3: $D > 0$. This means that $d\theta/d\omega$ has two distinct real zeroes and hence Q has a local maximum at ξ_- and a local minimum at $\xi_+, \xi_- < \xi_+$.

a) $\xi_- < \xi_+ \leq -2$: It is clear that Q is monotonically increasing on $[-2, 2]$ and hence (A49) and (A50) still hold.

b) $2 \leq \xi_- < \xi_+$: Q is monotonically increasing on $[-2, 2]$ and (A49) as well as (A50) is true.

c) $\xi_- \leq -2, 2 \leq \xi_+$: Q is monotonically decreasing on $[-2, 2]$ and

$$\sigma_+ = \begin{cases} a_3 Q(-2) & \text{when } a_3 > 0 \\ a_3 Q(2) & \text{when } a_3 < 0 \end{cases} \tag{A51}$$

$$\sigma_- = \begin{cases} a_3 Q(2) & \text{when } a_3 > 0 \\ a_3 Q(-2) & \text{when } a_3 < 0. \end{cases} \tag{A52}$$

d) $\xi_- \leq -2 \leq \xi_+ \leq 2$: In this case, Q has a local minimum at ξ_+ and hence

$$\sigma_+ = \begin{cases} a_3 \max[Q(-2), Q(2)] & \text{when } a_3 > 0 \\ a_3 Q(\xi_+) & \text{when } a_3 < 0 \end{cases}$$

$$\sigma_- = \begin{cases} a_3 Q(\xi_+) & \text{when } a_3 > 0 \\ a_3 \max[Q(-2), Q(2)] & \text{when } a_3 < 0. \end{cases}$$

e) $-2 \leq \xi_- \leq 2 \leq \xi_+$ Since Q has a local maximum at ξ_-,

$$\sigma_+ = \begin{cases} a_3 Q(\xi_-) & \text{when } a_3 > 0 \\ a_3 \min[Q(-2), Q(2)] & \text{when } a_3 < 0 \end{cases}$$

$$\sigma_- = \begin{cases} a_3 \min[Q(-2), Q(2)] & \text{when } a_3 > 0 \\ a_3 Q(\xi_-) & \text{when } a_3 < 0. \end{cases}$$

f) $-2 \leq \xi_- < \xi_+ \leq 2$: Since Q has a local minimum as well as a local maximum within $[-2, 2]$,

$$\sigma_+ = \begin{cases} a_3 \max[Q(\xi_-), Q(2)] & \text{when } a_3 > 0 \\ a_3 \min[Q(-2), Q(\xi_+)] & \text{when } a_3 < 0 \end{cases}$$

$$\sigma_- = \begin{cases} a_3 \min[Q(-2), Q(\xi_+)] & \text{when } a_3 > 0 \\ a_3 \max[Q(\xi_-), Q(2)] & \text{when } a_3 < 0. \end{cases}$$

Combining all these cases, one obtains the relations given at the bottom of the page, where $f_\pm$ and $h_\pm$ are defined by (77)–(79). □

APPENDIX VI
PROOF OF COROLLARY 1

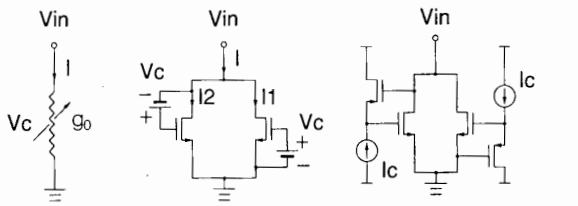
One can show that the *right derivative*,

$$D^+\|\boldsymbol{v}(t)\| = \lim_{\substack{h \to 0 \\ h > 0}} \frac{\|\boldsymbol{v}(t) + h\dot{\boldsymbol{v}}(t)\| - \|\boldsymbol{v}(t)\|}{h}$$

$$= \left(\lim_{\substack{h \to 0 \\ h > 0}} \frac{\|\boldsymbol{v}(t+h)\| - \|\boldsymbol{v}(t)\|}{h} \right)$$

$$\sigma_+(g_0, g_1, g_2, g_3) = \begin{cases} f_+ & \text{when } g_3 > 0 \text{ and (case 1 or case 2 or case 3-a or case 3-b)} \\ & \text{or } g_3 < 0 \text{ and case 3-c} \\ f_- & \text{when } g_3 < 0 \text{ and (case 1 or case 2 or case 3-a or case 3-b)} \\ & \text{or } g_3 > 0 \text{ and case 3-c} \\ \max[f_+, f_-] & \text{when } g_3 > 0 \text{ and case 3-d} \\ & \text{or } g_3 < 0 \text{ and case 3-e} \\ h_+ & \text{when } g_3 < 0 \text{ and case 3-d} \\ h_- & \text{when } g_3 > 0 \text{ and case 3-e} \\ \max[f_+, h_-] & \text{when } g_3 > 0 \text{ and case 3-f} \\ \max[f_-, h_+] & \text{when } g_3 < 0 \text{ and case 3-f} \end{cases}$$

$$\sigma_-(g_0, g_1, g_2, g_3) = \begin{cases} f_+ & \text{when } g_3 > 0 \text{ and case 3-c} \\ & \text{or } g_3 < 0 \text{ and (case 1 or case 2 or case 3-a or case 3-b)} \\ f_- & \text{when } g_3 < 0 \text{ and case 3-c} \\ & \text{or } g_3 > 0 \text{ and (case 1 or case 2 or case 3-a or case 3-b)} \\ \min[f_+, f_-] & \text{when } g_3 > 0 \text{ and case 3-d} \\ & \text{or } g_3 < 0 \text{ and case 3-e} \\ h_+ & \text{when } g_3 > 0 \text{ and case 3-d} \\ h_- & \text{when } g_3 < 0 \text{ and case 3-e} \\ \min[f_+, h_-] & \text{when } g_3 < 0 \text{ and case 3-f} \\ \min[f_-, h_+] & \text{when } g_3 > 0 \text{ and case 3-f} \end{cases}$$



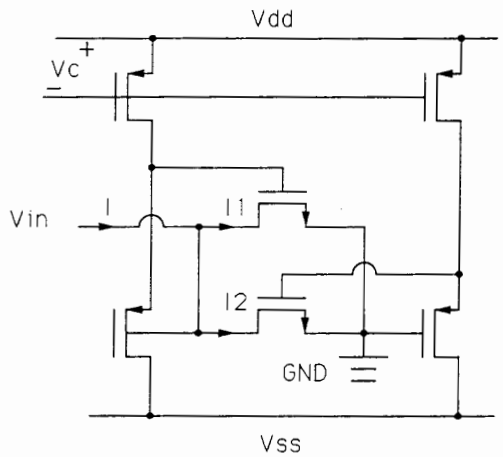
$I1 = K[(Vc - Vt)\,Vin - \frac{1}{2}Vin^2]$ (in triode region)

$I2 = K[\,(Vc + Vin - Vt)\,Vin - \frac{1}{2}\,Vin^2]$

$I = I1 + I2 = 2K[Vc - Vt]\,Vin$

I = go Vin
where go = 2K[Vc - Vt]

(a)



(b)

Fig. 14. Variable conductance g_0. (a) v_c controls the value of g_0. (b) Actual implementation.

exists despite the fact that $\|\boldsymbol{v}(t)\|$ it is *not* differentiable. Since

$$\frac{d\boldsymbol{v}(t)}{dt} = \boldsymbol{B}^{-1}\boldsymbol{A}\boldsymbol{v}(t) + \boldsymbol{B}^{-1}\boldsymbol{u}$$

and since

$$\begin{aligned}&\left\|\boldsymbol{v}(t) + h\boldsymbol{B}^{-1}\boldsymbol{A}\boldsymbol{v}(t) + h\boldsymbol{B}^{-1}\boldsymbol{u}\right\| \\ &\le \left\|\boldsymbol{1} + h\boldsymbol{B}^{-1}\boldsymbol{A}\right\|\|\boldsymbol{v}(t)\| + h\left\|\boldsymbol{B}^{-1}\boldsymbol{u}\right\|\end{aligned}$$

one has

$$\begin{aligned}D^+\|\boldsymbol{v}(t)\| = \lim_{\substack{h\to 0\\ h=0}}\left(\frac{\left\|\boldsymbol{1} + h\boldsymbol{B}^{-1}\boldsymbol{A}\right\| - 1}{h}\right)\|\boldsymbol{v}(t)\| \\ + \left\|\boldsymbol{B}^{-1}\boldsymbol{u}\right\| \qquad \text{(A53)}\end{aligned}$$

where the matrix norm is induced by the Euclidian norm:

$$\left\|\boldsymbol{1} + h\boldsymbol{B}^{-1}\boldsymbol{A}\right\| = \max_{\|\boldsymbol{v}=1\|}\left\|\left(\boldsymbol{1} + h\boldsymbol{B}^{-1}\boldsymbol{A}\right)\boldsymbol{v}\right\|.$$

One can easily show that the right-hand limit of the first term in (A53) also exists. Denoting this limit by

$$m\left(\boldsymbol{B}^{-1}\boldsymbol{A}\right) := \lim_{\substack{h\to 0\\ h>0}}\frac{\left\|\boldsymbol{1} + h\boldsymbol{B}^{-1}\boldsymbol{A}\right\| - 1}{h}$$

one has the right *differential inequality:*

$$\begin{aligned}D^+\|\boldsymbol{v}(t)\| \le m\left(\boldsymbol{B}^{-1}\boldsymbol{A}\right)\|\boldsymbol{v}(t)\| + \left\|\boldsymbol{B}^{-1}\boldsymbol{u}\right\|, \\ \|\boldsymbol{v}(0)\| = 0.\end{aligned}$$

It is not difficult to show that a solution of a differential inequality is bounded by the solution of the corresponding differential equation:

$$\begin{aligned}\frac{dw}{dt} = m\left(\boldsymbol{B}^{-1}\boldsymbol{A}\right)w + \left\|\boldsymbol{B}^{-1}\boldsymbol{u}\right\|, \\ w(0) = 0.\end{aligned}$$

Therefore

$$\|\boldsymbol{v}(t)\| \le \frac{1}{m\left(\boldsymbol{B}^{-1}\boldsymbol{A}\right)}\left[\exp\left(m\left(\boldsymbol{B}^{-1}\boldsymbol{A}\right)t - 1\right)\right]\left\|\boldsymbol{B}^{-1}\boldsymbol{u}\right\|. \qquad \text{(A54)}$$

Similarly, the *left derivative* satisfies

$$\begin{aligned}D^-\|\boldsymbol{v}(t)\| &\ge \lim_{\substack{h\to 0\\ h<0}}\left(\frac{\left\|\boldsymbol{1} + h\boldsymbol{B}^{-1}\boldsymbol{A}\right\| - 1}{h}\right)\|\boldsymbol{v}(t)\| \\ &\quad + \left\|\boldsymbol{B}^{-1}\boldsymbol{u}\right\| \\ &= -m\left(\boldsymbol{B}^{-1}\boldsymbol{A}\right)\|\boldsymbol{v}(t)\| + \left\|\boldsymbol{B}^{-1}\boldsymbol{u}\right\|,\end{aligned}$$

$$\|\boldsymbol{v}(0)\| = 0,$$

which yields

$$\begin{aligned}\|\boldsymbol{v}(t)\| \ge \frac{1}{-m\left(-\boldsymbol{B}^{-1}\boldsymbol{A}\right)} \\ \cdot \left[\exp\left(-m\left(-\boldsymbol{B}^{-1}\boldsymbol{A}\right)t\right) - 1\right]\left\|\boldsymbol{B}^{-1}\boldsymbol{u}\right\|. \qquad \text{(A55)}\end{aligned}$$

Finally, it is known [24] that

$$m\left(\boldsymbol{B}^{-1}\boldsymbol{A}\right) = \text{max. eigenvalue of } \boldsymbol{B}^{-1}\boldsymbol{A} \qquad \text{(A56)}$$

and hence

$$-m\left(-\boldsymbol{B}^{-1}\boldsymbol{A}\right) = \text{min. eigenvalue of } \boldsymbol{B}^{-1}\boldsymbol{A}. \qquad \text{(A57)}$$

It follows from (66), (84), and (A56) that

$$m\left(\boldsymbol{B}^{-1}\boldsymbol{A}\right) \le \frac{\sigma_+}{\eta_+}. \qquad \text{(A58)}$$

Similarly

$$-m\left(-\boldsymbol{B}^{-1}\boldsymbol{A}\right) \ge \frac{\sigma_-}{\eta_-}. \qquad \text{(A59)}$$

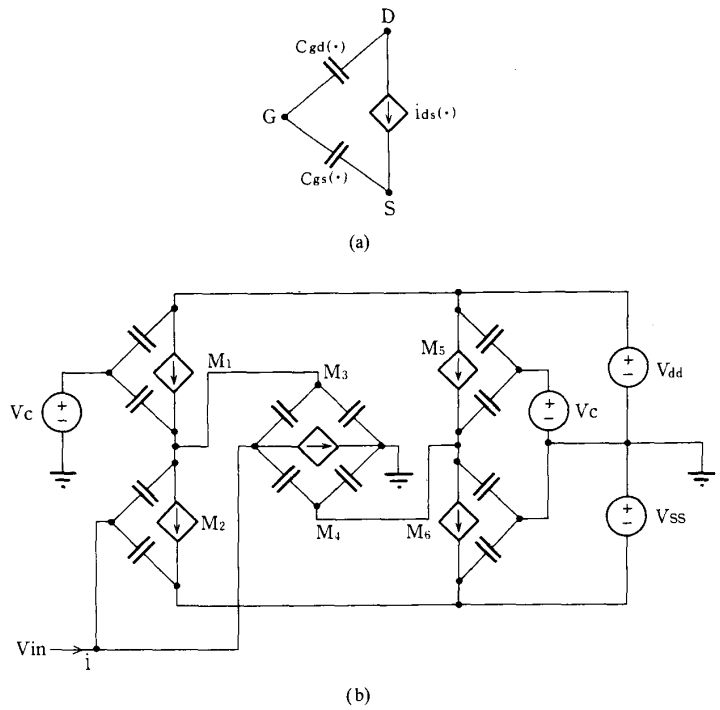


Fig. 15. Equivalent circuit of the g_0 circuit. (a) Equivalent circuit of an NMOS transistor; $i_{ds}(\cdot)$ indicates that the controlled current source is nonlinear, and $c_{gs}(\cdot)$ and $c_{gd}(\cdot)$ stand for nonlinear capacitors. (b) Equivalent circuit of Fig. 14(b).

Substituting (A58) and (A59) into (A54) and (A55), one has the desired bounds. □

APPENDIX VII

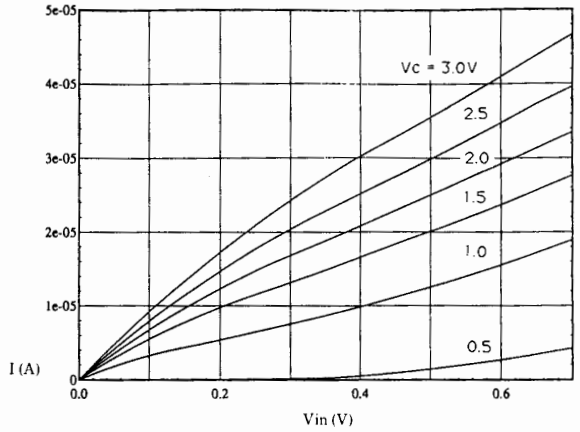
This appendix tries to justify the model given by (1)–(4). There are two aspects that must be examined:

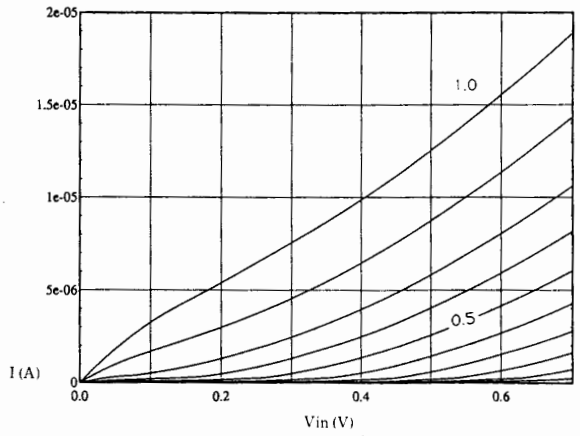
i) resistive part $g_0, g_1, \cdots, g_m$;
ii) capacitive part $c_0, c_1, \cdots, c_m$.

Although these parameters are implementation dependent, we can give a fairly reasonable account of them by checking the Gaussian-like convolver chip [1], where $m = 2$. Let us first look at Fig. 14, which implements g_0. Fig. 14(a) shows how g_0 can be made variable by controlling v_c, while Fig. 14(b) shows the actual implementation. In order to examine how this circuitry affects the resistive as well as the capacitive part of the model, one naturally has to have an equivalent circuit of each transistor. While a resistive part of an MOS transistor can be described by a simple nonlinear model, the capacitive part is known to be difficult to model [25]. In some cases it is described as a nonlinear distributed parameter element [26], and in some other cases it is described as a nonlinear, nonreciprocal multiterminal capacitor [27]. In many practical situations, parasitic capacitors are reciprocal and each is regarded as constant in each of the operating regions (cutoff, triode, and saturation) [25], [28], although they are still nonlinear, i.e., piecewise constant. (One has to be careful about the charge conservation because the incremental capacitance is discontinuous.) In many cases, a zero bulk charge is assumed. Fig. 15(a) gives such an equivalent circuit, where $i_{ds}(\cdot)$ indicates that the (controlled) current source is nonlinear, and $c_{gs}(\cdot)$ (resp. $c_{gd}(\cdot)$) represents nonlinear gate–source (resp. gate–drain) capacitor. A similar circuit can be given for a PMOS. Fig. 15(b) shows an equivalent circuit of Fig. 14(b) using Fig. 15(a). In order to examine the resistive part of the circuit, open-circuit all the capacitors. Fig. 16(a) shows the SPICE-simulated v_{in}–i characteristics while Fig. 16(b) gives measured characteristics which verify that the resistive part behaves in a sufficiently linear manner within the operating range. It should be noted that no small-signal argument is used. Namely, the linearity of the v_{in}–i characteristics does not mean that each transistor operates linearly. In fact, the four PMOS transistors are designed to operate in the saturation region.

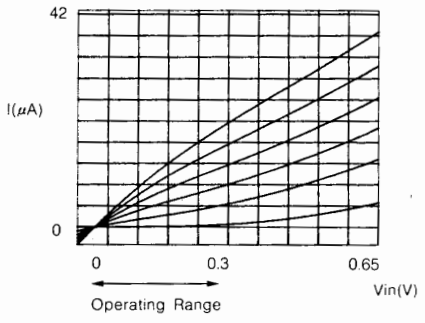
Next let us look at Fig. 17(a), which implements g_2, where $R_2 > 0$ is a p-well resistor and the remaining circuit realizes a negative impedance converter, where a triangle stands for a standard transconductance amplifier. Parts (b) and (c) of Fig. 17 give SPICE simulated and measured characteristics, respectively. In [1], g_1 is realized by a p-well resistor. Fig. 18 shows a SPICE simulation of a spatial impulse response at the *transistor* level. The reader is referred to [1] for measured impulse responses.

The capacitive part of the circuit needs more care to examine. In order to evaluate c_0, let us first check the g_0 circuit. To this end, open-circuit the current sources and short-circuit the voltage sources of Fig. 15(b) and obtain Fig. 19(a).



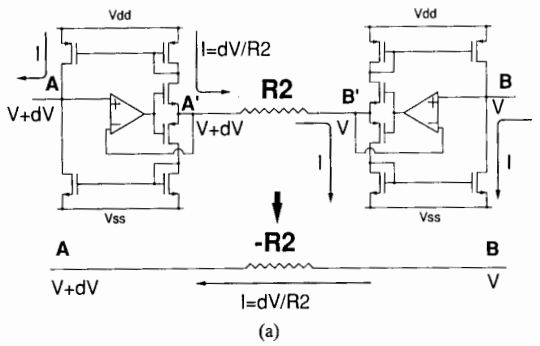


(a)



(b)

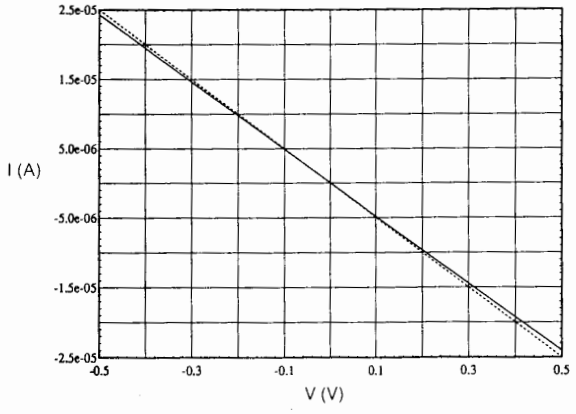
Fig. 16. The $v_{\text{in}} - i$ characteristics of Fig. 14(b). (a) SPICE simulated. (b) Measured.



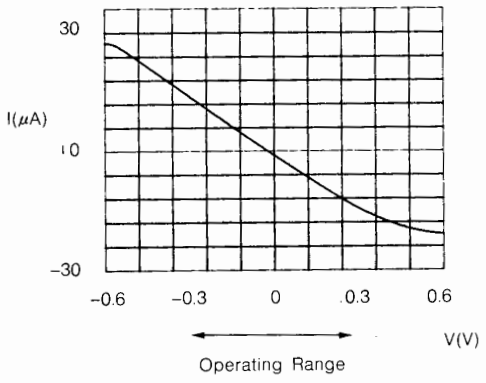
(a)

——— Actual Negative Resistor

— — — Ideal Negative Resistor



(b)



(c)

Fig. 17. Negative conductance g_2. (a) Circuitry. (b) SPICE simulated. (c) Measured.

That the resistive part behaves linearly does *not* guarantee that the capacitive part also behaves linearly. However, the pair of NMOS's in the middle is designed to operate in the triode region while the rest is designed to operate in the saturation region. Since we are assuming that each capacitance is constant in each operating region, c_{gd}'s and c_{gs}'s can be regarded as constant so that one can compute the overall equivalent capacitance, say c_0', between the v_{in} terminal and the ground. Since Fig. 19(a) is reduced to Fig. 19(b), one has

$$c_0' = \frac{(c_{02} + c_{04})(c_{01} + c_{05})}{c_{01} + c_{02} + c_{04} + c_{05}} + \frac{c_{06}(c_{07} + c_{08} + c_{09})}{c_{06} + c_{07} + c_{08} + c_{09}} + c_{03} \tag{A60}$$

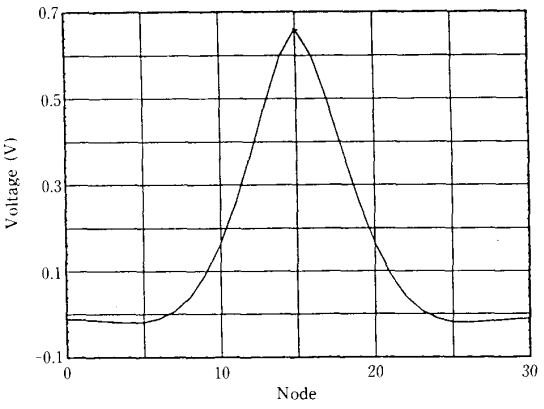


Fig. 18. SPICE simulated spatial impulse response at the transistor level where $1/g_1$ and $1/g_2$ are intended for 5 kΩ and −20 kΩ, respectively.

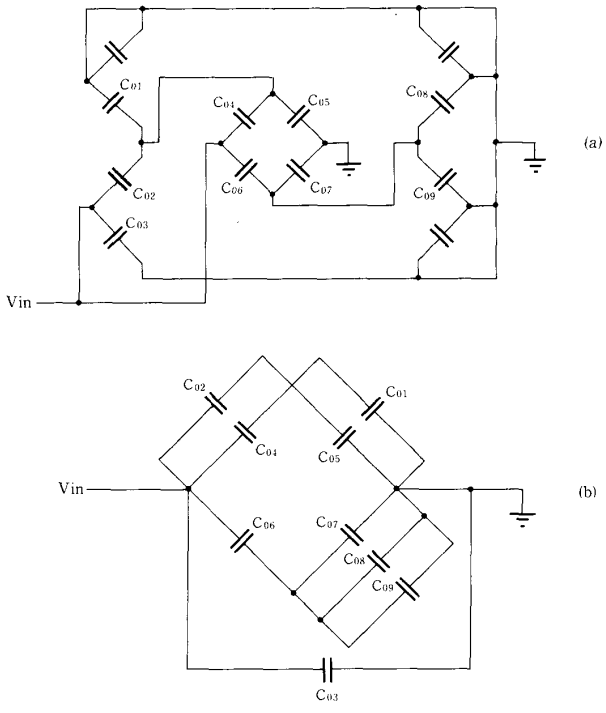


Fig. 19. Capacitive part of Fig. 15(b). (a) Original circuit. (b) Equivalent circuit.

Despite the fact that there are as many as nine capacitors contributing to c_0', the actual c_0' value would be very small. This stems from the fact that in the triode region, $c_{gs} = c_{gd} \approx (1/2)WLc_{ox}$ while in the saturation region $c_{gs} \approx (2/3)WLc_{ox}$, $c_{gd} \approx 0$ [25], [28], where W, L, and c_{ox} stand for the channel width, the channel length, and the capacitance (per unit area) of the oxide layer separating the gate from the channel. In this particular implementation, $W/L = 3/8$ (μm) for M_1 and M_5, $4/3$ for M_2 and M_4, and $7/2$ for M_2 and M_6, and $c_{ox} \approx 12 \times 10^{-4}$ pF/μm^2 in the present process. Since g_1 is a p-well resistor, its substrate is connected to v_{dd}. Thus there is a (distributed) diffusion capacitance between each node to v_{dd} (not between two nodes). In discussing the capacitive part of a circuit, one short-circuits voltage source as was done in the g_0 circuit. Therefore, this diffusion capacitance, say c_0'', contributes to c_0. The value of c_0'' would be larger than c_0' because (a) the area of the g_1 in this particular implementation is larger (36×20 μm^2) and (b) diffusion capacitance is the sum of a term proportional to the area and a term proportional to the peripheral length [28].

As for the contribution to c_0 from the g_2 circuit, there are two factors: (a) the parasitic capacitors of MOS transistors and (b) the p-well diffusion capacitance of $R_2 > 0$ (see Fig. 17(a)). The former can be calculated by using the same argument as the one used to compute c_0', while the latter can be estimated using the argument used to discuss the g_1 diffusion capacitance c_0''. If we call the resulting composite capacitance c_0''', the total capacitance between each node and the ground would be $c_0 = c_0' + c_0'' + c_0'''$.

Since conductance g_1 is implemented by a p-well, c_1 naturally represents associated parasitic capacitance between each node to its immediate neighbor. It should be noted, however, that c_1 appears in off-diagonal elements of $\boldsymbol{B}$.

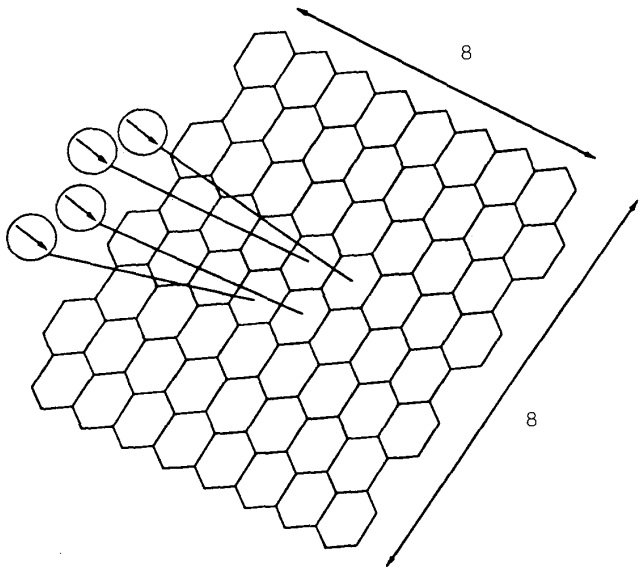
Finally, using the same argument, one can compute the composite capacitance c_2 from each node to its second nearest neighbor. The parasitic capacitor c_2 also appears in off-diagonal elements of $\boldsymbol{B}$.

It follows from (56) that $\boldsymbol{B}$ satisfies the diagonal dominance so that all eigenvalues are (strictly) positive. Naturally, in an actual implementation, $\boldsymbol{B}$ cannot be exactly symmetric. However, eigenvalues being strictly positive is an "open" condition, i.e., small variations of parameters do not destroy the property. We will leave quantitative estimates of those parasitic capacitances for a future paper. We will simply remark that $c_0 = 0.1$ pF used in Fig. 4 would not be too unrealistic.

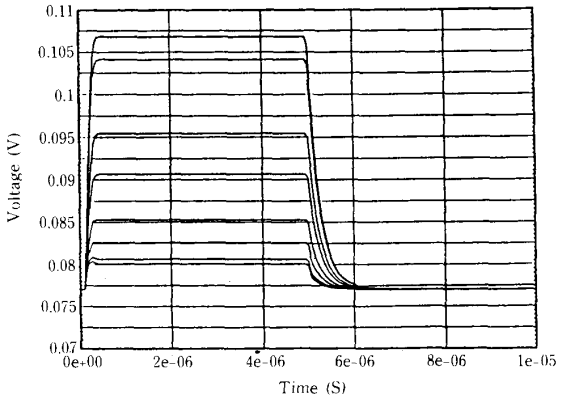
Fig. 20 shows a simulation result at the *transistor* level on SPICE where $1/g_0$, $1/g_1$, and $1/g_2$ are intended to be 200 kΩ, 5 kΩ, and −20 kΩ, respectively. A subnetwork of 8×8 is simulated (on a Cray) where a step current of duration 5 μs is injected into the four nodes as indicated in Fig. 20(a). Fig. 20(b) shows the voltage responses of the eight nodes on the fourth row. Although the above arguments are far from being complete, we believe that our model is sufficient for the present purpose.

ACKNOWLEDGMENT

The authors greatly appreciate the stimulating discussions with A. A. Abidi and J. L. White of UCLA. Thanks are also due to H. Kokubu of Kyoto University, H. Oka of Ryukoku University, M. W. Hirsch of U. C. Berkeley, M. Kando of the Science University of Tokyo, and A. Hio, K. Tanaka and Y. Takaku of Waseda University for discussions. The reviewers' comments were also very helpful. A part of this work was done while the second author was at UCLA.



(a)



(b)

Fig. 20. SPICE simulated temporal step responses at the transistor level. (a) 8 × 8 array is simulated where step current of duration 5 μs is injected to the four nodes as indicated. (b) Voltage responses at the eight nodes in the fourth row.

REFERENCES

[1] H. Kobayashi, J. L. White, and A. A. Abidi, "An active resistor network for Guassian filtering of images," *IEEE J. Solid-State Circuits,* vol. 26, pp. 738–748, May 1991.

[2] T. Poggio, H. Voohees, and A. Yuille, "A regularized solution to edge detection," AI Memo, MIT, Cambridge, MA, May 1985.

[3] T. Poggio, V. Torre, and C. Koch, "Computational vision and regularization theory," *Nature,* vol. 317, pp. 314–319, Sept. 1985.

[4] S. Grossberg, *Adaptive Brain (II).* Amsterdam: North Holland, 1987.

[5] J. G. Dougman, "Image analysis and compact coding by oriented 2D Gabor primitives," *SPIE,* vol. 758, pp. 19–30, 1987.

[6] D. L. Standley and J. L. Wyatt, Jr., "Stability criterion for lateral inhibition and related networks that is robust in the presence of integrated circuit parasitics," *IEEE Trans. Circuits Syst.,* vol. 36, pp. 675–681, May 1989.

[7] C. Mead, *Analog VLSI and Neural Systems.* Reading, MA: Addison-Wesley, 1989.

[8] C. Mead and M. Mahowald, "A silicon model of early visual processing," *Neural Networks,* vol. 1, no. 1, pp. 91–97, 1988.

[9] J. Harris, "An analog VLSI chip for thin-plate surface interpolation," presented at *IEEE Conf. Neural Info. Proc. Systems—Natural and Synthetic,* 1988.

[10] J. Hutchinson, C. Koch, J. Luo, and C. Mead, "Computing motion using analog and binary resistive network," *IEEE Computer,* vol. 21, pp. 52–63, Mar. 1988.

[11] J. Harris, C. Koch, J. Luo, and J. Wyatt, Jr., "Resistive fuses: Analog hardware for detecting discontinuities in early vision," in *Analog VLSI Implementation of Neural Systems.* Norwell, MA: Kluwer Academic, 1989.

[12] A. Lumsdaine, J. Wyatt, and I. Elfadel, "Nonlinear analog networks for image smoothing and segmentation," in *Proc. IEEE ISCAS,* 1990, pp. 987–991.

[13] S. C. Liu and J. Harris, "Generalized smoothing networks in early vision," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* 1989, pp. 184–191.

[14] B. Mathur, S. C. Liu, and H. T. Wang, "Analog neural networks for focal-plane image processing," *SPIE,* vol. 1242, pp. 141–151, 1990.

[15] T. Poggio and C. Koch, "Ill-posed problems in early vision: From computation theory to analog networks," *Proc. Roy. Soc. London, Ser. B226,* pp. 303–323, 1985.

[16] M. W. Hirsch and S. Smale, *Differential Equations, Dynamical Systems and Linear Algebra.* New York: Academic, 1974.

[17] T. Kamoto, Y. Akazawa, and M. Shinagawa, "An 8-bit 2-ns monolithic

DAC," *IEEE J. Solid-State Circuits,* vol. 23, no. 1, pp. 142–146, 1988.
[18] E. L. Allgower, "Criteria for positive definiteness of some band matrices," *Numer. Math.,* vol. 16, pp. 157–162, 1970.
[19] F. R. Gantmacher, *The Theory of Matrices.* New York: Chelsea, 1960.
[20] Y. Togawa and T. Matsumoto, "On the topological testability conjecture for analogue fault diagnosis problems," *IEEE Trans. Circuits Syst.,* vol. 31, pp. 147–158, Feb. 1984.
[21] A. V. Oppenheim and R. W. Schafer, *Digital Signal Processing.* Englewood Cliffs, NJ: Prentice Hall, 1975.
[22] J. Moody, "Dynamics of lateral interaction networks," in *Proc. 1990 IJCNN,* vol. III, 1990, pp. 483–486.
[23] S. Amari, "Dynamical study of formation of cortical maps," in *Dynamic Interactions in Neural Networks I,* M. Arbib and S. Amari Eds. New York: Springer, 1989, pp. 15–34.
[24] W. A. Coppel, *Stability and Asymptotic Behavior of Differential Equations.* Boston, MA: Heath, 1965.
[25] L. A. Glasser and R. W. Dobberpuhl, *The Design and Analysis of VLSI Circuits.* Reading, MA: Addison-Wesley, 1985.
[26] J. J. Paulos and D. A. Antoniadis, "Limitations of quasi-static capacitance models for the MOS transistors," *IEEE Electron Devices Lett.,* vol. EDL-4, pp. 221–224, July 1983.
[27] D. E. Ward and R. W. Dutton, "Charge-oriented model for MOS transistor capacitors," *IEEE J. Solid-State Circuits,* vol. SC-13, pp. 703–708, Oct. 1978.
[28] N. Weste and K. Eshraghian, *Principles of CMOS VLSI Design.* Reading, MA: Addison-Wesley, 1985.

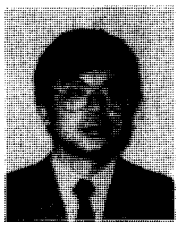
Takashi Matsumoto (M'71–SM'83–F'85) was born in Tokyo, Japan, on March 30, 1944. He received the B.Eng. degree in electrical engineering from Waseda University, Tokyo, Japan, the M.S. degree in applied mathematics from Harvard University, Cambridge, MA, and the Dr.Eng. degree in electrical engineering from Waseda University in 1966, 1970, and 1973, respectively.

At present he is Professor of Electrical Engineering, at Waseda University. From 1977 to 1979 he was on leave with the Department of Electrical Engineering and Computer Sciences, University of California at Berkeley. His research interests include nonlinear networks and neural networks.

Within the IEEE, he served as Associate Editor of the TRANSACTIONS ON CIRCUITS AND SYSTEMS and is a member of the Board of Governors of the Circuits and Systems Society. . He was a Guest Editor for the TRANSACTIONS ON CIRCUITS AND SYSTEMS (July 1988) and the PROCEEDINGS OF THE IEEE (August 1987). He is a member of the Circuits and Systems Society's Nonlinear Circuits and Systems Technical Committee and Neural Networks Technical Committee. He serves on the editorial board of the PROCEEEDINGS OF THE IEEE and is Vice Chairperson of the Tokyo Chapter of the Circuits and Systems Society.

Dr. Matsumoto is a member of the Institute of Electronics, Information and Communication Engineering (the IEICE—Japan's counterpart to the IEEE). He was chairperson of the Nonlinear Problems Technical Committee and was on the editorial boards of the *Transactions of the IEICE* and the *Journal of the IEICE.*

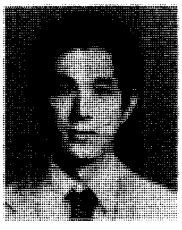
In addition, he serves on the editorial board of the journal *Circuits, Systems and Signal Processing* (Elsevier).



Haruo Kobayashi (S'88–M'90) was born in Utsunomiya, Japan, in 1958. He received the B.S. and M.S. degrees in information physics and mathematical engineering from the University of Tokyo in 1980 and 1982 respectively. From 1987 to 1989, he was at the University of California, Los Angeles, where he received the M.S. degree in electrical engineering in 1989.

He joined Yokogawa Electric Corporation, Tokyo, Japan, in 1982, where he has been engaged in research and development work on an FFT analyzer, a mini-supercomputer, and an LSI tester. His recent research interests include analog CMOS IC design and neural networks.

Mr. Kobayashi is a member of the Institute of Electronics, Information and Communication Engineers of Japan and the Society of Instrument and Control Engineers of Japan.



Yoshio Togawa was born in Tokyo, Japan, on January 5, 1953. He received the B.Sc., M.Sc., and Dr.Sc. degrees in mathematics from Waseda University, Tokyo, Japan, in 1975, 1977, and 1981, respectively.

Since 1977, he has been with the Science University of Tokyo, where his research has dealt mainly with dynamical systems.

Two-Dimensional Spatio-Temporal Dynamics of Analog Image Processing Neural Networks

Haruo Kobayashi, *Member, IEEE,* Takashi Matsumoto, *Fellow, IEEE,* and Jun Sanekata

Abstract— **A typical analog image-processing neural network consists of a two-dimensional (2-D) array of simple processing elements. When it is implemented with CMOS LSI, two dynamics issues naturally arise:**

1) **Parasitic capacitors of MOS transistors induce temporal dynamics. Since a processed image is given as the stable equilibrium point of temporal dynamics, a temporally unstable chip is unusable.**
2) **Because of the array structure, the node voltage distribution induces spatial dynamics, and the node voltage distribution could behave in a wild manner, e.g., oscillatory, which is undesirable for image-processing purposes.**

A discussion of these issues for one-dimensional cases is found in [1]. This paper extends its results to 2-D cases and also derives several explicit formulas and relationships for the 2-D dynamics, which are useful for the design and analysis of the class of networks of interest. Specifically, the following are derived: i) explicit spatial and temporal stability conditions and their equivalency, ii) spatial impulse responses, iii) spatial frequency responses, iv) power consumption, v) time constants, vi) relationships between spatial frequency responses and stability, vii) relationships between power consumption and stability, viii) relationships between spatial impulse responses and the discrete Fourier transform of network parameters, ix) relationships between spatial impulse responses and the inverse Z-transform of a transfer function, x) relationships between spatial frequency responses and time constants, xi) relationships between spatial frequency responses and equivalent circuits, xii) the characteristics of stable and unstable network dynamics, and xiii) hexagonal as well as square grid network dynamics.

I. Introduction

THIS study has been motivated by spatial versus temporal stability issues of analog image-processing neuro chips. The image-smoothing neuro chip [2] implemented by one of the authors, for instance, consists of a regular array of photosensors with conductances $g_0 > 0, g_1 > 0, g_2 < 0$ (Fig. 1). We refer the reader to [2] for the chip details. Since the chip involves negative conductances g_2, both spatial and temporal stability issues naturally arise. There are two intriguing elements. First, our earlier numerical experiments suggested that generally a neuro chip is temporally stable if and only if it is spatially stable, where spatial stability means that the node voltage distribution behaves "properly." Second, spatial dynamics naturally induces a discrete linear dynamical system

Manuscript received May 5, 1993; revised March 6, 1994, June 21, 1994, and November 3, 1994.

H. Kobayashi is with Teratec Corporation, Tokyo 180 Japan, on temporary leave from Yokogawa Electric Corp.

T. Matsumoto and J. Sanekata are with the Department of Electrical Engineering, Waseda University, Tokyo 169 Japan.

IEEE Log Number 9410198.

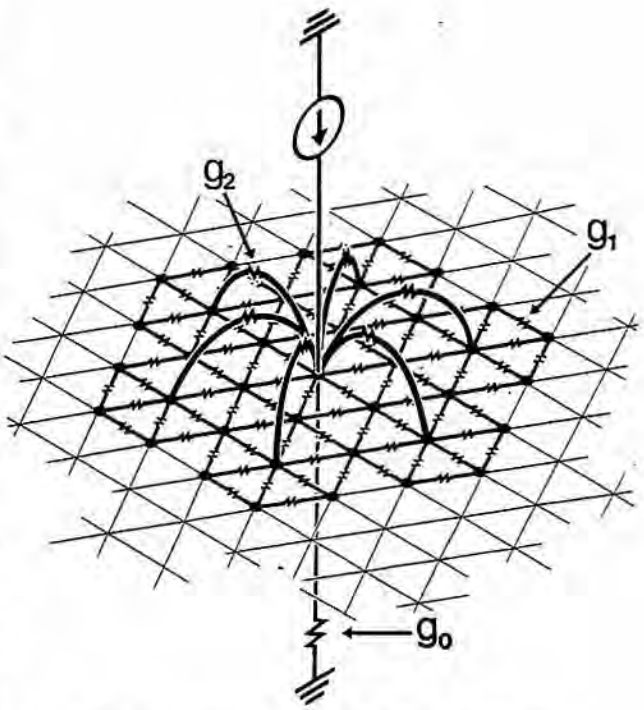


Fig. 1. The image-smoothing neuro chip. Only one unit is shown.

so that its stability should be checked by its eigenvalues. "A discrete linear dynamical system is stable if and only if all the eigenvalues lie inside the unit circle of the complex plane." This statement turned out to be false. Namely, due to the noncausal nature of the dynamics, if λ is an eigenvalue, so is $1/\lambda$, and hence the stability condition for causal linear systems is never satisfied.

Most of the fundamental issues involving these two elements have been settled in [1] for one-dimensional (1-D) array cases. For instance, a network is temporally stable if and only if it is spatially stable, except for a set of Lebesgue measure zero in the parameter space. Another fundamental result was that a network is spatially stable if and only if the eigenvalues of the dynamics are off the unit circle, even though they can be outside the unit circle. These results are far from trivial. One of the reasons that makes these results crucial is the boundary conditions associated with the finiteness of a network. Even if the eigenvalue conditions are satisfied, solutions can oscillate or explode if the boundary conditions are inappropriate.

Although our results in [1] are completely rigorous, the results are for 1-D cases only. The purpose of this paper is two-fold:

where $k_1 = 0, 1, 2, \cdots, N_1 - 1$ and $k_2 = 0, 1, 2, \cdots, N_2 - 1$.

ii) The associated eigenvectors are given by

$$\boldsymbol{c}_{N_1,k_1,N_2,k_2} = (c_{N_1,k_1}^{N_2,0}, c_{N_1,k_1}^{N_2,k_2}, c_{N_1,k_1}^{N_2,2k_2}, \cdots c_{N_1,k_1}^{N_2,(N_2-1)k_2})^T \tag{7.2}$$

and

$$\boldsymbol{s}_{N_1,k_1,N_2,k_2} = (s_{N_1,k_1}^{N_2,0}, s_{N_1,k_1}^{N_2,k_2}, s_{N_1,k_1}^{N_2,2k_2}, \cdots s_{N_1,k_1}^{N_2,(N_2-1)k_2})^T \tag{7.3}$$

where $c_{N_1,k_1}^{N_2,k_2}$ is given by (3.4) and $s_{N_1,k_1}^{N_2,k_2}$ is given by

$$\begin{aligned} s_{N_1,k_1}^{N_2,k_2} := \Bigg(&\sin\left(2\pi\frac{k_2}{N_2}\right), \sin\left(2\pi\left(\frac{k_1}{N_1}+\frac{k_2}{N_2}\right)\right), \\ &\sin\left(2\pi\left(\frac{2k_1}{N_1}+\frac{k_2}{N_2}\right)\right), \cdots, \\ &\sin\left(2\pi\left(\frac{(N_1-1)k_1}{N_1}+\frac{k_2}{N_2}\right)\right)\Bigg). \end{aligned} \tag{7.4}$$

Remark: In general, an eigenvalue has a unique associated eigenvector. In this case, however, since the matrixes are symmetric, each eigenvalue has a multiplicity of two (except for the case $k_1 = k_2 = 0$), i.e., $\lambda_{N_1,k_1,N_2,k_2} = \lambda_{N_1,N_1-k_1,N_2,N_2-k_2}$, and then each eigenvalue has two associated eigenvectors.

iii) When $\boldsymbol{A_b}$ is nonsingular, $\boldsymbol{A_b}^{-1}$ is also symmetric block circulant and its eigenvalues are given by

$$\lambda_{N_1,k_1,N_2,k_2}^{-1} = \frac{1}{\displaystyle\sum_{p=-m_1}^{m_1}\sum_{q=-m_2}^{m_2} a_{p,q}\cos\left(2\pi\left(\frac{k_1 p}{N_1}+\frac{k_2 q}{N_2}\right)\right)}$$

and the associated eigenvectors are also given by (7.2) and (7.3).

Lemma 2—Diagonalization, Exponential and Inverse of block circulant matrix: i) A block circulant matrix $\boldsymbol{A_b}$ with circulant blocks can be diagonalized by Fourier matrixes $\boldsymbol{F}_{N_1}, \boldsymbol{F}_{N_2}$ and their conjugate transpose $\boldsymbol{F}_{N_1}^*, \boldsymbol{F}_{N_2}^*$

$$\boldsymbol{A_b} = (\boldsymbol{F}_{N_1} \otimes \boldsymbol{F}_{N_2})^* \boldsymbol{\Lambda_b} (\boldsymbol{F}_{N_1} \otimes \boldsymbol{F}_{N_2})$$

where

$$\begin{aligned} \boldsymbol{\Lambda_b} := \operatorname{diag}\,(&\lambda_{N_1,0,N_2,0}, \lambda_{N_1,0,N_2,1}, \cdots, \\ &\lambda_{N_1,0,N_2,N_2-1}; \lambda_{N_1,1,N_2,0}, \lambda_{N_1,1,N_2,1}, \cdots, \\ &\lambda_{N_1,1,N_2,N_2-1}; \cdots \lambda_{N_1,N_1-1,N_2,0}, \\ &\lambda_{N_1,N_1-1,N_2,1}, \cdots, \lambda_{N_1,N_1-1,N_2,N_2-1}) \end{aligned}$$

and $\lambda_{N_1,k_1,N_2,k_2}$ is an eigenvalue of $\boldsymbol{A_b}$.

ii) If $\boldsymbol{A_b}$ is a block circulant matrix with circulant blocks, and if

$$\boldsymbol{A_b} = (\boldsymbol{F}_{N_1} \otimes \boldsymbol{F}_{N_2})^* \boldsymbol{\Lambda_b} (\boldsymbol{F}_{N_1} \otimes \boldsymbol{F}_{N_2})$$

then the exponential of $\boldsymbol{A_b}$, $e^{\boldsymbol{A_b}}$, is also a block circulant with circulant blocks and is given by

$$e^{\boldsymbol{A_b}} = (\boldsymbol{F}_{N_1} \otimes \boldsymbol{F}_{N_2})^* e^{\boldsymbol{\Lambda_b}} (\boldsymbol{F}_{N_1} \otimes \boldsymbol{F}_{N_2})$$

where

$$\begin{aligned} e^{\boldsymbol{\Lambda_b}} = \operatorname{diag}\,(&e^{\lambda_{N_1,0,N_2,0}}, e^{\lambda_{N_1,0,N_2,1}}, \cdots, \\ &e^{\lambda_{N_1,0,N_2,N_2-1}}; e^{\lambda_{N_1,1,N_2,0}}, e^{\lambda_{N_1,1,N_2,1}}, \cdots, \\ &e^{\lambda_{N_1,1,N_2,N_2-1}}; \cdots e^{\lambda_{N_1,N_1-1,N_2,0}}, \\ &e^{\lambda_{N_1,N_1-1,N_2,1}}, \cdots, e^{\lambda_{N_1,N_1-1,N_2,N_2-1}}). \end{aligned}$$

iii) If $\boldsymbol{A_b}$ is a nonsingular block circulant matrix with circulant blocks, and if

$$\boldsymbol{A_b} = (\boldsymbol{F}_{N_1} \otimes \boldsymbol{F}_{N_2})^* \boldsymbol{\Lambda_b} (\boldsymbol{F}_{N_1} \otimes \boldsymbol{F}_{N_2})$$

then $\boldsymbol{A_b}^{-1}$ is also a block circulant matrix with circulant blocks and is given by

$$\boldsymbol{A_b}^{-1} = (\boldsymbol{F}_{N_1} \otimes \boldsymbol{F}_{N_2})^* \boldsymbol{\Lambda_b}^{-1} (\boldsymbol{F}_{N_1} \otimes \boldsymbol{F}_{N_2})$$

where

$$\begin{aligned} \boldsymbol{\Lambda_b}^{-1} = \operatorname{diag}\,(&\lambda_{N_1,0,N_2,0}^{-1}, \lambda_{N_1,0,N_2,1}^{-1}, \cdots, \\ &\lambda_{N_1,0,N_2,N_2-1}^{-1}; \lambda_{N_1,1,N_2,0}^{-1}, \lambda_{N_1,1,N_2,1}^{-1}, \cdots, \\ &\lambda_{N_1,1,N_2,N_2-1}^{-1}; \cdots; \lambda_{N_1,N_1-1,N_2,0}^{-1}, \\ &\lambda_{N_1,N_1-1,N_2,1}^{-1}, \cdots, \lambda_{N_1,N_1-1,N_2,N_2-1}^{-1}). \end{aligned}$$

ACKNOWLEDGMENT

The authors would like to thank Y. Togawa of Science University of Tokyo for valuable discussions. Thanks are also due to the reviewers for their careful reading of the manuscript and their detailed comments.

REFERENCES

[1] T. Matsumoto, H. Kobayashi, and Y. Togawa, "Spatial versus temporal stability issues in image processing neuro chips," *IEEE Trans. Neural Networks*, vol. 3, no. 4, pp. 540–569, July 1992.

[2] H. Kobayashi, J. L. White, and A. A. Abidi, "An active resistor network for Gaussian filtering of images," *IEEE J. Solid-State Circuits*, vol. 26, no. 5, pp. 738–748, May 1991.

[3] J. L. White and A. N. Willson, Jr., "On the equivalence of spatial and temporal stability for translation invariant linear resistive networks," *IEEE Trans. Circuits Syst.-I*, vol. 39, no. 9, pp. 734–743, Sep. 1992.

[4] T. Poggio, H. Voorhees, and A. Yuille, "A regularized solution to edge detection," MIT AI Lab., Memo 833, 1985.

[5] B. E. Shi and L. O. Chua, "Resistive grid image filtering: Input/output analysis via CNN network," *IEEE Trans. Circuits Syst.-I*, vol. 39, no. 7, pp. 531–548, July 1992.

[6] D. G. Kelly, "Stability in contractive nonlinear neural networks," *IEEE Trans. Biomedical Eng.*, vol. 37, no. 3, pp. 231–241, Mar. 1990.

[7] C. Mead, *Analog VLSI and Neural Systems*. Reading, MA: Addison-Wesley, 1989.

[8] D. Standley and J. Wyatt, "Stability criterion for lateral inhibition and related networks that is robust in the presence of integrated circuit parasitics," *IEEE Trans. Circuits Syst.-I*, vol. 36, no. 5, pp. 675–681, May 1989.

[9] P. J. Davis, *Circulant Matrixes*. New York: Wiley, 1979.

[10] J. L. White and A. A. Abidi, "Active resistor networks as 2-D sampled data filters," *IEEE Trans. Circuits Syst.-I*, vol. 39, no. 9, pp. 724–733, Sept. 1992.

1) First, the stability results obtained in [1] are extended rigorously to two-dimensional 2-D cases, for which several new ideas are necessary.
2) Second, this paper discusses the 2-D dynamics issues and derives the following explicit formulas and relationships which are useful for designing and analyzing the class of filters described: i) spatial impulse responses, ii) spatial frequency responses, iii) power consumption, iv) time constants, v) relationships between spatial frequency responses and stability, vi) relationships between power consumption and stability, vii) relationships between spatial impulse responses and the discrete Fourier transform of network parameters, viii) relationships between spatial impulse responses and the inverse Z-transform of a transfer function, ix) relationships between spatial frequency responses and time constants, x) relationships between spatial frequency responses and equivalent circuits, xi) the characteristics of stable and unstable network dynamics, and xii) hexagonal as well as square grid network dynamics.

Related Works: The paper by White and Wilson [3] is vitally related to the present one. Some of the results in [3] regarding stability are similar to ours [1) mentioned above], even though the two approaches are completely different and there are several different issues addressed; e.g., [3] discusses the relationship between boundary conditions and temporal stability while [1] and this paper discuss the relationship between boundary conditions and spatial stability. In addition, this paper describes not only stability issues but also dynamics issues, and our approach leads to several explicit results [2) mentioned above].

Our approach in this paper is a systematic exploitation of the block circulant network structure for 2-D cases and the circulant network structure for 1-D cases. Speaking roughly, a block circulant network has a "two-torus" (doughnut) structure while a circulant network has a "ring" structure. Precise definitions will be given later. This exploitation of the block circulant and circulant structures has been used in several other previous works, e.g., [4]–[6]. In the present paper, however, since actual chips are not block circulant, our results for block circulant and circulant networks would be of little value unless block circulant and circulant networks behave in a manner similar to the networks that are uniform block and uniform band, respectively (see Section II-B for a precise definition). Also the networks include active elements (negative conductances) and thus the boundary conditions are crucial as already shown in [1] for 1-D cases. Therefore, a careful examination must be done to determine whether this approach can be of any use. It is shown that as the network size grows, responses of stable block circulant and circulant networks behave in a manner similar to those of stable uniform block and uniform band networks, respectively, which makes the approach valid.

We also remark that these theoretical results have important and practical consequences for the design and implementation of the class of filters with the analog CMOS LSI. Another example of the stability study for the analog early vision chip is the work of Standley and Wyatt [8] who investigated the stability conditions of Mead's chips [7] because they could be unstable in certain conditions. Another chip [2] which motivated the present study has negative conductances and thus it can easily be unstable. If the analog network is unstable, it is unusable, and thus these stability issues are important in analog neuro chips as well as in many other analog circuits.

II. Formulation

A. Circulant Networks

Let us first explain circulant networks before explaining block circulant networks. Consider a 1-D network with N nodes numbered zero through $N-1$, where each node k is connected with a current source u_k, a (possibly negative) conductance g_0, a capacitance c_0 to ground, and (possibly negative) conductances g_l and capacitances c_l to the lth nearest neighbor nodes ($l = 1, 2, \cdots, m$). We neglect inductance components here because most analog image-processing neuro chips are implemented by CMOS [2], [7], where inductances are practically zero. The network is called circulant if the rightmost and leftmost nodes are connected (ring-shaped). Fig. 2 shows a circulant network where $m = 2$. Observe that the Kirchoff current law (KCL) at node k reads

$$\begin{aligned}\frac{d}{dt}\{&(c_0+2c_1+2c_2)v_k - c_1(v_{k+1}+v_{k-1})\\ &- c_2(v_{k+2}+v_{k-2})\}\\ &= -(g_0+2g_1+2g_2)v_k + g_1(v_{k+1}+v_{k-1})\\ &+ g_2(v_{k+2}+v_{k-2}) + u_k.\end{aligned}$$

Since the network is circulant, node zero is connected to node $N-1$ as a nearest neighbor node and hence the state equation of the network where $m = 2$ is given by

$$\boldsymbol{B}\frac{d}{dt}\boldsymbol{v} = \boldsymbol{A}\boldsymbol{v} + \boldsymbol{u}$$

where

$$\begin{aligned}&\boldsymbol{v} := (v_0, v_1, \cdots, v_{N-1})^T \in \mathcal{R}^N,\\ &\boldsymbol{u} := (u_0, u_1, \cdots, u_{N-1})^T \in \mathcal{R}^N,\\ &\boldsymbol{A} := \{A(i,j)\} \in R^{N\times N}, \qquad i,j = 0,1,\cdots,N-1,\\ &A(i,j) := \begin{cases} a_0 & \text{when } i = j\\ a_1 & \text{when } (i-j) \bmod_N = \pm 1\\ a_2 & \text{when } (i-j) \bmod_N = \pm 2\\ 0 & \text{otherwise}\end{cases}\\ &a_0 := -g_0 - 2(g_1+g_2), \quad a_1 := g_1, \quad a_2 := g_2,\\ &\boldsymbol{B} := \{B(i,j)\} \in R^{N\times N}, \qquad i,j = 0,1,\cdots,N-1,\\ &B(i,j) := \begin{cases} b_0 & \text{when } i = j\\ b_1 & \text{when } (i-j) \bmod_N = \pm 1\\ b_2 & \text{when } (i-j) \bmod_N = \pm 2\\ 0 & \text{otherwise}\end{cases}\\ &b_0 := c_0 + 2(c_1+c_2), \quad b_1 := -c_1, \quad b_2 := -c_2.\end{aligned}$$

Thus, for a general m, matrixes $\boldsymbol{A}$ and $\boldsymbol{B}$ are of the forms

$$\begin{aligned}&\boldsymbol{A} := \{A(i,j)\} \in R^{N\times N}, \qquad i,j = 0,1,\cdots,N-1,\\ &A(i,j) := \begin{cases} a_k & \text{when } (i-j) \bmod_N = \pm k, \quad k = 0,\cdots,m\\ 0 & \text{otherwise}\end{cases}\end{aligned}$$

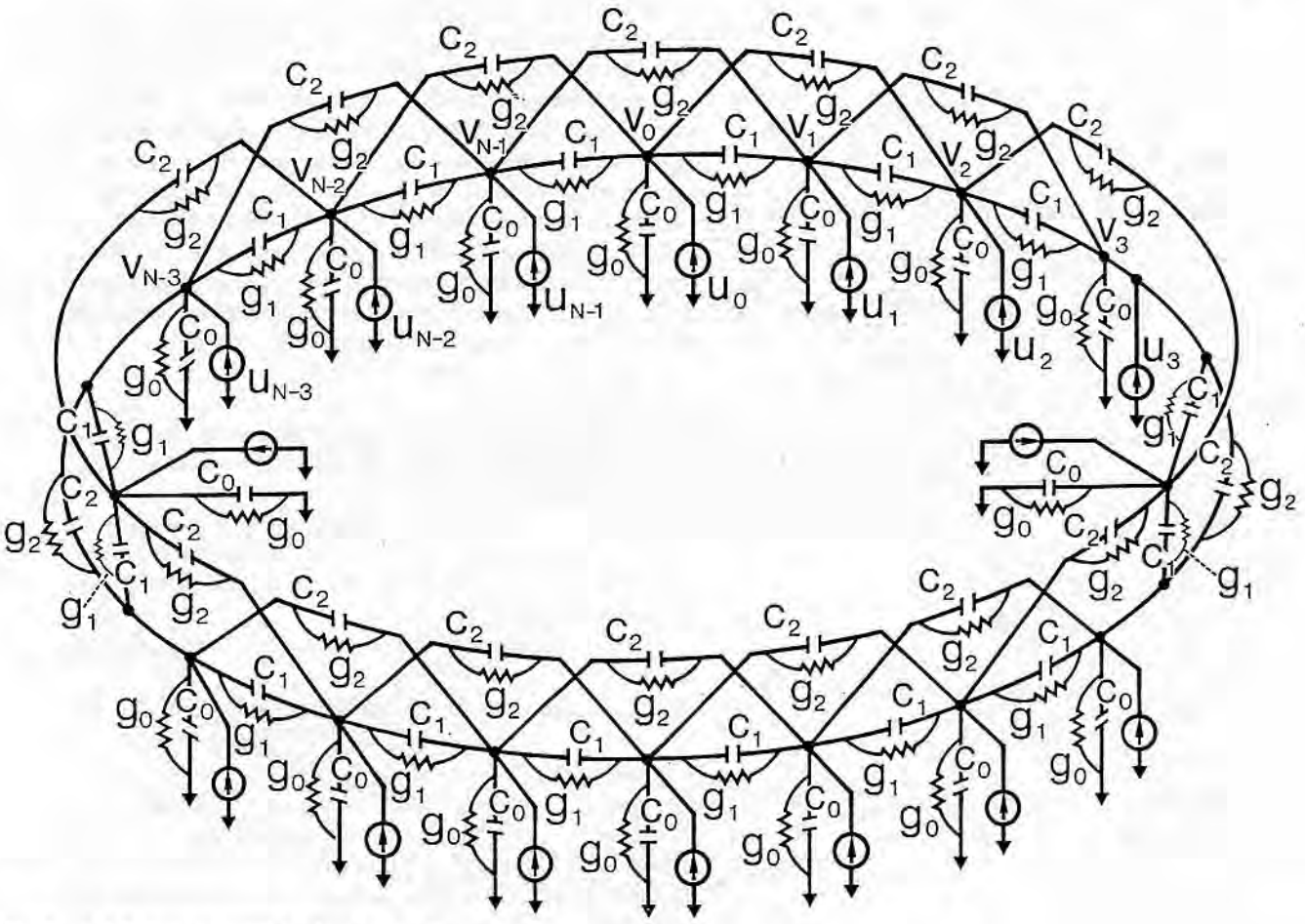


Fig. 2. A 1-D circulant network where $m = 2$.

$$a_0 := -g_0 - 2\sum_{p=1}^{m} g_p, \quad a_p := g_p \, (p = 1, 2, \cdots, m) \quad (2.1)$$

$$\boldsymbol{B} := \{B(i,j)\} \in R^{N\times N}, \qquad i,j = 0, 1, \cdots, N-1,$$

$$B(i,j) := \begin{cases} b_k & \text{when } (i-j) \bmod_N = \pm k, \quad k = 0, \cdots, m \\ 0 & \text{otherwise} \end{cases}$$

$$b_0 := c_0 + 2\sum_{p=1}^{m} c_p, \quad b_p := -c_p \, (p = 1, 2, \cdots, m). \quad (2.2)$$

The name "circulant network" comes from the fact that matrixes of the forms (2.1) and (2.2) are called (symmetric) circulant [9]. If nodes zero and $N-1$ are disconnected and the disconnected resistors are connected to ground [3], (2.1) and (2.2) should be replaced by

$$\boldsymbol{A} := \{A(i,j)\} \in R^{N\times N}, \qquad i,j = 0, 1, \cdots, N-1,$$

$$A(i,j) := \begin{cases} a_k & \text{when } i-j = \pm k, \quad k = 0, \cdots, m \\ 0 & \text{otherwise} \end{cases}$$

$$\boldsymbol{B} := \{B(i,j)\} \in R^{N\times N}, \qquad i,j = 0, 1, \cdots, N-1,$$

$$B(i,j) := \begin{cases} b_k & \text{when } i-j = \pm k, \quad k = 0, \cdots, m \\ 0 & \text{otherwise} \end{cases}$$

which are uniform band matrixes, and the corresponding network is called a uniform band network. The networks discussed in [1], [3], and [10] are of this type.

B. Block Circulant Networks

Now let us consider a 2-D network with $N_1 \times N_2$ nodes numbered (0, 0) through $(N_1 - 1, N_2 - 1)$, where each node (k_1, k_2) is excited by a current source u_{k_1,k_2} and has a (possibly negative) conductance g_0 and a capacitance c_0 to ground, and (possibly negative) conductances g_{l_1,l_2} and capacitances c_{l_1,l_2} to nodes $(k_1 + l_1, k_2 + l_2)$ for $l_1 = 0, \pm 1, \pm 2, \cdots \pm m_1, l_2 = 0, \pm 1, \pm 2, \cdots \pm m_2$ except for $l_1 = l_2 = 0$. Note that $g_{l_1,l_2} = g_{-l_1,-l_2}$ and $c_{l_1,l_2} = c_{-l_1,-l_2}$ because node (k_1, k_2) connects to node (k_1+l_1, k_2+l_2) with g_{l_1,l_2} and c_{l_1,l_2} whereas node (k_1+l_1, k_2+l_2) connects to node $((k_1+l_1)-l_1, (k_2+l_2)-l_2))$, i.e., node (k_1, k_2) with $g_{-l_1,-l_2}$ and $c_{-l_1,-l_2}$ and hence $g_{l_1,l_2} = g_{-l_1,-l_2}$ and $c_{l_1,l_2} = c_{-l_1,-l_2}$. The network is said to be block circulant if the rightmost and leftmost nodes are connected together and the top and bottom nodes are connected together, and thus the network is of a torus structure. Fig. 3 shows a block circulant network where $m_1 = m_2 = 1$. Since the KCL equation at node (k_1, k_2) reads

$$\begin{aligned} \frac{d}{dt}\{&(c_0 + 2c_{1,1} + 2c_{1,0} + 2c_{1,-1} + 2c_{0,1})v_{k_1,k_2} \\ &- c_{1,1}(v_{k_1+1,k_2+1} + v_{k_1-1,k_2-1}) \\ &- c_{1,0}(v_{k_1+1,k_2} + v_{k_1-1,k_2}) \\ &- c_{1,-1}(v_{k_1+1,k_2-1} + v_{k_1-1,k_2+1}) \\ &- c_{0,1}(v_{k_1,k_2+1} + v_{k_1,k_2-1})\} \end{aligned}$$

$$
\begin{aligned}
&= -(g_0 + 2g_{1,1} + 2g_{1,0} + 2g_{1,-1} + 2g_{0,1})v_{k_1,k_2} \\
&\quad + g_{1,1}(v_{k_1+1,k_2+1} + v_{k_1-1,k_2-1}) \\
&\quad + g_{1,0}(v_{k_1+1,k_2} + v_{k_1-1,k_2}) \\
&\quad + g_{1,-1}(v_{k_1+1,k_2-1} + v_{k_1-1,k_2+1}) \\
&\quad + g_{0,1}(v_{k_1,k_2+1} + v_{k_1,k_2-1}) + u_{k_1,k_2}
\end{aligned}
$$

so that the state equation of the network for general m_1 and m_2 is given by

$$
\boldsymbol{B_b}\frac{d}{dt}\boldsymbol{v_b} = \boldsymbol{A_b}\boldsymbol{v_b} + \boldsymbol{u_b} \tag{2.3}
$$

where

$$
\begin{aligned}
&\boldsymbol{v_b} := \{v'_k\}^T \in \mathcal{R}^{N_1N_2}, \quad \boldsymbol{u_b} := \{u'_k\}^T \in \mathcal{R}^{N_1N_2}, \\
&\qquad k = 0, 1, \cdots, N_1N_2 - 1, \\
&v'_k := v_{k_1,k_2}, \quad u'_k := u_{k_1,k_2}, \quad k := k_1N_2 + k_2, \\
&\qquad k_1 = 0, 1, \cdots, N_1 - 1, \\
&\qquad k_2 = 0, 1, \cdots, N_2 - 1, \\
&\boldsymbol{A_b} := \{A(i,j)\} \in \mathcal{R}^{N_1N_2 \times N_1N_2}, \\
&\qquad i, j = 0, 1, \cdots, N_1 - 1, \\
&A(i,j) := \begin{cases} \boldsymbol{A}_l & \text{when } (i-j) \bmod_{N_1} = l, \\ & \qquad l = 0, \pm 1, \cdots, \pm m_1 \\ 0 & \text{otherwise} \end{cases} \in R^{N_2 \times N_2}
\end{aligned}
$$

and for $l = 0, \pm 1, \pm 2, \cdots, \pm m_1$

$$
\begin{aligned}
&\boldsymbol{A}_l := \{A_l(i,j)\} \in \mathcal{R}^{N_2 \times N_2}, \qquad i, j = 0, 1, \cdots, N_2 - 1, \\
&\boldsymbol{A}_l := \begin{cases} a_{l,k} & \text{when } (i-j) \bmod_{N_2} = k, \\ & \qquad k = 0, \pm 1, \cdots, \pm m_2 \\ 0 & \text{otherwise} \end{cases} \\
&a_{0,0} := -\sum_{p=-m_1}^{m_1} \sum_{q=-m_2}^{m_2} g_{p,q}, \qquad g_{0,0} := g_0, \\
&a_{p,q} := g_{p,q}, \quad a_{p,q} = a_{-p,-q}, \quad (p = 0, \pm 1, \pm 2, \cdots, \pm m_1; \\
&\qquad q = 0, \pm 1, \pm 2, \cdots, \pm m_2; \quad (p,q) \neq (0,0)), \\
&\boldsymbol{B_b} := \{\boldsymbol{B}(i,j)\} \in \mathcal{R}^{N_1N_2 \times N_1N_2}, \\
&\qquad i, j = 0, 1, \cdots, N_1 - 1, \\
&\boldsymbol{B}(i,j) := \begin{cases} \boldsymbol{B}_l & \text{when } (i-j) \bmod_{N_1} = l, \\ & \qquad l = 0, \pm 1, \cdots, \pm m_1 \\ 0 & \text{otherwise} \end{cases} \in R^{N_2 \times N_2}
\end{aligned} \tag{2.4}
$$

and for $l = 0, \pm 1, \pm 2, \cdots, \pm m_1$

$$
\begin{aligned}
&\boldsymbol{B}_l := \{B_l(i,j)\} \in \mathcal{R}^{N_2 \times N_2}, \\
&\qquad i, j = 0, 1, \cdots, N_2 - 1, \\
&B_l(i,j) := \begin{cases} b_{l,k} & \text{when } (i-j) \bmod_{N_2} = k, \\ & \qquad k = 0, \pm 1, \cdots, \pm m_2 \\ 0 & \text{otherwise} \end{cases} \\
&b_{0,0} := \sum_{p=-m_1}^{m_1} \sum_{q=-m_2}^{m_2} c_{p,q}, \qquad c_{0,0} := c_0, \\
&b_{p,q} := -c_{p,q}, \quad b_{p,q} = b_{-p,-q}, \quad (p = 0, \pm 1, \pm 2, \cdots, \\
&\qquad \pm m_1; \quad q = 0, \pm 1, \pm 2, \cdots, \pm m_2; \\
&\qquad (p,q) \neq (0,0)).
\end{aligned} \tag{2.5, 2.6}
$$

Matrixes $\boldsymbol{A_b}$ and $\boldsymbol{B_b}$ are called (symmetric) block circulant with circulant blocks [9].

Circulant and block circulant matrixes enjoy many interesting properties [9] and we will fully exploit these properties. Note that a network being circulant or block circulant is equivalent to its boundary conditions being periodic. We would like to emphasize that boundary conditions are critical for the class of problems we are discussing. Namely, even if the eigenvalue conditions are satisfied for the spatial dynamics, inappropriate boundary conditions may force the solution to blow up (see [1]). It is extremely difficult, if not impossible, to obtain analytical conditions for stability as well as other properties for uniform block networks, while several interesting analytical results can be obtained for block circulant networks.

On the other hand, when $\boldsymbol{A}_l$ and $\boldsymbol{B}_l$ $(l = 0, \pm 1, \cdots, \pm m_1)$ are uniform band matrixes and they are assigned in $\boldsymbol{A_b}$ and $\boldsymbol{B_b}$ in uniform bands, respectively, matrixes $\boldsymbol{A_b}$ and $\boldsymbol{B_b}$ are called uniform block and the corresponding network is said to be a uniform block network.

We further remark that all the results for 2-D networks described in this paper are also valid in 1-D cases by specifying $N_2 = 1$ and $m_2 = 0$ and the results are consistent with those for 1-D networks in [1].

Standing Assumptions: We will hereafter assume the following:

i) Capacitance matrix $\boldsymbol{B_b}$ is positive definite for all N_1, N_2, and

ii) $a_{0,0} < 0$.

Assumption i) is very mild because, for example, if all capacitances $c_0, c_{0,1}, \cdots c_{m_1,m_2}$ are positive, $\boldsymbol{B_b}$ is positive definite. The reason for assuming positive definiteness of $\boldsymbol{B_b}$ for all N_1 and N_2 is to derive analytical stability conditions as is done in [1]. Regarding assumption ii), if $a_{0,0} > 0$, the system matrix $\boldsymbol{A_b}$ cannot be negative definite, and the network is always temporally unstable and is therefore unusable.

III. Spatial Dynamics

Note that the distribution of the node voltage $\boldsymbol{v}$ in (2.3) at an equilibrium is given by

$$
\boldsymbol{A_b}\boldsymbol{v_b} + \boldsymbol{u_b} = 0. \tag{3.1}
$$

Let us first consider the spatial impulse response. If the input current is injected at node (0, 0) while no currents are injected at the other nodes, then the resultant voltage distribution is called the spatial impulse response of the network. Recall that the spatial impulse response completely characterizes a linear spatial filter network because it is identical to the convolution kernel such that any output is obtained by convolution of the input with the kernel. Suppose that the system matrix $\boldsymbol{A_b}$ is nonsingular and the input $\boldsymbol{u_b}$ is a spatial impulse $\boldsymbol{\delta_b}$

$$
\boldsymbol{\delta_b} := (1, 0, 0, \cdots, 0)^T \in \mathcal{R}^{N_1N_2}.
$$

Then the spatial impulse response is obtained by

$$
\boldsymbol{v_b} = -\boldsymbol{A_b}^{-1}\boldsymbol{\delta_b}. \tag{3.2}
$$

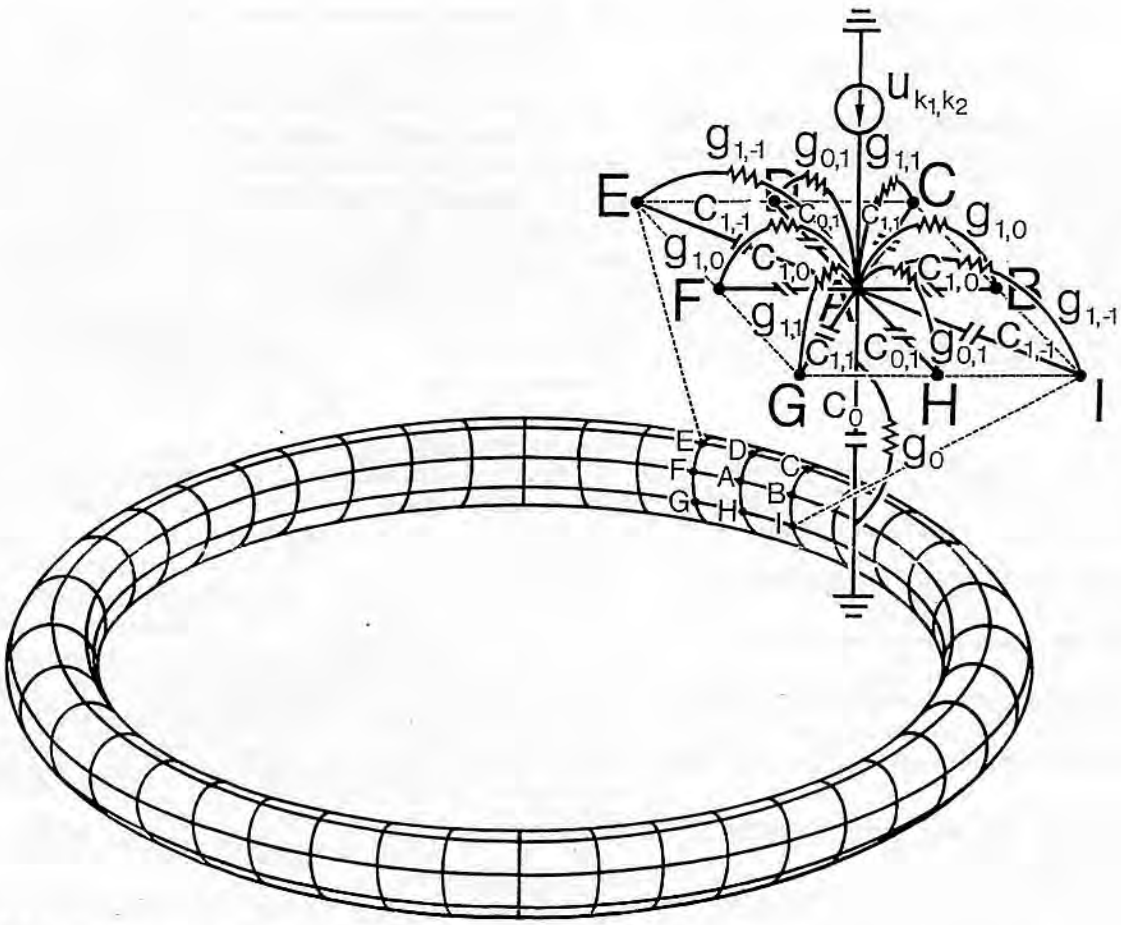


Fig. 3. A 2-D block circulant network where $m_1 = m_2 = 1$.

Proposition 1: Assume that $\boldsymbol{A_b}$ is nonsingular and let

$$\boldsymbol{c}_{N_1,k_1,N_2,k_2} := (\boldsymbol{c}_{N_1,k_1}^{N_2,0}, \boldsymbol{c}_{N_1,k_1}^{N_2,k_2}, \boldsymbol{c}_{N_1,k_1}^{N_2,2k_2}, \cdots \boldsymbol{c}_{N_1,k_1}^{N_2,(N_2-1)k_2})^T \tag{3.3}$$

where

$$\begin{aligned}\boldsymbol{c}_{N_1,k_1}^{N_2,k_2} := \Bigg(&\cos\left(2\pi\frac{k_2}{N_2}\right), \cos\left(2\pi\left(\frac{k_1}{N_1}+\frac{k_2}{N_2}\right)\right),\\ &\cos\left(2\pi\left(\frac{2k_1}{N_1}+\frac{k_2}{N_2}\right)\right), \cdots,\\ &\cos\left(2\pi\left(\frac{(N_1-1)k_1}{N_1}+\frac{k_2}{N_2}\right)\right)\Bigg)\end{aligned} \tag{3.4}$$

$k_1 = 0, 1, 2, \cdots, N_1 - 1$ and $k_2 = 0, 1, 2, \cdots, N_2 - 1$. Then the spatial impulse response (3.2) has the following explicit representation

$$\begin{aligned}\boldsymbol{v_b} = -\boldsymbol{A_b}^{-1}\boldsymbol{\delta_b} = \frac{-1}{N_1N_2}\times\\ \sum_{k_1=0}^{N_1-1}\sum_{k_2=0}^{N_2-1}\frac{1}{\sum_{p=-m_1}^{m_1}\sum_{q=-m_2}^{m_2} a_{p,q}\cos\left(2\pi\left(\frac{k_1p}{N_1}+\frac{k_2q}{N_2}\right)\right)}\\ \cdot \boldsymbol{c}_{N_1,k_1,N_2,k_2}.\end{aligned} \tag{3.5}$$

Proof: To prove Proposition 1, note that

$$\begin{aligned}&\frac{1}{N_1N_2}\sum_{k_1=0}^{N_1-1}\sum_{k_2=0}^{N_2-1}\cos\left(2\pi\left(\frac{k_1n_1}{N_1}+\frac{k_2n_2}{N_2}\right)\right)\\ &\quad = \begin{cases}1 & (n_1 = n_2 = 0)\\ 0 & \text{otherwise.}\end{cases}\end{aligned}$$

Thus the input impulse is represented by

$$\boldsymbol{\delta_b} = \frac{1}{N_1N_2}\sum_{k_1=0}^{N_1-1}\sum_{k_2=0}^{N_2-1}\boldsymbol{c}_{N_1,k_1,N_2,k_2}$$

and hence it follows from Lemma 1 (see Appendix) that

$$\begin{aligned}\boldsymbol{v_b} &= \frac{-1}{N_1N_2}\boldsymbol{A_b}^{-1}\sum_{k_1=0}^{N_1-1}\sum_{k_2=0}^{N_2-1}\boldsymbol{c}_{N_1,k_1,N_2,k_2}\\ &= \frac{-1}{N_1N_2}\sum_{k_1=0}^{N_1-1}\sum_{k_2=0}^{N_2-1}\frac{1}{\lambda_{N_1,k_1,N_2,k_2}}\boldsymbol{c}_{N_1,k_1,N_2,k_2}.\end{aligned} \tag{3.6}$$

□

We would like to point out that the spatial impulse response of a block circulant network is closely related to the discrete Fourier transform (DFT) and the inverse discrete Fourier transform (IDFT) of circuit parameters. Recall that the spatial

impulse response of a block circulant network is given by

$$\boldsymbol{v_b} = -\boldsymbol{A_b}^{-1}\boldsymbol{\delta_b}$$

provided that $\boldsymbol{A_b}$ is a nonsingular circulant matrix. Then $-\boldsymbol{A_b}^{-1}\boldsymbol{\delta_b}$ can be obtained by the 2-D DFT of circuit parameters using the following steps.

Step 1) Compute the eigenvalues of $\boldsymbol{A_b}$, $\lambda_{N_1,0,N_2,0}$, $\lambda_{N_1,0,N_2,1}$, $\cdots$ $\lambda_{N_1,N_1-1,N_2,N_2-1}$ with the 2-D IDFT of parameters

$$\{\overbrace{\boldsymbol{a}_0, \boldsymbol{a}_1, \cdots, \boldsymbol{a}_{m_1}, \boldsymbol{0}, \cdots, \boldsymbol{0}, \boldsymbol{a}_{-m_1}, \cdots, \boldsymbol{a}_{-1}}^{N_1}\}$$

where

$$\boldsymbol{a_k} = (\overbrace{a_{k,0}, a_{k,1}, \cdots, a_{k,m_2}, 0, \cdots, 0, a_{k,-m_2}, \cdots, a_{k,-1}}^{N_2})^T$$
$$k = 0, \pm 1, \pm 2, \cdots, \pm m_1.$$

Step 2) Obtain the reciprocal of the eigenvalues, i.e., $\{\lambda^{-1}_{N_1,0,N_2,0}, \lambda^{-1}_{N_1,0,N_2,1}, \cdots, \lambda^{-1}_{N_1,N_1-1,N_2,N_2-1}\}$.

Step 3) Compute the spatial impulse response by the 2-D DFT of $\{\lambda^{-1}_{N_1,0,N_2,0}, \lambda^{-1}_{N_1,0,N_2,1}, \cdots, \lambda^{-1}_{N_1,N_1-1,N_2,N_2-1}\}$.

Note that the 2-D DFT and IDFT are defined as follows [11]: let $f(n_1, n_2) = f(n_1 + N_1, n_2 + N_2)$, where n_1 and n_2 are integers and let $F(k_1, k_2)$ be the DFT of $f(n_1, n_2)$. Then

$$F(k_1, k_2) := \sum_{n_1=0}^{N_1-1} \sum_{n_2=0}^{N_2-1} f(n_1, n_2) W_1^{n_1 k_1} W_2^{n_2 k_2}$$

where $W_1 := e^{-j(2\pi/N_1)}$, $W_2 := e^{-j(2\pi/N_2)}$, $F(k_1, k_2) = F(k_1 + N_1, k_2 + N_2)$, and k_1, k_2 are integers. Conversely, $f(n_1, n_2)$ is the IDFT of $F(k_1, k_2)$, and

$$f(n_1, n_2) = \frac{1}{N_1 N_2} \sum_{k_1=0}^{N_1-1} \sum_{k_2=0}^{N_2-1} F(k_1, k_2) W_1^{-n_1 k_1} W_2^{-n_2 k_2}.$$

We also remark that in general the 2-D explicit impulse response is very difficult to obtain [12].

There is an interesting way of looking at (3.5). Consider the input $\boldsymbol{u_b}$ given by

$$\boldsymbol{u_b} = \alpha \boldsymbol{c}_{N_1,k_1,N_2,k_2} + \beta \boldsymbol{s}_{N_1,k_1,N_2,k_2}, \quad \text{where } \alpha, \beta \text{ are constants.} \tag{3.7}$$

Then it follows from Lemma 1 that

$$\begin{aligned}\boldsymbol{v_b} &= -\boldsymbol{A_b}^{-1}(\alpha \boldsymbol{c}_{N_1,k_1,N_2,k_2} + \beta \boldsymbol{s}_{N_1,k_1,N_2,k_2}) \\ &= \frac{-1}{\lambda_{N_1,k_1,N_2,k_2}}(\alpha \boldsymbol{c}_{N_1,k_1,N_2,k_2} + \beta \boldsymbol{s}_{N_1,k_1,N_2,k_2}) \end{aligned} \tag{3.8}$$

where $\boldsymbol{c}_{N_1,k_1,N_2,k_2}, \boldsymbol{s}_{N_1,k_1,N_2,k_2}$ and $\lambda_{N_1,k_1,N_2,k_2}$ are given by (7.2), (7.3), and (7.1), respectively. If we regard (3.7) as a spatially periodic input function with an angle frequency $(2\pi k_1/N_1, 2\pi k_2/N_2)$, then (3.8) is the spatial frequency response of the network. Thus (3.8) can be interpreted in the following manner: for a spatially periodic input with a spatial angle frequency $(2\pi k_1/N_1, 2\pi k_2/N_2)$.

1) The gain of the spatial frequency response is $|1/\lambda_{N_1,k_1,N_2,k_2}|$, while

2) The phase of the spatial frequency response is

$$\begin{cases} 0 & \text{when } \lambda_{N_1,k_1,N_2,k_2} < 0 \\ \pi & \text{when } \lambda_{N_1,k_1,N_2,k_2} > 0. \end{cases}$$

Next, let us discuss the spatial stability of block circulant networks. Note first that Proposition 1 says that the spatial impulse response is well defined if $\boldsymbol{A_b}$ is nonsingular for a fixed network size $N_1 \times N_2$. Formula (3.5) naturally suggests that the invertibility of $\boldsymbol{A_b}$ is not enough for the spatial stability for all N_1, N_2. A problem arises when $\lambda_{N_1,k_1,N_2,k_2}$ gets smaller and smaller as the network size $N_1 \times N_2$ grows because then $\boldsymbol{v_b}$ blows up as $N_1 \times N_2$ grows, which is inappropriate for image processing purposes. To discuss the spatial stability issue valid for all N_1, N_2 as is done in [1], we will use the following definition:

Definition—Spatial Stability for a 2-D Network: The 2-D block circulant network described by (3.1) is spatially stable if, for any $\boldsymbol{u_b}$ bounded uniformly in N_1, N_2, there is a unique $\boldsymbol{v_b}$ which is bounded uniformly in N_1, N_2 and satisfies (3.1).

Of course, that $\boldsymbol{u_b}$ (respectively, $\boldsymbol{v_b}$) bounded uniformly in N_1, N_2 means $||\boldsymbol{u_b}||$ (respectively, $||\boldsymbol{v_b}||$) is bounded uniformly in N_1, N_2. Thus this definition means that if the input has a finite energy $||\boldsymbol{u_b}||^2$ bounded uniformly in N_1, N_2, the output energy $||\boldsymbol{v_b}||^2$ is also finite and bounded uniformly in N_1, N_2.

Proposition 2: A 2-D block circulant network described by (3.1) is spatially stable if and only if $\boldsymbol{A_b}$ is negative definite uniformly in N_1, N_2.

Proof: $\Leftarrow$) This is clear from the definition of uniform negative definiteness.

$\Rightarrow$) Suppose that $\boldsymbol{A_b}$ is not negative definite uniformly in N_1, N_2, then there are two cases.

Case 1) There are N_1, k_1, N_2, and k_2 with $\lambda_{N_1,k_1,N_2,k_2} > 0$.

Case 2) For an arbitrary $\epsilon > 0$, there are N_1, k_1, N_2 and k_2, which satisfy $|\lambda_{N_1,k_1,N_2,k_2}| < \epsilon$ even if $\lambda_{N_1,k_1,N_2,k_2} < 0$ for all N_1, k_1, N_2 and k_2.

Case 2) apparently contradicts the requirement for uniform invertibility. To check case 1, let

$$f(x_1, x_2) := \sum_{p=-m_1}^{m_1} \sum_{q=-m_2}^{m_2} a_{p,q} \cos(2\pi(px_1 + qx_2)). \tag{3.9}$$

Then the eigenvalues of $\boldsymbol{A_b}$ are given by $\lambda_{N_1,k_1,N_2,k_2} = f(k_1/N_1, k_2/N_2)$. Due to assumption ii) ($a_{0,0} < 0$), $\boldsymbol{A_b}$ cannot be positive definite and therefore $f(x_{1a}, x_{2a}) < 0$ for some x_{1a}, x_{2a}. Case 1 means that there are x_{1b} and x_{2b} such that $f(x_{1b}, x_{2b}) > 0$. Then since $f(x_1, x_2)$ is continuous, by virtue of the mean value theorem there are infinitely many N_1, k_1, N_2 and k_2 with $|f(k_1/N_1, k_2/N_2)| < \epsilon$. This also contradicts the requirement for uniform invertibility. □

The following fact gives an explicit if-and-only-if condition for spatial stability in terms of network parameters. The proof immediately follows from Proposition 2 and (7.1) and (3.9).

Proposition 3: The 2-D block circulant network described by (3.1) is spatially stable if and only if

$$\sigma_+ := \max_{x_1, x_2 \in [0,1] \times [0,1]} \sum_{p=-m_1}^{m_1} \sum_{q=-m_2}^{m_2} a_{p,q} \cdot \cos(2\pi(px_1 + qx_2)) < 0. \tag{3.10}$$

We call σ_+ a stability indicator function.

Remark: The spatial stability condition above is consistent with the classical stability condition for noncausal discrete variable IIR filters where the network extends indefinitely. To see this, recall that the classical stability condition demands that all poles of their transfer functions are off the unit circle [12]. For example, let us consider a 2-D network where the rightmost and leftmost nodes are disconnected, the top and bottom nodes are also disconnected, and $N_1, N_2 \to \infty$. The spatial transfer function of the 2-D network can be defined in terms of

$$\frac{V(z_1, z_2)}{U(z_1, z_2)} = \frac{-1}{H(z_1, z_2)}$$

where $U(z_1, z_2)$ and $V(z_1, z_2)$ are 2-D Z-transform of u_{k_1,k_2} and v_{k_1,k_2}, respectively, [11], and

$$H(z_1, z_2) := \sum_{p=-m_1}^{m_1} \sum_{q=-m_2}^{m_2} a_{p,q} z_1^p z_2^q.$$

If the transfer function has a pole on the unit circle $|z_1| = |z_2| = 1$, there is a 2-D spatial frequency (θ_1, θ_2) which satisfies $H(e^{j\theta_1}, e^{j\theta_2}) = 0$, i.e., $|1/H(e^{j\theta_1}, e^{j\theta_2})| = \infty$, which gives the reader an intuitive interpretation of spatial instability. Since $H(e^{j\theta_1}, e^{j\theta_2}) = \sigma_+$, it is clear that this violates the stability condition of our new definition. Conversely, if all of the transfer function poles are off the unit circle, $H(e^{j\theta_1}, e^{j\theta_2})$ is nonzero for all θ_1, θ_2, which means $H(e^{j\theta_1}, e^{j\theta_2}) < 0$ for all θ_1, θ_2 or $H(e^{j\theta_1}, e^{j\theta_2}) > 0$ for all θ_1, θ_2. Due to our standing assumption ii), the second possibility is excluded and it yields $H(e^{j\theta_1}, e^{j\theta_2}) = \sigma_+ < 0$ for all θ_1, θ_2, which satisfies our stability condition.

Example 1: Let us consider a 2-D network where $m_1 = m_2 = 1$ and interconnection conductances are g_0, $g_1 := g_{0,1} = g_{0,-1} = g_{1,0} = g_{-1,0} = g_{1,1} = g_{-1,-1} = g_{1,-1} = g_{-1,1}$ (Fig. 4). The stability indicator function σ_+ is given by

$$\begin{aligned}
\sigma_+ = &\max_{x_1,x_2\in[0,1]\times[0,1]} \{-(g_0 + 8g_1) + 2g_1(\cos(2\pi x_1) \\
&+ \cos(2\pi x_2) + \cos(2\pi(x_1 + x_2)) \\
&+ \cos(2\pi(x_1 - x_2)))\} \\
= &\begin{cases} -(g_0 + 8g_1) + 8g_1 & \text{when } g_1 \geq 0 \\ & \text{at } x_1 = x_2 = 0 \\ -(g_0 + 8g_1) - 4g_1 & \text{when } g_1 < 0 \\ & \text{at } x_1 = 0, x_2 = 0.5 \\ & \text{or } x_1 = 0.5, x_2 = 0 \end{cases} \\
= &\begin{cases} -g_0 & \text{when } g_1 \geq 0 \\ & \text{at } x_1 = x_2 = 0 \\ -g_0 - 12g_1 & \text{when } g_1 < 0 \\ & \text{at } x_1 = 0, x_2 = 0.5 \\ & \text{or } x_1 = 0.5, x_2 = 0. \end{cases}
\end{aligned} \tag{3.11}$$

Example 2: Let us consider a 2-D block circulant network with interconnection conductances $g_0, g_{1h} := g_{0,1} = g_{0,-1}$, $g_{1v} := g_{1,0} = g_{-1,0}$, $g_{2h} := g_{0,2} = g_{0,-2}$ and $g_{2v} := g_{2,0} = g_{-2,0}$

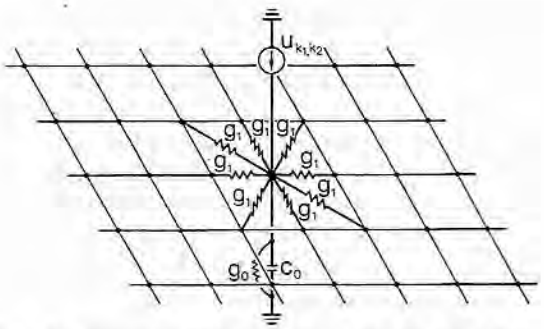


Fig. 4. A 2-D network of Examples 1 and 3. Only one unit is shown.

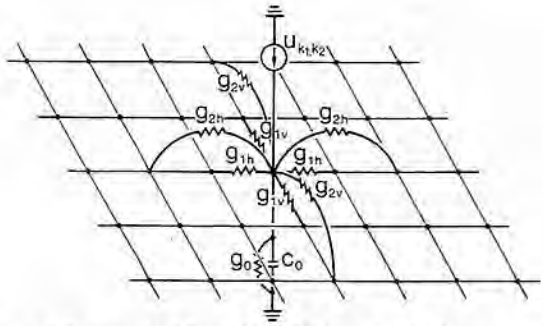


Fig. 5. A 2-D network of Example 2. Only one unit is shown.

(Fig. 5). The stability indicator function σ_+ is given by

$$\begin{aligned}
\sigma_+ = &\max_{x_1,x_2\in[0,1]\times[0,1]} \{-(g_0 + 2g_{1h} + 2g_{1v} + 2g_{2h} + 2g_{2v}) \\
&+ 2g_{1h}\cos(2\pi x_1) + 2g_{1v}\cos(2\pi x_2) \\
&+ 2g_{2h}\cos(4\pi x_1) + 2g_{2v}\cos(4\pi x_2)\} \\
= &-(g_0 + 2g_{1h} + 2g_{1v} + 2g_{2h} + 2g_{2v}) \\
&+ 2\max_{x_1\in[0,1]} \{g_{1h}\cos(2\pi x_1) + g_{2h}\cos(4\pi x_1)\} \\
&+ 2\max_{x_2\in[0,1]} \{g_{1v}\cos(2\pi x_2) + g_{2v}\cos(4\pi x_2)\}
\end{aligned}$$

where

$$\begin{aligned}
&2\max_{x_1\in[0,1]} \{g_{1h}\cos(2\pi x_1) + g_{2h}\cos(4\pi x_1)\} \\
&= \begin{cases} 2|g_{1h}| + 2g_{2h} & \text{when } g_{2h} \geq 0 \\ & \text{or } \left(g_{2h} < 0 \text{ and } \left|\frac{g_{1h}}{g_{2h}}\right| \geq 4\right) \\ -2g_{2h} - \frac{g_{1h}^2}{4g_{2h}} & \text{when } g_{2h} < 0 \\ & \text{and } \left|\frac{g_{1h}}{g_{2h}}\right| \leq 4 \end{cases}
\end{aligned}$$

and

$$\begin{aligned}
&2\max_{x_2\in[0,1]} \{g_{1v}\cos(2\pi x_2) + g_{2v}\cos(4\pi x_2)\} \\
&= \begin{cases} 2|g_{1v}| + 2g_{2v} & \text{when } g_{2v} \geq 0 \\ & \text{or } \left(g_{2v} < 0 \text{ and } \left|\frac{g_{1v}}{g_{2v}}\right| \geq 4\right) \\ -2g_{2v} - \frac{g_{1v}^2}{4g_{2v}} & \text{when } g_{2v} < 0 \\ & \text{and } \left|\frac{g_{1v}}{g_{2v}}\right| \leq 4. \end{cases}
\end{aligned}$$

Since the 2-D stability condition (3.10) is completely analytical and simple, it will be very useful to check the spatial stability of the network. One has to remember, however, that this result is for block circulant networks instead of uniform block networks. Thus, the 2-D stability test (3.10) will be of little value unless one shows that solutions for block circulant networks behave in a manner similar to those for uniform block networks. The following is an important fact which says that if a block circulant network is spatially stable, then as the network size $N_1 \times N_2$ grows, the impulse response far from the node where current is injected tends to zero so that the response behaves in a manner similar to that of a "properly" behaved solution of a uniform block network.

Proposition 4: Suppose that a block circulant network is spatially stable and current is injected to node (0, 0) while no currents are applied to the other nodes.

1) When $N_1 \to \infty$, the response v_{n_1,n_2} approaches zero for node (n_1, n_2) far from node (0, 0) with respect to the n_1-coordinate. In other words, for arbitrary $c_1 \geq 0$ and $\epsilon_1 > 0$, there is an $M > 0$ such that $|v_{n_1,n_2}| < \epsilon_1$ for all $N_1 > M$, where

$$\begin{aligned} \frac{N_1}{2} - c_1 \leq n_1 \leq \frac{N_1}{2} + c_1 \quad (n_1\text{: even}), \\ \frac{N_1 - 1}{2} - c_1 \leq n_1 \leq \frac{N_1 + 1}{2} + c_1 \quad (n_1\text{: odd}), \\ 0 \leq n_2 \leq N_2 - 1. \end{aligned}$$

2) When $N_2 \to \infty$, the response v_{n_1,n_2} approaches zero for node (n_1, n_2) far from node (0, 0) with respect to the n_2-coordinate. In other words, for arbitrary $c_2 \geq 0$ and $\epsilon_2 > 0$, there is an $M > 0$ such that $|v_{n_1,n_2}| < \epsilon_2$ for all $N_2 > M$, where

$$\begin{aligned} 0 \leq n_1 \leq N_1 - 1, \quad \frac{N_2}{2} - c_2 \leq n_2 \\ \leq \frac{N_2}{2} + c_2 \quad (n_2\text{: even}), \\ \frac{N_2 - 1}{2} - c_2 \leq n_2 \leq \frac{N_2 + 1}{2} + c_2 \quad (n_2\text{: odd}). \end{aligned}$$

3) When $N_1, N_2 \to \infty$, the response v_{n_1,n_2} approaches zero for node (n_1, n_2) far from node (0, 0) with respect to the n_1-coordinate and the n_2-coordinate. In other words, for arbitrary $c_1, c_2 \geq 0$ and $\epsilon > 0$, there is an $M > 0$ such that $|v_{n_1,n_2}| < \epsilon$ for all $N_1, N_2 > M$, where

$$\begin{aligned} \frac{N_1}{2} - c_1 \leq n_1 \leq \frac{N_1}{2} + c_1 \quad (n_1\text{: even}), \\ \frac{N_1 - 1}{2} - c_1 \leq n_1 \leq \frac{N_1 + 1}{2} + c_1 \quad (n_1\text{: odd}), \\ 0 \leq n_2 \leq N_2 - 1 \end{aligned}$$

or

$$\begin{aligned} 0 \leq n_1 \leq N_1 - 1, \quad \frac{N_2}{2} - c_2 \leq n_2 \\ \leq \frac{N_2}{2} + c_2 \quad (n_2\text{: even}), \\ \frac{N_2 - 1}{2} - c_2 \leq n_2 \leq \frac{N_2 + 1}{2} + c_2 \quad (n_2\text{: odd}). \end{aligned}$$

Proof: Recall $f(x_1, x_2)$ defined by (3.9) and let

$$g(x_1, x_2, x_3, x_4) := \frac{\cos(2\pi(x_3 + x_4))}{f(x_1, x_2)}.$$

If the network is spatially stable, $f(x_1, x_2)$ is uniformly continuous and bounded away from zero so that $1/f(x_1, x_2)$ is uniformly continuous in $x_1, x_2 \in [0,1] \times [0,1]$. Then $g(x_1, x_2, x_3, x_4)$ is also uniformly continuous in $x_1, x_2, x_3, x_4 \in [0,1] \times [0,1] \times [0,1] \times [0,1]$. Hence, for an arbitrary $\epsilon > 0$ there is a $\delta > 0$ which satisfies

$$\begin{aligned} |g(x_1, x_2, x_3, x_4) - g(x_1 + \Delta x_1, x_2 + \Delta x_2, x_3 \\ + \Delta x_3, x_4 + \Delta x_4)| < \epsilon \end{aligned} \tag{3.12}$$

for all $x_1, x_2, x_3, x_4 \in [0,1] \times [0,1] \times [0,1] \times [0,1]$, $|\Delta x_1| < \delta, |\Delta x_2| < \delta, |\Delta x_3| < \delta$ and $|\Delta x_4| < \delta$. Also, if the network is stable, there is a $\xi > 0$ so that

$$|g(x_1, x_2, x_3, x_4)| < \xi \tag{3.13}$$

for all $x_1, x_2, x_3, x_4 \in [0,1] \times [0,1] \times [0,1] \times [0,1]$. Letting

$$a_1 := n_1 - \frac{N_1}{2}, \quad a_2 := n_2 - \frac{N_2}{2}$$

we obtain the impulse response at node (n_1, n_2) from (3.5)

$$\begin{aligned} v_{n_1,n_2} &= \frac{-1}{N_1 N_2} \sum_{k_1=0}^{N_1-1} \sum_{k_2=0}^{N_2-1} \frac{1}{\lambda_{N_1,k_1,N_2,k_2}} \\ &\quad \cdot \cos\left(2\pi\left(\frac{k_1 n_1}{N_1} + \frac{k_2 n_2}{N_2}\right)\right) \\ &= \frac{-1}{N_1 N_2} \sum_{k_1=0}^{N_1-1} \sum_{k_2=0}^{N_2-1} \frac{1}{f(k_1/N_1, k_2/N_2)} \\ &\quad \cdot \cos\left(2\pi\left(\frac{k_1}{N_1}\left(a_1 + \frac{N_1}{2}\right)\right.\right. \\ &\quad \left.\left. + \frac{k_2}{N_2}\left(a_2 + \frac{N_2}{2}\right)\right)\right) \\ &= \frac{-1}{N_1 N_2} \sum_{k_1=0}^{N_1-1} \sum_{k_2=0}^{N_2-1} \frac{(-1)^{k_1+k_2}}{f(k_1/N_1, k_2/N_2)} \\ &\quad \cdot \cos\left(2\pi\left(\frac{k_1 a_1}{N_1} + \frac{k_2 a_2}{N_2}\right)\right) \\ &= \frac{-1}{N_1 N_2} \sum_{k_1=0}^{N_1-1} \sum_{k_2=0}^{N_2-1} (-1)^{k_1+k_2} \\ &\quad g\left(\frac{k_1}{N_1}, \frac{k_2}{N_2}, \frac{k_1 a_1}{N_1}, \frac{k_2 a_2}{N_2}\right). \end{aligned} \tag{3.14}$$

We will prove only i) of Proposition 4, since proofs of ii) and iii) are similar.

When N_1 Is Even: From (3.14) we obtain

$$\begin{aligned} |v_{n_1,n_2}| \leq \frac{1}{N_1 N_2} \sum_{l_1=0}^{N_1/2-1} \sum_{k_2=0}^{N_2-1} \left| g\left(\frac{2l_1}{N_1}, \frac{k_2}{N_2}, \frac{2l_1}{N_1} a_1, \frac{k_2}{N_2} a_2\right)\right. \\ \left. - g\left(\frac{2l_1 + 1}{N_1}, \frac{k_2}{N_2}, \frac{2l_1 + 1}{N_1} a_1, \frac{k_2}{N_2} a_2\right)\right|. \end{aligned} \tag{3.15}$$

Noting that

$$\frac{N_1}{2} - c_1 \le n_1 \le \frac{N_1}{2} + c_1 \Leftrightarrow -c_1 \le a_1 \le c_1 \tag{3.16}$$

it follows from (3.12) that

$$\left| g(x_1, x_2, x_3, x_4) - g\left(x_1 + \frac{1}{N_1}, x_2, x_3 + \frac{a_1}{N_1}, x_4\right)\right| < \epsilon \tag{3.17}$$

for all N_1 with $N_1 > \max(c_1/\delta, 1/\delta)$ and all a_1, which satisfies (3.16). Then from (3.15) and (3.17) we obtain

$$|v_{n_1,n_2}| < \frac{1}{N_1 N_2}\frac{N_1}{2} N_2 \epsilon = \frac{\epsilon}{2} < \epsilon.$$

When N_1 is odd:

From (3.14) we obtain

$$\begin{aligned} |v_{n_1,n_2}| \le \frac{1}{N_1 N_2} \sum_{k_2=0}^{N_2-1} \Bigg[\Bigg\{ \sum_{l_1=0}^{(N_1-3)/2} & \cdot \left| g\left(\frac{2l_1}{N_1}, \frac{k_2}{N_2}, \frac{2l_1 a_1}{N_1}, \frac{k_2 a_2}{N_2}\right) \right. \\ & \left. - g\left(\frac{2l_1+1}{N_1}, \frac{k_2}{N_2}, \frac{2l_1+1}{N_1} a_1, \frac{k_2 a_2}{N_2}\right)\right| \Bigg\} \\ & + \left| g\left(\frac{N_1-1}{N_1}, \frac{k_2}{N_2}, \frac{N_1-1}{N_1} a_1, \frac{k_2 a_2}{N_2}\right)\right| \Bigg]. \end{aligned} \tag{3.18}$$

Noting that

$$\begin{aligned} &\frac{N_1-1}{2} - c_1 \le n_1 \le \frac{N_1+1}{2} + c_1 \Leftrightarrow \\ &-\left(c_1 + \frac{1}{2}\right) \le a_1 \le c_1 + \frac{1}{2} \end{aligned} \tag{3.19}$$

it follows from (3.12) and (3.13) that

$$\left| g(x_1, x_2, x_3, x_4) - g\left(x_1 + \frac{1}{N_1}, x_2, x_3 + \frac{a_1}{N_1}, x_4\right)\right| < \epsilon \tag{3.20}$$

$$|g(x_1, x_2, x_3, x_4)| < \xi < \frac{N_1 \epsilon}{2} \tag{3.21}$$

for all N_1 with $N_1 > \max(c_1/\delta, 1/\delta, 2\xi/\epsilon)$ and all a_1, which satisfies (3.19). Then from (3.18), (3.20) and (3.21), we obtain

$$|v_{n_1,n_2}| < \frac{1}{N_1 N_2} N_2 \left(\frac{N_1-1}{2}\epsilon + \frac{N_1}{2}\epsilon\right) < \epsilon$$

which proves 1).

One can prove 2) of Proposition 4 in a similar manner. It is clear that if 1) and 1) are valid, so is 3). □

It is known in [12] that the spatial impulse response of the 2-D network in [12] with infinite size is obtained by the inverse Z-transform [11] of the transfer function, and we will discuss the relation between this and our results.

Proposition 5: The impulse response of a stable 2-D block circulant network converges to the one obtained by the inverse Z-transform as the network size grows.

Proof: If the network is spatially stable, there are $\delta_1, \delta_2 > 0$ with $-\delta_1 < 1/H(e^{j\theta_1}, e^{j\theta_2}) < -\delta_2$ for all θ_1, θ_2 and thus the impulse response of the transfer function $-1/H(z_1, z_2)$ is given by

$$\begin{aligned} v^{\infty}_{k_1,k_2} &:= \frac{-1}{(2\pi j)^2} \oint \oint \frac{z_1^{k_1-1} z_2^{k_2-1}}{H(z_1, z_2)} dz_1\, dz_2 \\ &= \frac{-1}{(2\pi j)^2} \int_{|z_1|=1} \int_{|z_2|=1} \frac{z_1^{k_1-1} z_2^{k_2-1}}{H(z_1, z_2)} dz_1\, dz_2 \\ &= \frac{-1}{(2\pi j)^2} \int_0^{2\pi} \int_0^{2\pi} \frac{e^{j\theta_1(k_1-1)} e^{j\theta_2(k_2-1)}}{H(e^{j\theta_1}, e^{j\theta_2})} \\ &\quad \cdot j e^{j\theta_1} j e^{j\theta_2}\, d\theta_1\, d\theta_2 \\ &= \frac{-1}{(2\pi)^2} \int_0^{2\pi} \int_0^{2\pi} \frac{e^{j\theta_1 k_1} e^{j\theta_2 k_2}}{H(e^{j\theta_1}, e^{j\theta_2})} d\theta_1\, d\theta_2. \end{aligned} \tag{3.22}$$

Recall that from (3.5) the impulse response at node (k_1, k_2) of a 2-D block circulant network with size $N_1 \times N_2$ is given by

$$\begin{aligned} v_{k_1,k_2} &:= \frac{-1}{N_1 N_2} \sum_{p=0}^{N_1-1} \sum_{q=0}^{N_2-1} \frac{\cos\left(2\pi\left(\frac{k_1 p}{N_1} + \frac{k_2 q}{N_2}\right)\right)}{H(e^{j2\pi p/N_1}, e^{j2\pi q/N_2})} \\ &= \frac{-1}{N_1 N_2} \sum_{p=0}^{N_1-1} \sum_{q=0}^{N_2-1} \frac{e^{j2\pi(k_1 p/N_1)} e^{j2\pi(k_2 q/N_2)}}{H(e^{j2\pi p/N_1}, e^{j2\pi q/N_2})} \\ &= \frac{-1}{(2\pi)^2} \sum_{p=0}^{N_1-1} \sum_{q=0}^{N_2-1} \frac{e^{j\theta_{1,p} k_1} e^{j\theta_{2,q} k_2}}{H(e^{j\theta_{1,p}}, e^{j\theta_{2,q}})} \\ &\quad \cdot (\theta_{1,p+1} - \theta_{1,p})(\theta_{2,q+1} - \theta_{2,q}) \end{aligned} \tag{3.23}$$

where

$$\theta_{1,p} := \frac{2\pi}{N_1} p, \quad \theta_{2,q} := \frac{2\pi}{N_2} q.$$

Moreover, note that

$$\begin{aligned} 0 = \theta_{1,0} < \theta_{1,1} < \theta_{1,2} < \cdots < \theta_{1,N_1} = 2\pi, \\ 0 = \theta_{2,0} < \theta_{2,1} < \theta_{2,2} < \cdots < \theta_{2,N_2} = 2\pi \end{aligned}$$

and

$$\begin{aligned} \theta_{1,p+1} - \theta_{1,p} \to 0, \theta_{2,q+1} - \theta_{2,q} \to 0, \text{ as } N_1, N_2 \to \infty \\ \text{for } p = 0, 1, \cdots N_1 - 1, q = 0, 1, \cdots, N_2 - 1. \end{aligned}$$

If the network is stable, $1/H(e^{j\theta_1}, e^{j\theta_2})$ is continuous for $\theta_1 \times \theta_2 \in [0, 2\pi] \times [0, 2\pi]$ so that the summation in (3.23) converges to the integral in (3.22) as $N_1, N_2 \to \infty$. □

Remark: It cannot be overemphasized that Propositions 4 and 5 say that even though a 2-D block circulant network has a special boundary condition (i.e., periodic), once the stability condition (3.10) is satisfied, it behaves quite properly as the network size tends to ∞. If $\boldsymbol{A_b}$ is a uniform block matrix instead of a block circulant matrix, it is difficult to derive an analytic stability condition.

For image processing, input is naturally not always an impulse. One can show a result similar to Proposition 4 when the input is a general image, provided that its extent is uniformly bounded as N_1 and N_2 grow (see [1] for the 1-D case).

Now let us consider the 1-D cases, i.e., $N_2 = 1, m_2 = 0$ and the network size is $N_1 \times 1$. Recall our spatial stability results

for uniform band networks [1], where the system matrix $\boldsymbol{A_b}$ is a uniform band matrix instead of a circulant matrix. Then $\boldsymbol{A_b v_b} + \boldsymbol{u_b} = 0$ can be recast as $\boldsymbol{x}_{k+1} = \boldsymbol{F x}_k + \boldsymbol{y}_k$. It is shown in [1] that if and only if $\boldsymbol{F}$ is hyperbolic, the network is spatially stable. Since the function σ_+ defined by (3.10) is the same as the one defined in [1], Proposition 3 is consistent with the results therein. Why then, the new definition? A major obstacle in extending the 1-D spatial stability results obtained in [1] to 2-D cases lies in the fact that it is difficult to derive an $\boldsymbol{F}$-matrix. Our new definition can handle the 2-D network as well as the 1-D network.

Note that Propositions 4 and 5 suggest that the spatial impulse response of a spatially stable circulant network is almost insensitive to changes in N_1, N_2, the network size. On the other hand, the following example demonstrates that the spatial impulse response of a spatially unstable network can be very sensitive to changes in N_1, N_2 as well as network parameter values and boundary conditions.

Example 3: Let us consider again 2-D networks where $m_1 = m_2 = 1$ and interconnection conductances are $g_0, g_1 := g_{0,1} = g_{0,-1} = g_{1,0} = g_{-1,0} = g_{1,1} = g_{-1,-1} = g_{1,-1} = g_{-1,1}$ (Fig. 4). The stability indicator function σ_+ is given by (3.11). Fig. 6(a) shows the spatial impulse response of a 2-D block circulant network with $1/g_0 = 200\,\mathrm{k\Omega}, 1/g_1 = 20\,\mathrm{k\Omega}$ and network size 61×61, while Fig. 6(b) shows that of a 2-D block circulant network with the same g_0 and g_1 values and network size 57×57, where $\sigma_+ < 0$ and the networks are spatially stable. Fig. 6(c) shows that of a 2-D uniform block network with the same g_0 and g_1 values and network size 61×61. Fig. 6(d) shows that of a 2-D block circulant network with $1/g_0 = 200\,\mathrm{k\Omega}, 1/g_1 = 21\,\mathrm{k\Omega}$ and network size 61×61 where $\sigma_+ < 0$ and the network is spatially stable. We see that these responses in Fig. 6 are similar.

On the other hand, Fig. 7(a) shows the response of a 2-D block circulant network with $1/g_0 = -200\,\mathrm{k\Omega}, 1/g_1 = 200\,\mathrm{k\Omega}$ and network size 61×61, while Fig. 7(b) shows that of a 2-D block circulant network with the same g_0 and g_1 values and network size 57×57, where $\sigma_+ > 0$ and the networks are spatially unstable. Fig. 7(c) shows that of a 2-D uniform block network with the same g_0 and g_1 values and network size 61×61. Fig. 7(d) shows that of a 2-D block circulant network with $1/g_0 = -200\,\mathrm{k\Omega}, 1/g_1 = 210\,\mathrm{k\Omega}$ and network size 61×61 where $\sigma_+ > 0$ and the network is spatially unstable. We see that these responses in Fig. 7 are very different.

This simulation suggests that the impulse responses of stable 2-D networks are relatively insensitive to changes in boundary conditions and circuit parameter values as well as network size while those of unstable 2-D networks can be very sensitive to them. We will explain why this happens. Recall (see (3.6)) that the spatial impulse response is given by

$$\boldsymbol{v_b} = \frac{-1}{N_1 N_2} \sum_{k_1=0}^{N_1-1} \sum_{k_2=0}^{N_2-1} \frac{1}{\lambda_{N_1,k_1,N_2,k_2}} \boldsymbol{c}_{N_1,k_1,N_2,k_2} \tag{3.24}$$

provided that $\boldsymbol{A_b}$ is invertible. Also recall that

$$f(x_1, x_2) := \sum_{p=-m_1}^{m_1} \sum_{q=-m_2}^{m_2} a_{p,q} \cos\left(2\pi(px_1 + qx_2)\right),$$

$$\lambda_{N_1,k_1,N_2,k_2} = f(k_1/N_1, k_2/N_2). \tag{3.25}$$

Since the networks are spatially unstable, there are x_{z1}, x_{z2} which satisfy $f(x_{z1}, x_{z2}) = 0$. Then there are k_{z1}, k_{z2} such that $k_{z1}/N_1 \approx x_{z1}, k_{z2}/N_2 \approx x_{z2}, f(k_{z1}/N_1, k_{z2}/N_2) \approx 0$. By changing the network size N_1, N_2 by a small amount, $f(k_{z1}/N_1, k_{z2}/N_2)$ also changes only by a small value. Since $f(k_{z1}/N_1, k_{z2}/N_2) \approx 0$, however, the change of $1/f(k_{z1}/N_1, k_{z2}/N_2) (= 1/\lambda_{N_1,k_{z1},N_2,k_{z2}})$ can be drastic in terms of sign and amplitude. Noting (3.24) and (3.25), $(1/\lambda_{N_1,k_{z1},N_2,k_{z2}})\boldsymbol{c}_{N_1,k_{z1},N_2,k_{z2}}$ becomes a dominant component in the spatial impulse response and thus the spatial impulse response can be changed drastically by varying the network size N_1, N_2. The sensitivity of the spatial impulse response to changes in network parameters and boundary conditions can be similarly explained.

IV. Temporal Dynamics

Recall the definition of the temporal stability of 1-D networks given in [1].

Definition—Temporal Stability for a 2-D Network: The 2-D network given by (2.3) is temporally stable if $\boldsymbol{A_b}$ is negative definite for all N_1, N_2.

Proposition 6: Recall (2.4), (2.5), (2.6) and define

$$\mu_{N_1,k_1,N_2,k_2} := \sum_{p=-m_1}^{m_1} \sum_{q=-m_2}^{m_2} b_{p,q} \cdot \cos\left(2\pi\left(\frac{k_1 p}{N_1} + \frac{k_2 q}{N_2}\right)\right),$$
$$\text{where } k_1 = 0, 1, \cdots, N_1 - 1,$$
$$k_2 = 0, 1, \cdots, N_2 - 1.$$

Also recall

$$\lambda_{N_1,k_1,N_2,k_2} := \sum_{p=-m_1}^{m_1} \sum_{q=-m_2}^{m_2} a_{p,q} \cdot \cos\left(2\pi\left(\frac{k_1 p}{N_1} + \frac{k_2 q}{N_2}\right)\right),$$
$$\text{where } k_1 = 0, 1, \cdots, N_1 - 1,$$
$$k_2 = 0, 1, \cdots, N_2 - 1.$$

i) Consider the temporal dynamics (2.3) with the initial condition $\boldsymbol{v_{b0}} := \boldsymbol{v_b}(0)$. Then $\boldsymbol{v_b}(t)$ is explicitly given by

$$\begin{aligned}\boldsymbol{v_b}(t) &= e^{\boldsymbol{B_b^{-1}} \boldsymbol{A_b} t} \boldsymbol{v_{b0}} + \int_0^t e^{\boldsymbol{B_b^{-1}} \boldsymbol{A_b}(t-\tau)} \boldsymbol{u_b}(\tau)\, d\tau \\ &= (\boldsymbol{F}_{N_1} \otimes \boldsymbol{F}_{N_2})^* \operatorname{diag}\left(e^{-t/T_{N_1,0,N_2,0}}, e^{-t/T_{N_1,0,N_2,1}}, \right. \\ &\qquad \left. e^{-t/T_{N_1,0,N_2,2}}, \cdots, e^{-t/T_{N_1,N_1-1,N_2,N_2-1}}\right) \\ &\quad \cdot (\boldsymbol{F}_{N_1} \otimes \boldsymbol{F}_{N_2}) \boldsymbol{v_{b0}} \\ &\quad + \int_0^t (\boldsymbol{F}_{N_1} \otimes \boldsymbol{F}_{N_2})^* \operatorname{diag}\left(e^{-(t-\tau)/T_{N_1,0,N_2,0}}, \right. \\ &\qquad e^{-(t-\tau)/T_{N_1,0,N_2,1}}, e^{-(t-\tau)/T_{N_1,0,N_2,2}}, \cdots, \\ &\qquad \left. e^{-(t-\tau)/T_{N_1,N_1-1,N_2,N_2-1}}\right)(\boldsymbol{F}_{N_1} \otimes \boldsymbol{F}_{N_2}) \boldsymbol{u_b}(\tau)\, d\tau\end{aligned} \tag{4.1}$$

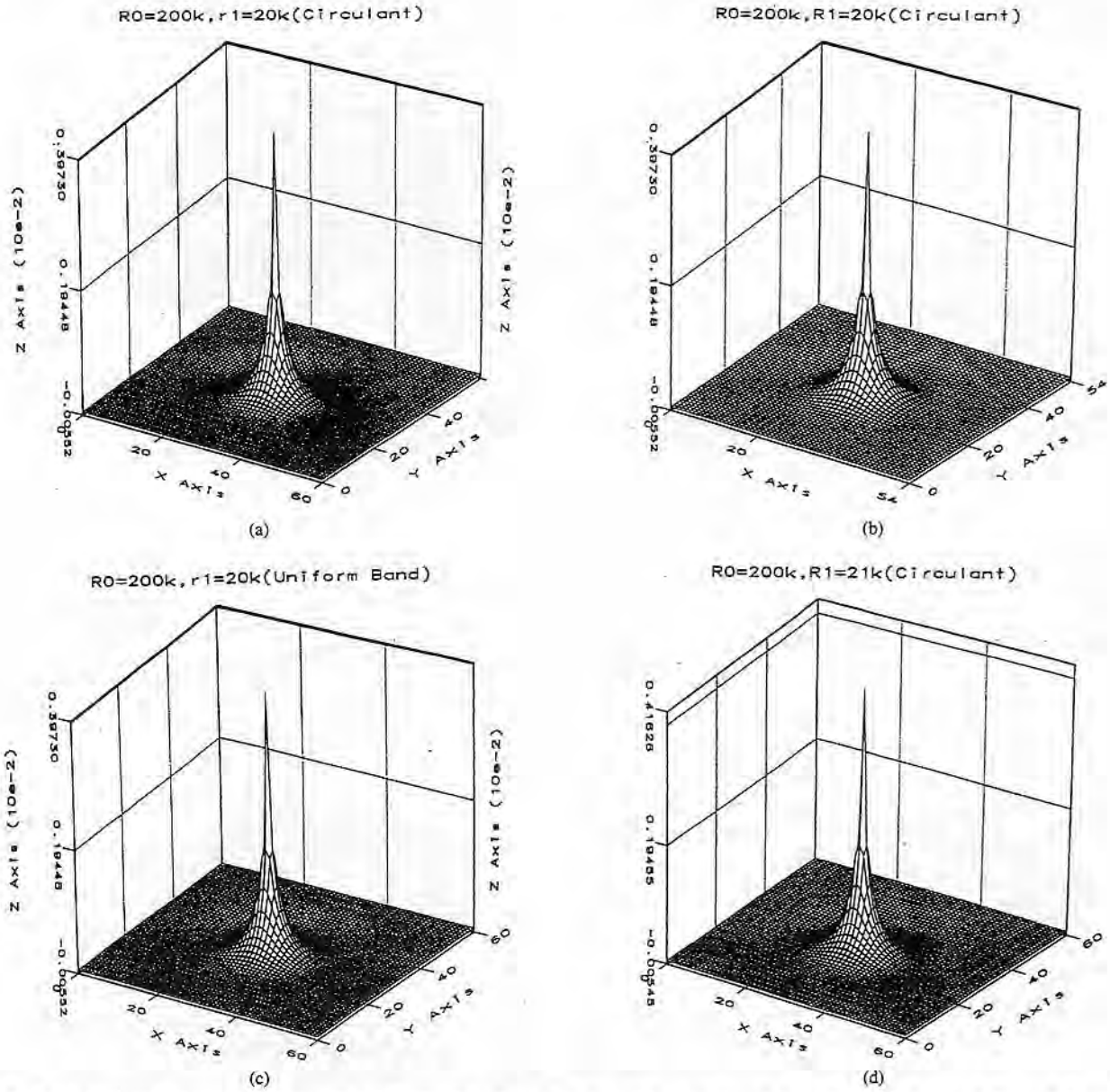


Fig. 6. Spatial impulse responses of 2-D spatially stable networks with $g_0, g_1 := g_{0,1} = g_{0,-1} = g_{1,0} = g_{-1,0} = g_{1,1} = g_{-1,-1} = g_{1,-1} = g_{-1,1}$ which corresponds to the network in Fig. 4. 1 μA current is applied to the center node. (a) Response of a block circulant network with $1/g_0 = 200\,\mathrm{k}\Omega, 1/g_1 = 20\,\mathrm{k}\Omega$ and network size 61×61. (b) Response of a block circulant network with $1/g_0 = 200\,\mathrm{k}\Omega, 1/g_1 = 20\,\mathrm{k}\Omega$ and network size 55×55. (c) Response of a uniform block network with $1/g_0 = 200\,\mathrm{k}\Omega, 1/g_1 = 20\,\mathrm{k}\Omega$ and network size 61×61. (d) Response of a block circulant network with $1/g_0 = 200\,\mathrm{k}\Omega, 1/g_1 = 21\,\mathrm{k}\Omega$ and network size 61×61.

where

$$T_{N_1,k_1,N_2,k_2} := -\frac{\mu_{N_1,k_1,N_2,k_2}}{\lambda_{N_1,k_1,N_2,k_2}}. \tag{4.2}$$

$\boldsymbol{F}_{N_1}$ and $\boldsymbol{F}_{N_2}$ are Fourier matrixes with sizes $N_1 \times N_1$ and $N_2 \times N_2$, respectively [9]. $\otimes$ denotes a Kronecker product. diag means a diagonal matrix. $(\boldsymbol{F}_{N_1} \otimes \boldsymbol{F}_{N_2})^*$ represents the conjugate transpose of $(\boldsymbol{F}_{N_1} \otimes \boldsymbol{F}_{N_2})$. And a Fourier matrix $\boldsymbol{F}_N$ with size $N \times N$ is defined by

$$\boldsymbol{F}_N := \{F(i,j)\} \in \mathcal{C}^{N\times N}, \qquad i,j = 0,1,\cdots,N_1 - 1$$

where

$$F(i,j) := \frac{1}{\sqrt{N}} W^{-ij}, \quad W := e^{j(2\pi/N)}$$

and $\boldsymbol{F}_{N_1} \otimes \boldsymbol{F}_{N_2}$ is given by

$$\boldsymbol{F}_{N_1} \otimes \boldsymbol{F}_{N_2} := \{\boldsymbol{F}(i,j)\} \in \mathcal{C}^{N_1N_2\times N_1N_2}, \quad i,j = 0,1,\cdots,N_1 - 1$$

where

$$\boldsymbol{F}(i,j) := \frac{1}{\sqrt{N_1}} W_1^{-ij} \boldsymbol{F}_{N_2} \in \mathcal{C}^{N_2\times N_2}, \quad W_1 := e^{j(2\pi/N_1)},$$
$$W_2 := e^{j(2\pi/N_2)}.$$

ii) The temporal dynamics is stable if and only if

$$\sum_{p=-m_1}^{m_1} \sum_{q=-m_2}^{m_2} a_{p,q} \cos(2\pi(py_1 + qy_2)) < 0$$
$$\text{for all rational } y_1, y_2, \quad 0 \le y_1, y_2 < 1. \tag{4.3}$$

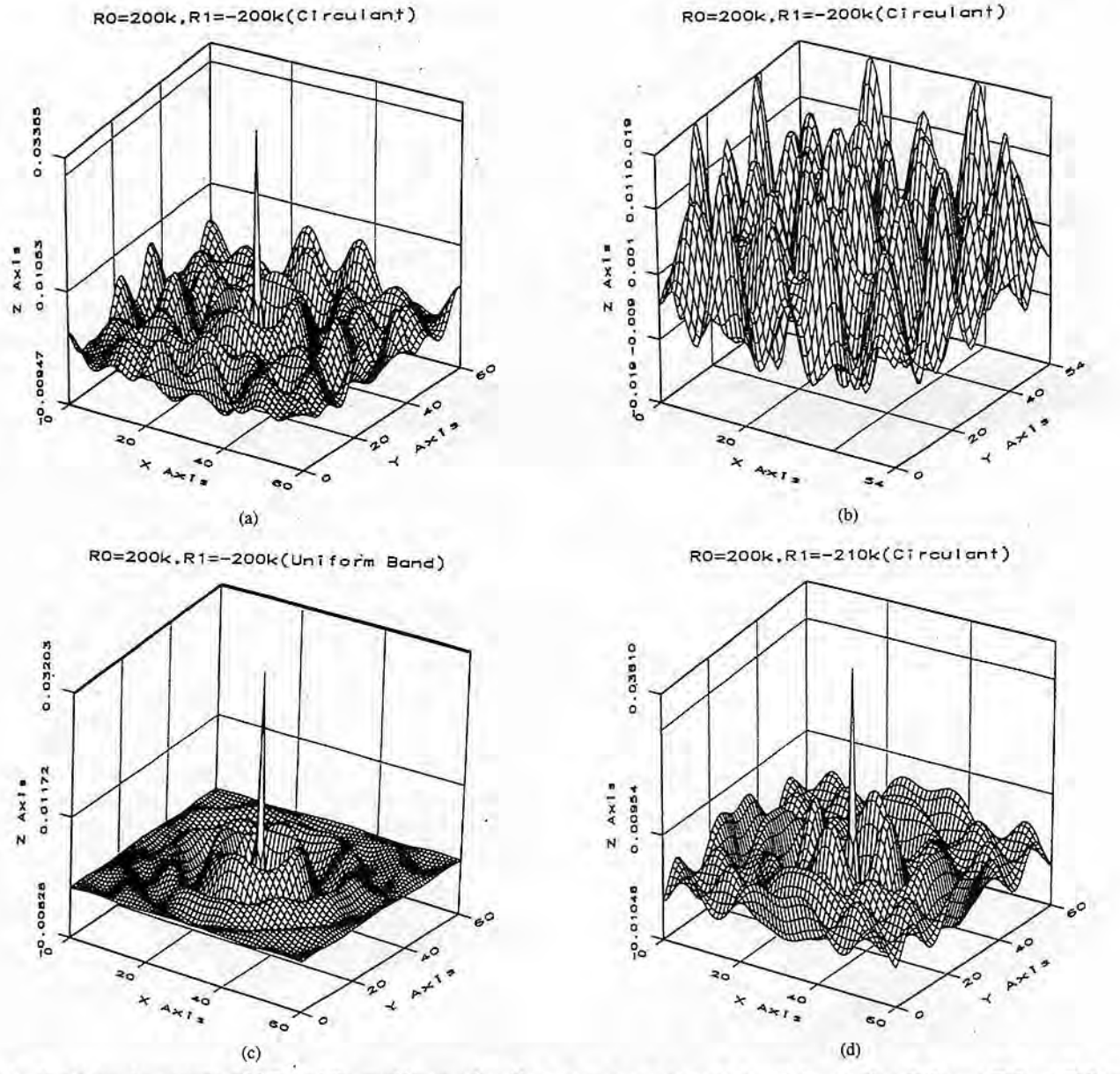


Fig. 7. Spatial impulse responses of 2-D spatially unstable networks with $g_0, g_1 := g_{0,1} = g_{0,-1} = g_{1,0} = g_{-1,0} = g_{1,1} = g_{-1,-1} = g_{1,-1} = g_{-1,1}$ which corresponds to the network in Fig. 4. (a) Response of a block circulant network with $1/g_0 = -200\,\mathrm{k}\Omega, 1/g_1 = 200\,\mathrm{k}\Omega$ and network size 61×61. (b) Response of a block circulant network with $1/g_0 = -200\,\mathrm{k}\Omega, 1/g_1 = 200\,\mathrm{k}\Omega$ and network size 55×55. (c) Response of a uniform block network with $1/g_0 = -200\,\mathrm{k}\Omega, 1/g_1 = 200\,\mathrm{k}\Omega$ and network size 61×61. (d) Response of a block circulant network with $1/g_0 = -200\,\mathrm{k}\Omega, 1/g_1 = 210\,\mathrm{k}\Omega$ and network size 61×61.

Proof: i) The solution of state equation (2.3) with the initial condition $\boldsymbol{v}_{\boldsymbol{b}0} := \boldsymbol{v}_{\boldsymbol{b}}(0)$ is given by

$$\boldsymbol{v}_{\boldsymbol{b}}(t) = e^{\boldsymbol{B}_{\boldsymbol{b}}^{-1}\boldsymbol{A}_{\boldsymbol{b}}t}\boldsymbol{v}_{\boldsymbol{b}0} + \int_0^t e^{\boldsymbol{B}_{\boldsymbol{b}}^{-1}\boldsymbol{A}_{\boldsymbol{b}}(t-\tau)}\boldsymbol{u}_{\boldsymbol{b}}(\tau)\,d\tau. \quad (4.4)$$

From Lemma 2-i), iii), and vi) (see Appendix), it follows that

$$\begin{aligned}\boldsymbol{B}_{\boldsymbol{b}}^{-1}\boldsymbol{A}_{\boldsymbol{b}}t &= (\boldsymbol{F}_{N_1}\otimes\boldsymbol{F}_{N_2})^*\,\mathrm{diag}\,(\lambda_{N_1,0,N_2,0}, \lambda_{N_1,0,N_2,1}, \cdots,\\ &\quad \lambda_{N_1,0,N_2,N_2-1}; \lambda_{N_1,1,N_2,0}, \lambda_{N_1,1,N_2,1}, \cdots,\\ &\quad \lambda_{N_1,1,N_2,N_2-1}; \cdots, \lambda_{N_1,N_1-1,N_2,0},\\ &\quad \lambda_{N_1,N_1-1,N_2,1}, \cdots, \lambda_{N_1,N_1-1,N_2,N_2-1})\\ &\quad \cdot (\boldsymbol{F}_{N_1}\otimes\boldsymbol{F}_{N_2}) \times (\boldsymbol{F}_{N_1}\otimes\boldsymbol{F}_{N_2})^*\\ &\quad \cdot \mathrm{diag}\,(\mu_{N_1,0,N_2,0}, \mu_{N_1,0,N_2,1}, \cdots,\\ &\quad \mu_{N_1,0,N_2,N_2-1}; \mu_{N_1,1,N_2,0}, \mu_{N_1,1,N_2,1}, \cdots,\\ &\quad \mu_{N_1,1,N_2,N_2-1}; \cdots \mu_{N_1,N_1-1,N_2,0},\\ &\quad \mu_{N_1,N_1-1,N_2,1}, \cdots, \mu_{N_1,N_1-1,N_2,N_2-1})\\ &\quad \cdot (\boldsymbol{F}_{N_1}\otimes\boldsymbol{F}_{N_2})t\\ &= (\boldsymbol{F}_{N_1}\otimes\boldsymbol{F}_{N_2})^*\,\mathrm{diag}\,(-t/T_{N_1,0,N_1,0},\\ &\quad -t/T_{N_1,0,N_1,1}, -t/T_{N_1,0,N_2,2}, \cdots,\\ &\quad -t/T_{N_1,N_1-1,N_2,N_2-1})(\boldsymbol{F}_{N_1}\otimes\boldsymbol{F}_{N_2}).\end{aligned}$$

Using Lemma 2-ii), one obtains

$$e^{B_{\boldsymbol{b}}^{-1}A_{\boldsymbol{b}}t}v_{\boldsymbol{b}0} = (F_{N_1}\otimes F_{N_2})^* \operatorname{diag}(e^{-t/T_{N_1,0,N_2,0}}, e^{-t/T_{N_1,0,N_2,1}}, e^{-t/T_{N_1,0,N_2,2}}, \cdots, e^{-t/T_{N_1,N_1-1,N_2,N_2-1}})(F_{N_1}\otimes F_{N_2}) \cdot v_{\boldsymbol{b}0}. \tag{4.5}$$

Similarly, one has

$$\int_0^t e^{B_{\boldsymbol{b}}^{-1}A_{\boldsymbol{b}}(t-\tau)}u_{\boldsymbol{b}}(\tau)\,d\tau = \int_0^t (F_{N_1}\otimes F_{N_2})^* \times \operatorname{diag}(e^{-(t-\tau)/T_{N_1,0,N_2,0}}, e^{-(t-\tau)/T_{N_1,0,N_2,1}}, e^{-(t-\tau)/T_{N_1,0,N_2,2}}, \cdots, e^{-(t-\tau)/T_{N_1,N_1-1,N_2,N_2-1}})(F_{N_1}\otimes F_{N_2}) \cdot u_{\boldsymbol{b}}(\tau)\,d\tau. \tag{4.6}$$

Equation (4.1) follows from (4.4), (4.5), and (4.6).

ii) Recall the temporal stability definition of the 2-D network. The 2-D network is temporally stable if all the eigenvalues of $A_{\boldsymbol{b}}$ are negative, i.e., for all N_1, k_1, N_2, and k_2

$$\lambda_{N_1,k_1,N_2,k_2} = \sum_{p=-m_1}^{m_1}\sum_{q=-m_2}^{m_2} a_{p,q} \cdot \cos\left(2\pi\left(\frac{k_1 p}{N_1}+\frac{k_2 q}{N_2}\right)\right) < 0$$
$$\text{where } k_1 = 0,1,\cdots,N_1-1,\quad k_2 = 0,1,\cdots,N_2-1.$$

□

Statement i) shows that T_{N_1,k_1,N_2,k_2}'s defined by (4.2) can be interpreted as "time constants" of the RC network even though "time constant" in its strictest sense is defined only for a first order network. Recall the standing assumption i) (positive definiteness of $B_{\boldsymbol{b}}$) and note that μ_{N_1,k_1,N_2,k_2} is positive. Then we see that the temporal stability condition $\lambda_{N_1,k_1,N_2,k_2} < 0$ implies $T_{N_1,k_1,N_2,k_2} > 0$ for all k_1, k_2.

Remark: The assumption that "the parasitic capacitances of the resistors are the same for the similar resistors and thus $B_{\boldsymbol{b}}$ is a symmetric block circulant matrix" may be a rather crude approximation in actual LSI implementation. This is because parasitic capacitances of MOSFET's (metal-oxide semiconductor/field effect transistors) depend on their bias voltages and they can be mismatched during fabrication. We would like to point out, however, that this assumption led to the explicit time constant formula which can help a filter designer estimate the temporal behavior of the network.

Example 4: Consider a block circulant network where $m_1 = 2$, $m_2 = 0$, $1/g_1 := 1/g_{1,0} = 1/g_{-1,0} = 1/\cos(\sqrt{2})$ k Ω, $1/g_2 := 1/g_{2,0} = 1/g_{-2,0} = -4k\,\Omega$, $g_0 = -(2g_1 + 4g_2 + g_1^2/(4g_2))$, and the other $g_{i,j}$'s $= 0$. This network is spatially unstable but temporally stable. The stability indicator function σ_+ is given by

$$\begin{aligned}\sigma_+ &:= -(g_0 + 2g_1 + 2g_2) + 2\max_{x\in[0,1]}\cdot\{g_1\cos(2\pi x) + g_2\cos(4\pi x)\}\\ &= -(g_0 + 2g_1 + 2g_2) + 2\max_{x\in[0,1]}\cdot\left\{2g_2\left(\cos(2\pi x)+\frac{g_1}{4g_2}\right)^2 - \frac{g_1^2}{8g_2} - g_2\right\}\\ &= -\left(g_0 + 2g_1 + 4g_2 + \frac{g_1^2}{4g_2}\right) + 4\max_{x\in[0,1]}\cdot\left\{g_2\left(\cos(2\pi x)+\frac{g_1}{4g_2}\right)^2\right\}\\ &= 4\max_{x\in[0,1]}\left\{g_2\left(\cos(2\pi x)+\frac{g_1}{4g_2}\right)^2\right\} = 0\\ &\quad \text{at } x = \frac{1}{2\pi}\arccos\left(-\frac{g_1}{4g_2}\right) = \frac{\sqrt{2}}{2\pi}.\end{aligned}$$

Since $\sigma_+ = 0$, it is spatially unstable. Next let us check the temporal stability. The left-hand term of (4.3) is given by

$$\sum_{p=-2}^{2} a_{p,0}\cos(2\pi p y) = 4g_2\left(\cos(2\pi y)+\frac{g_1}{4g_2}\right)^2 \le 0 \tag{4.7}$$

which is zero if and only if $y = \sqrt{2}/2\pi$. Since y is restricted to a rational number, the left-hand side of (4.7) is always negative. Thus the network is temporally stable.

We will next study several more relationships between spatial and temporal dynamics. Let $v_{\boldsymbol{b}}(0) = 0$ and consider

$$u_{\boldsymbol{b}}(t) = (\alpha c_{N_1,k_1,N_2,k_2} + \beta s_{N_1,k_1,N_2,k_2})\cdot u(t),$$
$$\text{where } u(t) = \begin{cases}1 & \text{when } t \ge 0\\ 0 & \text{when } t < 0\end{cases}$$

α and β are constants, and $c_{N_1,k_1,N_2,k_2}, s_{N_1,k_1,N_2,k_2}$ are defined by (7.2) and (7.3), respectively. Then it follows from Proposition 6 that $v_{\boldsymbol{b}}(t), t \ge 0$ is given by

$$v_{\boldsymbol{b}}(t) = \frac{-1}{\lambda_{N_1,k_1,N_2,k_2}}(1 - e^{-t/T_{N_1,k_1,N_2,k_2}})u_{\boldsymbol{b}}(t)$$

which has a rather interesting interpretation: if there is a spatial angle frequency $(2\pi k_1/N_1, 2\pi k_2/N_2)$ for which the phase of the spatial frequency response is π, the network is temporally unstable because $\lambda_{N_1,k_1,N_2,k_2} \ge 0$; otherwise (i.e., when phase of each spatial frequency response is zero), it is temporally stable.

It is also interesting to see that the "gain" $|1/\lambda_{N_1,k_1,N_2,k_2}|$ of the spatial frequency response is proportional to the time constant $T_{N_1,k_1,N_2,k_2} = -\mu_{N_1,k_1,N_2,k_2}/\lambda_{N_1,k_1,N_2,k_2}$ of the temporal dynamics. Since μ_{N_1,k_1,N_2,k_2} and $\lambda_{N_1,k_1,N_2,k_2}$ are always real, we see that if a block circulant network is spatially stable (respectively, unstable), then the "phase" of the spatial frequency response is zero (respectively, π). We remark that a stable linear phase causal IIR (infinite impulse response) filter cannot be realized [11] but the block circulant network can be a stable zero phase noncausal IIR filter. The above fact

clarifies the relation between the phase of frequency response and the stability of such noncausal IIR filters.

If the input is given by $\boldsymbol{u_b} = \alpha \boldsymbol{c}_{N_1,k_1,N_2,k_2} + \beta \boldsymbol{s}_{N_1,k_1,N_2,k_2}$ [see (3.3)], the power dissipated by the network at an equilibrium is given by

$$\begin{aligned}\text{Power} &= \boldsymbol{u_b}^T \boldsymbol{v_b} = -(\alpha \boldsymbol{c}_{N_1,k_1,N_2,k_2} + \beta \boldsymbol{s}_{N_1,k_1,N_2,k_2})^T \boldsymbol{A_b}^{-1} \\ &\quad \cdot (\alpha \boldsymbol{c}_{N_1,k_1,N_2,k_2} + \beta \boldsymbol{s}_{N_1,k_1,N_2,k_2}) \\ &= \frac{-1}{\lambda_{N_1,k_1,N_2,k_2}} (\alpha \boldsymbol{c}_{N_1,k_1,N_2,k_2} + \beta \boldsymbol{s}_{N_1,k_1,N_2,k_2})^T \\ &\quad \cdot (\alpha \boldsymbol{c}_{N_1,k_1,N_2,k_2} + \beta \boldsymbol{s}_{N_1,k_1,N_2,k_2})\end{aligned}$$

where α and β are constants. Thus, when a network is spatially stable, $-\lambda^{-1}_{N_1,k_1,N_2,k_2}$ is positive and the power consumed by the network is positive, i.e., the network acts as a passive element at that spatial frequency. On the other hand, when the network is unstable, $-\lambda^{-1}_{N_1,k_1,N_2,k_2}$ is negative and the power consumed by the network is negative, i.e., the network acts as an active element at that spatial frequency.

Next consider the case that

$$\begin{aligned}\boldsymbol{u_b}(t) &= g(t)(\alpha \boldsymbol{c}_{N_1,k_1,N_2,k_2} + \beta \boldsymbol{s}_{N_1,k_1,N_2,k_2}), \\ \boldsymbol{v_b}(0) &= \gamma(\alpha \boldsymbol{c}_{N_1,k_1,N_2,k_2} + \beta \boldsymbol{s}_{N_1,k_1,N_2,k_2})\end{aligned}$$

where α, β, and γ are constants, and $g(t)$ is a real-valued function, such that $\int_0^t e^{-[(t-\tau)]/T_{N_1,k_1,N_2,k_2}} g(\tau)\, d\tau$ is welldefined for $t \geq 0$. Then $\boldsymbol{v_b}(t)$ is given by

$$\begin{aligned}\boldsymbol{v_b}(t) &= e^{\boldsymbol{B_b}^{-1}\boldsymbol{A_b}t} \boldsymbol{v_b}(0) + \int_0^t e^{\boldsymbol{B_b}^{-1}\boldsymbol{A_b}(t-\tau)} \boldsymbol{B_b}^{-1} \boldsymbol{u_b}(\tau)\, d\tau \\ &= \left(\gamma e^{-t/T_{N_1,k_1,N_2,k_2}} + \frac{1}{\mu_{N_1,k_1,N_2,k_2}} \right. \\ &\quad \left. \int_0^t \cdot e^{-[(t-\tau)]/T_{N_1,k_1,N_2,k_2}} g(\tau)\, d\tau \right) \\ &\quad \cdot (\alpha \boldsymbol{c}_{N_1,k_1,N_2,k_2} + \beta \boldsymbol{s}_{N_1,k_1,N_2,k_2}) \\ &= \xi(t)(\alpha \boldsymbol{c}_{N_1,k_1,N_2,k_2} + \beta \boldsymbol{s}_{N_1,k_1,N_2,k_2})\end{aligned}$$

where

$$\begin{aligned}\xi(t) := \gamma e^{-t/T_{N_1,k_1,N_2,k_2}} &+ \frac{1}{\mu_{N_1,k_1,N_2,k_2}} \\ &\int_0^t \cdot e^{-[(t-\tau)]/T_{N_1,k_1,N_2,k_2}} g(\tau)\, d\tau.\end{aligned}$$

Thus

$$\begin{aligned}\boldsymbol{B_b} \frac{d}{dt} \xi(t)(\alpha \boldsymbol{c}_{N_1,k_1,N_2,k_2} + \beta \boldsymbol{s}_{N_1,k_1,N_2,k_2}) \\ = \boldsymbol{A_b} \xi(t)(\alpha \boldsymbol{c}_{N_1,k_1,N_2,k_2} + \beta \boldsymbol{s}_{N_1,k_1,N_2,k_2}) \\ + g(t)(\alpha \boldsymbol{c}_{N_1,k_1,N_2,k_2} + \beta \boldsymbol{s}_{N_1,k_1,N_2,k_2}),\end{aligned}$$

$$\mu_{N_1,k_1,N_2,k_2} \frac{d}{dt} \xi(t) = \lambda_{N_1,k_1,N_2,k_2} \xi(t) + g(t). \tag{4.8}$$

Since (4.8) is the KCL equation in Fig. 8, a 2-D block circulant network, for example, in Fig. 4 is equivalent to the network shown in Fig. 8, when input $\boldsymbol{u_b}(t)$ and initial value $\boldsymbol{v_b}(0)$ are spatially sinusoidal. Since all the nodes are disconnected from each other in Fig. 8, one sees that $\boldsymbol{v_b} = -\lambda^{-1}_{N_1,k_1,N_2,k_2} \boldsymbol{u_b}$ at the equilibrium point and the time constant is $-\mu_{N_1,k_1,N_2,k_2}/\lambda_{N_1,k_1,N_2,k_2}$. Moreover, we see that the circuit theoretical interpretations of $\lambda_{N_1,k_1,N_2,k_2}$ and μ_{N_1,k_1,N_2,k_2} are that $-\lambda_{N_1,k_1,N_2,k_2}$ behaves as an effective conductance and that μ_{N_1,k_1,N_2,k_2} behaves as an effective capacitance of the network for spatially sinusoidal $\boldsymbol{u_b}(t)$ and $\boldsymbol{v_b}(0)$.

Example 5: The power of both the networks in Fig. 3 and Fig. 8 is given by

$$\begin{aligned}\text{Power} &= \boldsymbol{u_b}^T \boldsymbol{v_b} \\ &= \left(\mu_{N_1,k_1,N_2,k_2} \frac{d}{dt} \xi(t) - \lambda_{N_1,k_1,N_2,k_2} \xi(t) \right) \\ &\quad \cdot (\alpha \boldsymbol{c}_{N_1,k_1,N_2,k_2} + \beta \boldsymbol{s}_{N_1,k_1,N_2,k_2})^T \xi(t) \\ &\quad \cdot (\alpha \boldsymbol{c}_{N_1,k_1,N_2,k_2} + \beta \boldsymbol{s}_{N_1,k_1,N_2,k_2}) \\ &= \mu_{N_1,k_1,N_2,k_2} \frac{d}{dt} \boldsymbol{v_b}^T \boldsymbol{v_b} - \lambda_{N_1,k_1,N_2,k_2} \boldsymbol{v_b}^T \boldsymbol{v_b}.\end{aligned}$$

The first term of the last expression indicates the power stored in the capacitor in the form of $\frac{1}{2} \mu_{N_1,k_1,N_2,k_2} \boldsymbol{v_b}^T \boldsymbol{v_b}$ electrostatic potential energy, while the second term indicates the power dissipated by the ohmic conductance of the network.

V. Concluding Remarks

1) In this paper we have discussed the spatio-temporal stability and dynamics of 1-D and 2-D circulant networks. We conjecture that our results can be extended for m-dimensional ($m \geq 3$) cases which may have significance in higher dimensional image-processing problems.
2) We conjecture that, in general, 1-D and 2-D spatially stable networks are relatively insensitive to changes with respect to circuit parameters $\{a_{0,0}, a_{0,1}, \cdots, a_{m1,m2}\}$, network size and boundary conditions in spatial dynamics, whereas 1-D and 2-D spatially unstable networks are very sensitive to these. We showed in [1] that the spatial impulse responses of 1-D stable networks are almost insensitive to changes with respect to network size and boundary conditions. Propositions 4 and 5 suggest that those of 2-D stable networks are robust to network size. On the other hand, Section III shows that the responses of 1-D and 2-D unstable circulant networks are very sensitive to changes with respect to network size. Also, simulation results suggest that the responses of unstable networks may vary drastically to changes with respect to circuit parameters and boundary conditions (see Example 3).
3) Hexagonal sampling is known to be the most efficient sampling method in 2-D systems [12] and it is extensively used in image-processing neuro chips [2], [7]. Our results are directly applicable to hexagonal networks as well as square networks. For example, let us consider the image-smoothing neuro chip [2] which motivated

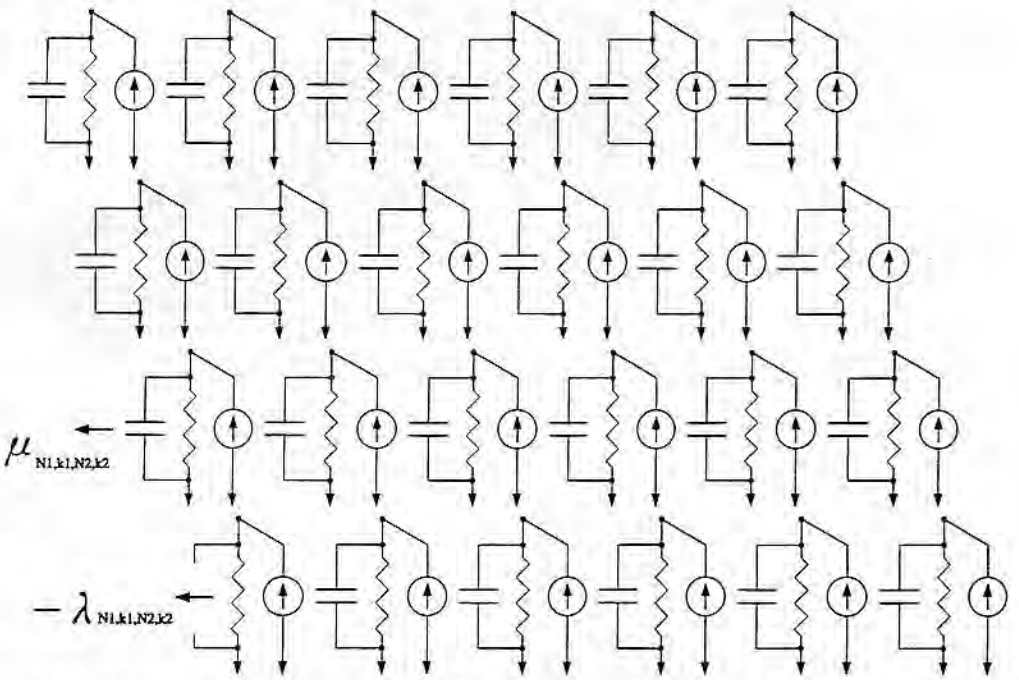


Fig. 8. An equivalent circuit for spatially sinusoidal input $\boldsymbol{u}_{\boldsymbol{b}}(t) = g(t)(\alpha \boldsymbol{c}_{N_1,k_1,N_2,k_2} + \beta \boldsymbol{s}_{N_1,k_1,N_2,k_2})$ and initial value $\boldsymbol{v}_{\boldsymbol{b}}(0) = \gamma\ (\alpha \boldsymbol{c}_{N_1,k_1,N_2,k_2} + \beta \boldsymbol{s}_{N_1,k_1,N_2,k_2})$. Each node has a conductance $-\lambda_{N_1,k_1,N_2,k_2}$ to ground and a capacitance μ_{N_1,k_1,N_2,k_2} to ground.

the present study. The structure of the network is shown in Fig. 1. There are two labeling conventions for a hexagonal grid: the standard one shown in Fig. 9(a) and the alternate one shown in Fig. 9(b). With the standard labeling convention, we obtain $g_0 > 0, g_1 := g_{1,0} = g_{0,1} = g_{1,-1} > 0, g_2 := g_{2,0} = g_{0,2} = g_{2,-2} < 0$ and $g_1 = 4|g_2|$. Thus the stability indicator function σ_+ is given by

$$\begin{aligned}\sigma_+ = \max_{x_1,x_2\in[0,1]\times[0,1]} \{&-g_0 - 6g_1 - 6g_2 + 2g_1(\cos(2\pi x_1) \\ &+ \cos(2\pi x_2) + 2g_1\cos(2\pi(x_1+x_2))) \\ &+ 2g_2(\cos(4\pi x_1) + \cos(4\pi x_2) \\ &+ 2\cos(4\pi(x_1+x_2)))\}.\end{aligned}$$

Noting that $g_2 = -g_1/4$, we have

$$\begin{aligned}\sigma_+ &= \max_{x_1,x_2\in[0,1]\times[0,1]} \{-g_0 - 6g_1 + \tfrac{6}{4}g_1 + 2g_1(\cos(2\pi x_1) \\ &\quad + \cos(2\pi x_2) + \cos(2\pi(x_1+x_2))) \\ &\quad - \tfrac{2}{4}g_1(\cos(4\pi x_1) \\ &\quad + \cos(4\pi x_2) + \cos(4\pi(x_1+x_2)))\} \\ &= \max_{x_1,x_2\in[0,1]\times[0,1]} \{-g_0 - g_1\{(1-\cos(2\pi x_1))^2 \\ &\quad + (1-\cos(2\pi x_2))^2 + (1-\cos(2\pi(x_1+x_2)))^2\}\} \\ &= -g_0 < 0 \quad \text{at } x_1 = x_2 = 0.\end{aligned}$$

Since the stability indicator function σ_+ is negative, the spatial and temporal stabilities are guaranteed.

4) This paper assumes that network conductance and capacitance are linear and a capacitor is located parallel to a conductance. In actual implementation, however, MOS transistors would be used for negative resistors [2]. Therefore there can be associated nonlinear parasitic capacitances at different locations. Moreover, conductances can be nonlinear. The differences between an ideal network and its actual circuit realization can easily be enough to push a stable network into instability, particularly if it is a fairly high-order system. The stability analysis of a network implemented with actual circuitry is important and left for future work.

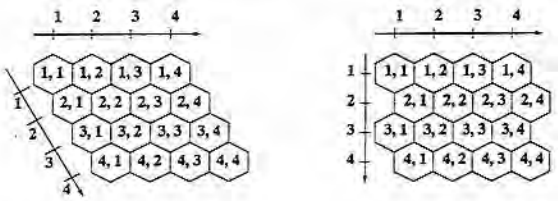


Fig. 9. Hexagonal grid labeling conventions. (a) Standard. (b) Alternate.

APPENDIX

This appendix gives some properties of circulant and block circulant matrixes [9].

Lemma 1—Eigenvalues and Eigenvectors of Symmetric Block Circulant Matrixes with Circulant Blocks: i) Given a network size $N_1 \times N_2$, eigenvalues $\lambda_{N_1,k_1,N_2,k_2}$ for a symmetric block circulant matrix $\boldsymbol{A}_{\boldsymbol{b}}$ with circulant blocks are given explicitly by

$$\lambda_{N_1,k_1,N_2,k_2} = \sum_{p=-m_1}^{m_1}\sum_{q=-m_2}^{m_2} a_{p,q}\cos\left(2\pi\left(\frac{k_1 p}{N_1} + \frac{k_2 q}{N_2}\right)\right) \tag{7.1}$$

[11] A. V. Oppenheim, and R. W. Schafer, *Digital Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1975.
[12] D. E. Dudgeon and R. M. Mersereau, *Multidimensional Digital Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1984.

Haruo Kobayashi (S'88–M'90) was born in Utsunomiya, Japan, in 1958 and received the B.S. and M.S. degrees in information physics and mathematical engineering from the University of Tokyo in 1980 and 1982, respectively, and the Dr. Eng., degree in electrical engineering from Waseda University in 1995.. From 1987 to 1989, he was at UCLA and received the M.S. degree in electrical engineering in 1989.

He joined Yokogawa Electric Corp. Tokyo, Japan, in 1982, where he has been engaged in the research and development related to measuring instruments and a mini-supercomputer. From 1994 he has been transferred to Teratec Corp. on temporary leave from Yokogawa Electric Corp., to develop ultra-high-speed ADCs with HBT; he is also a lecturer at Waseda University. His recent research interests include analog CMOS IC design and neural networks.

Dr. Kobayashi is a member of the Institute of Electronics, Information and Communication Engineers of Japan and the Society of Instrument and Control Engineers of Japan. He is a recipient of the 1994 Best Paper Award from the Japanese Neural Network Society.

Takashi Matsumoto (M'71-SM'83-F'85) received the B.Eng. degree in electrical engineering from Waseda University, Tokyo, Japan, the M.S. degree in applied mathematics from Harvard University, Cambridge, MA, and the Dr.Eng. degree in electrical engineering from Waseda University, Tokyo, Japan.

He is presently Professor and Chairperson of the Department of Electrical Engineering, Waseda University. His research interests include bifurcations/chaos and neural networks.

Dr. Matsumoto is currently the Chairperson of the IEEE CAS Society Tokyo Chapter. He chairs the Special Committee on "Chaos/Mathematics and New Technology" of the Institute of Electrical Engineers of Japan, where he organizes various workshops on bifurcations, chaos, fractals and their applications. He serves on the editorial board of *Circuits, Systems and Signal Processing*. He is a recipient of the 1994 Best Paper Award from the Japanese Neural Network Society.

Jun Sanekata was born in Tokyo, Japan, on November 30, 1967. He received the B.E. and M.E. degrees, respectively, in 1991 and 1993 in electrical engineering from Waseda University, Tokyo, Japan.

In April 1993, he joined Hitachi, Ltd., Tokyo, Japan.

Mr. Sanekata is a member of IEE, Japan.

No.8

2015年11月27日　STARCフォーラム2015学生ポスターセッション　於　新横浜

黄金比重み付けDA変換器の構成

澁谷　将平、小林　佑太朗、荒船　拓也、　小林　春夫
群馬大学大学院理工学府　電子情報部門
〒376-8515　群馬県桐生市天神町1-5-1　E-mail:t15804045@gunma-u.ac.jp

研究目的

研究背景・目的

自動車のエレクトロニクス化が著しく
車載用エレクトロニクス技術に大きな関心

車載用マイコンと組み合わせるADCへの要求が厳しい
➡ 逐次比較近似AD変換器

+冗長性

逐次比較近似AD変換器の冗長設計
➡ 高性能化・高速化

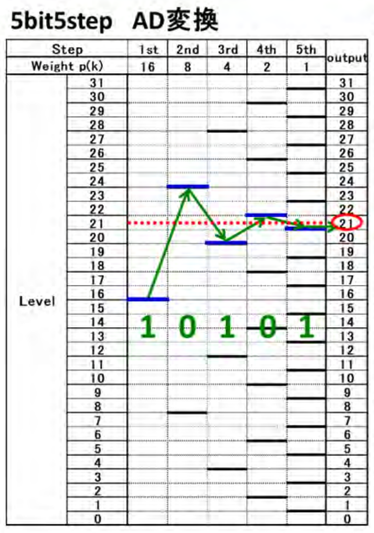
+整数論

従来のDACが使用不可

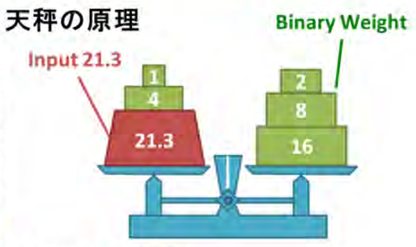
研究目的
逐次比較近似AD変換器の整数論を用いた冗長設計
➡ さらに高性能化・高速化

逐次比較近似(SAR)ADC

5bit5step　AD変換



天秤の原理



Dout = 10101

16 + 8 − 4 + 2 − 1 + 0.5 − 0.5= 21

出力値とディジタル表現が1対1に対応

一回の判定間違えは誤った出力に直結

冗長性を持つSAR ADC

冗長：余分や余裕のこと

SAR ADCに適用

時間の冗長性を利用
判定ステップ数を増加

5step ⇒ 6step など

ディジタルコードによる表現方法増加

SAR ADC ― 誤り耐性向上 / 変換速度向上

二進重み　1,2,4,8,16
↓
非二進重み　1,2,3,6,10,16

フィボナッチ数列で冗長性を実現！！

フィボナッチ数列を用いた冗長SAR ADC設計

フィボナッチ数列とは？

フィボナッチ数列

$$F_0 = 0$$
$$F_1 = 1$$
$$F_{n+2} = F_n + F_{n+1}$$

Leonardo Fibonacci
(伊:1170~1250年頃)

初項から計算していくと…
$0, 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, 233 \ldots$

隣り合う2項の比率を考えると…

$$\lim_{n\to\infty} \frac{F_n}{F_{n-1}} = 1.618033988749895 = \varphi$$

収束比率 φ
黄金比（約1.6進数）

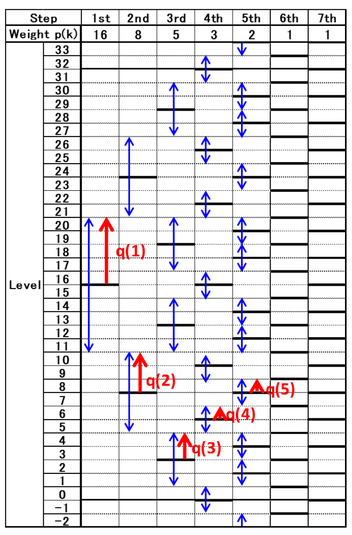
整数で1.6進数を表現可能

フィボナッチ数を用いたSAR ADC

2点の性質を新発見！
①補正可能範囲q(k)は
　必ずフィボナッチ数F_{M-k-1}になる
②補正可能範囲q(k)は
　必ず次のステップのq(k+1)に接する

性質②より…
➢ **信頼性の高い設計**
q(k)は最小のステップ数で広い範囲を補正可能

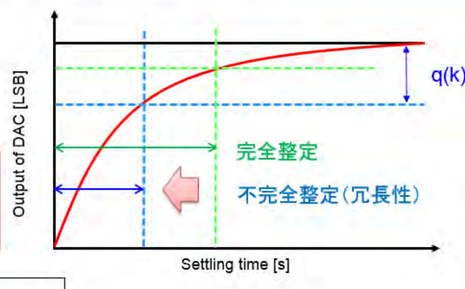
➢ **基数の基準**
冗長SAR ADCの基数基準は黄金比である

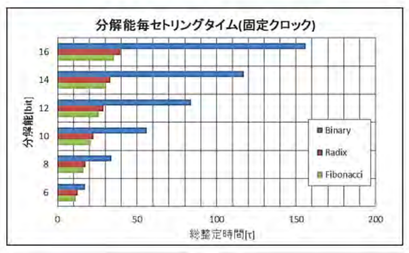


フィボナッチ手法による高速化

二進探索SAR ADC
-完全整定
➡ 変換時間 長

非二進重みSAR ADC
-不完全整低
➡ 変換時間 短





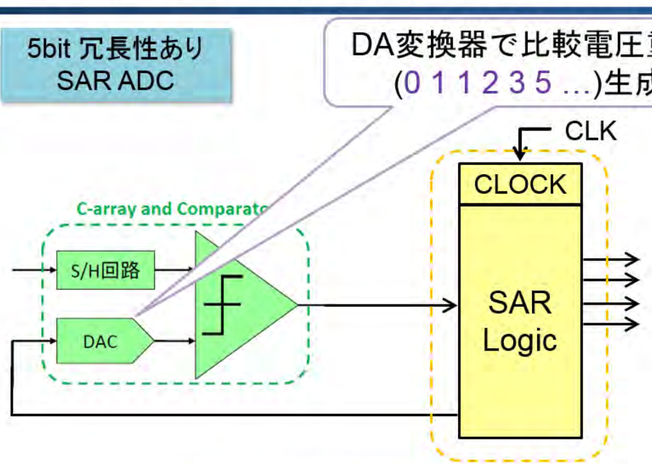
固定クロックで
フィボナッチSAR ADCは
Radix SAR ADCよりも高速！

黄金比重み付けDA変換器の構成

SAR ADCにおける黄金比重みDAC

5bit 冗長性あり
SAR ADC

DA変換器で比較電圧重み
(0 1 1 2 3 5 …)生成



DACでフィボナッチ数重み付け

DA変換器の新提案回路

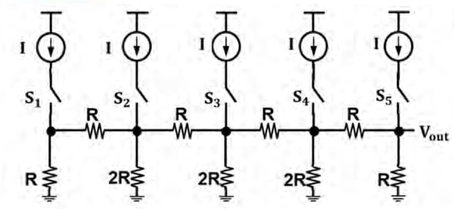
新しい発見！

R-2R抵抗ラダー回路
➡ 2進重みの電圧発生

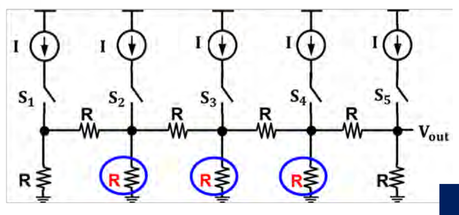
すべての抵抗がR

R-R抵抗ラダー回路
➡ フィボナッチ重みの電圧発生

容易にフィボナッチ
対応DA変換器実現可！

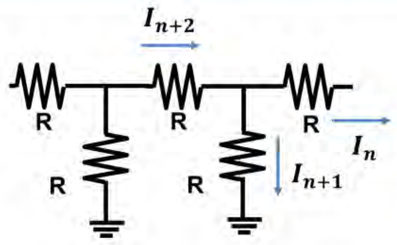


R-2R抵抗ラダー回路



R-R抵抗ラダー回路

フィボナッチ数列重み付けの原理



- $I_{n+2} = I_n + I_{n+1}$

フィボナッチ数列の原理

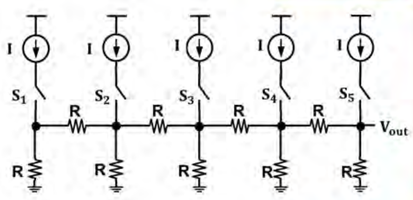
- $F_{n+2} = F_{n+1} + F_n$

0,1,1,2,3,5, …

R-Rラダー回路の電流分割は
フィボナッチ数列の定義と相似

フィボナッチ数列を出力する回路

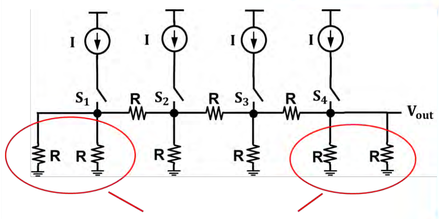
R終端回路



$$V_{out}(m) = \left(\frac{F_{2(n-m)+1}}{F_{2n}}\right) IR$$

フィボナッチの奇数項の出力 $= 1,2,5,13 \ldots$

R//R終端回路

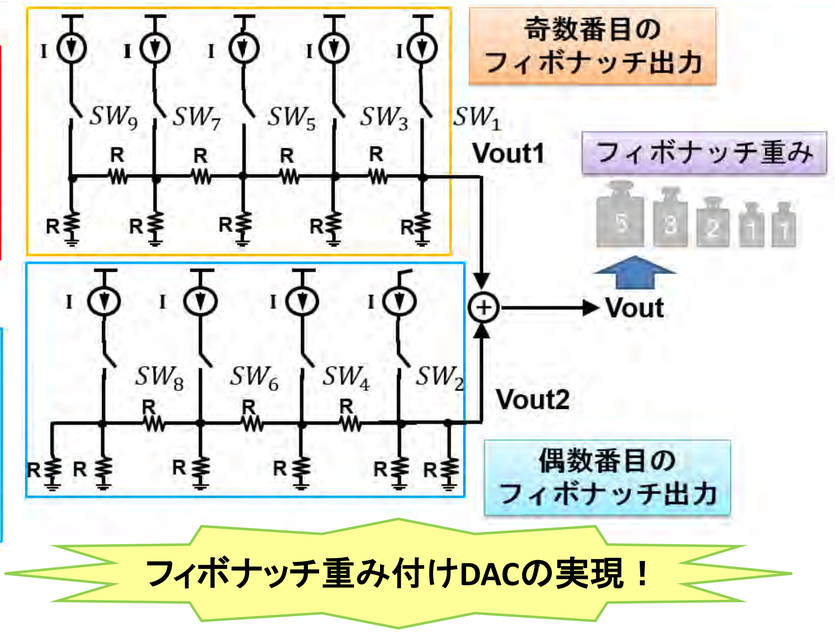


$$V_{out}(m) = \left(\frac{F_{2(n-m)}}{F_{2n+1}}\right) IR$$

フィボナッチの偶数項の出力 $= 1,3,8,21 \ldots$

両端の回路を　R　→　R//R

フィボナッチ重み付けR-RラダーDAC



フィボナッチ重み付けDACの実現！

まとめ

まとめ

◆フィボナッチ冗長設計対応
- 冗長設計による補正力がUP
- 整定時間の短縮

◆DACで重み付け可能
- SAR ロジック回路の簡略・小規模化

➡ 簡単な回路構成で実現可能!!

参考文献

[1].小林佑太朗、小林春夫
「逐次比較近似ADCの整数論に基づく冗長アルゴリズム設計」
電気学会、電子回路研究会、島根（2014年7月）
[2].Y. Kobayashi, H. Kobayashi
"SAR ADC Algorithm with Redundancy Based on Fibonacci Sequence"
International Conference on Analog VLSI Circuits,
Ho Chi Minh,Vietnam (Oct.,2014)

整数論に基づく無理数近似値アナログ信号生成回路

群馬大学 理工学部 電子情報理工学科

平井 愛統，
桑名杏奈，小林 春夫

Outline

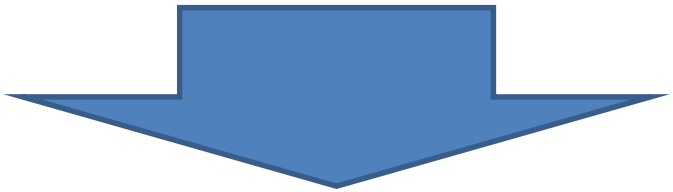
- 研究背景・目的
- R-r ラダーネットワーク
  - 合成抵抗の収束値
  - 貴金属数比のラダー , シミュレーション検証
  - $\sqrt{2}$の比のラダー, シミュレーション検証
- 各段の抵抗値が異なるラダー
  - 合成抵抗と連分数展開との対応
  - ネイピア数e,円周率π
- まとめ

Outline

- 研究背景･目的
- R-r ラダーネットワーク
  - 合成抵抗の収束値
  - 貴金属数比のラダー , シミュレーション検証
  - $\sqrt{2}$の比のラダー, シミュレーション検証
- 各段の抵抗値が異なるラダー
  - 合成抵抗と連分数展開との対応
  - ネイピア数e,円周率π
- まとめ

研究背景・目的

- IC内での抵抗値はばらつく
  「比精度」はよい（0.1% 程度の誤差）
- 無理数は連分数展開で表せる

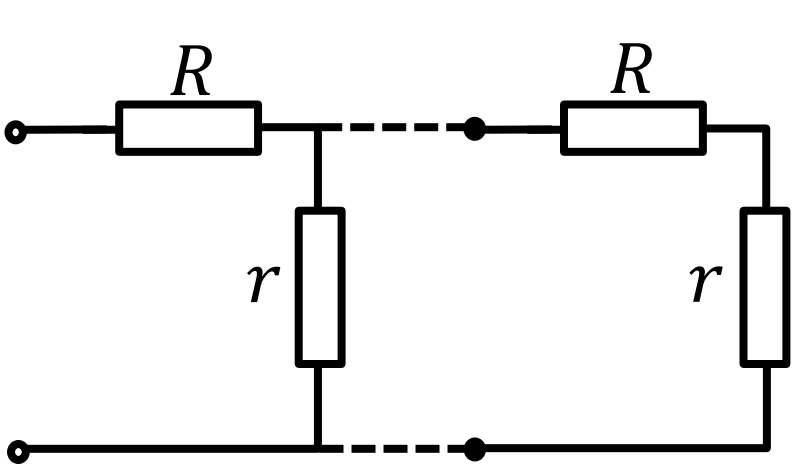


→ 整数比の抵抗でのネットワークで
無理数(近似)の比抵抗を構成
→無理数アナログ信号を生成

Outline

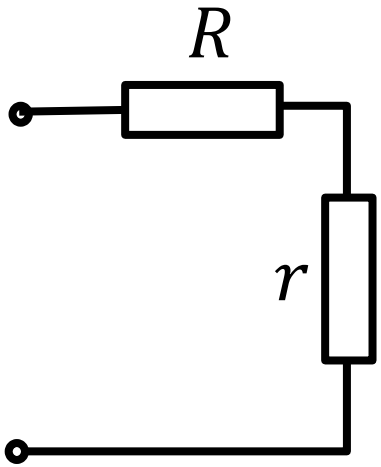
- 研究背景･目的
- R-r ラダーネットワーク
  - 合成抵抗の収束値
  - 貴金属数比のラダー , シミュレーション検証
  - $\sqrt{2}$の比のラダー, シミュレーション検証
- 各段の抵抗値が異なるラダー
  - 合成抵抗と連分数展開との対応
  - ネイピア数e,円周率π
- まとめ

R–rラダーの合成抵抗値



分解

接続



R-r 抵抗ラダー
：Rとrで構成される抵抗ラダー

R-r ラダーの単位

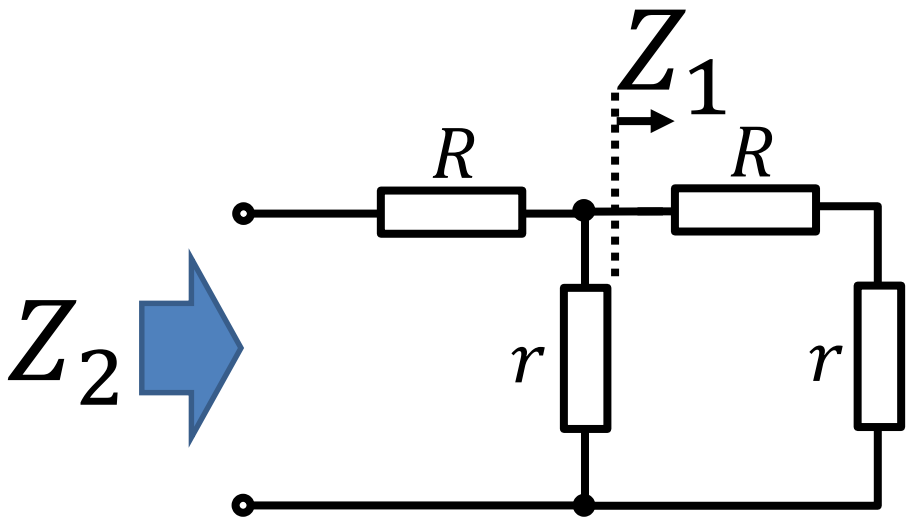
抵抗ラダーはRとrの単位に分割できる

R–rラダーの合成抵抗値

- 接続する”単位”を増やす

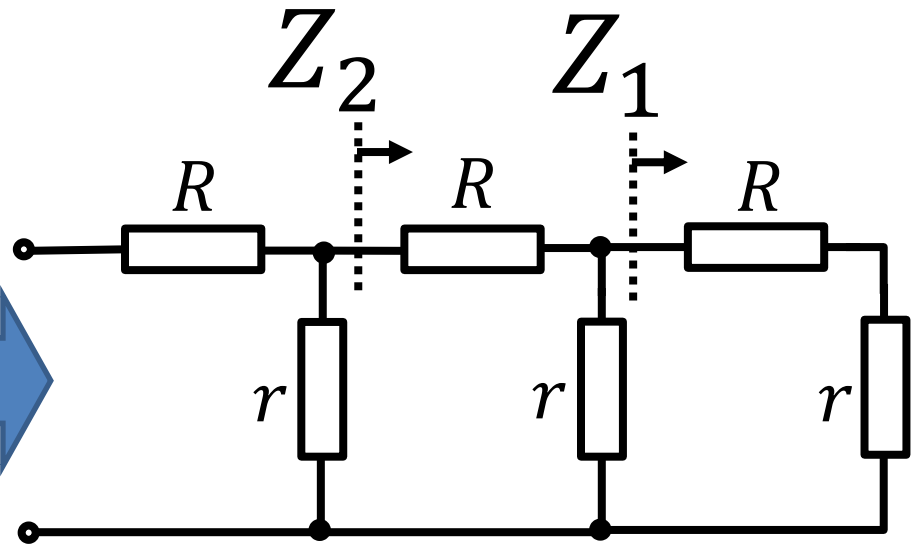
二段ラダーZ_2

$$Z_2 = R + \frac{r(R+r)}{r+(R+r)}$$



三段ラダーZ_3

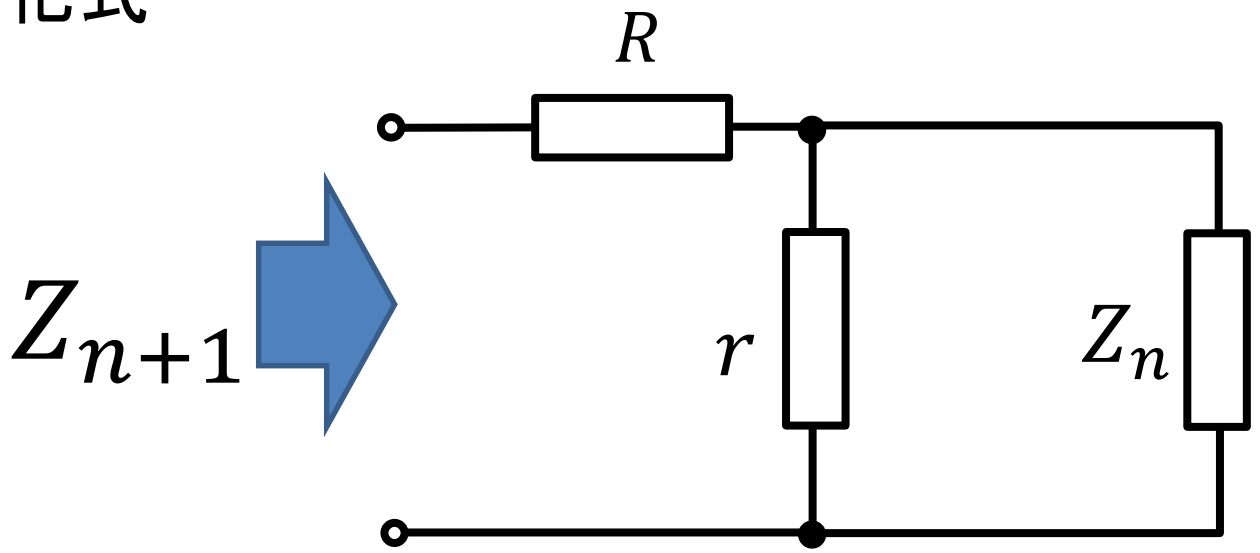
$$Z_3 = R + \frac{rZ_2}{r+Z_2}$$



n段R–rラダーの合成抵抗

$$Z_{n+1} = R + \frac{rZ_n}{r + Z_n}$$
$$= \frac{(r+R)Z_n + rR}{Z_n + r}$$

→Z_nに関する漸化式



n段ラダーの合成抵抗

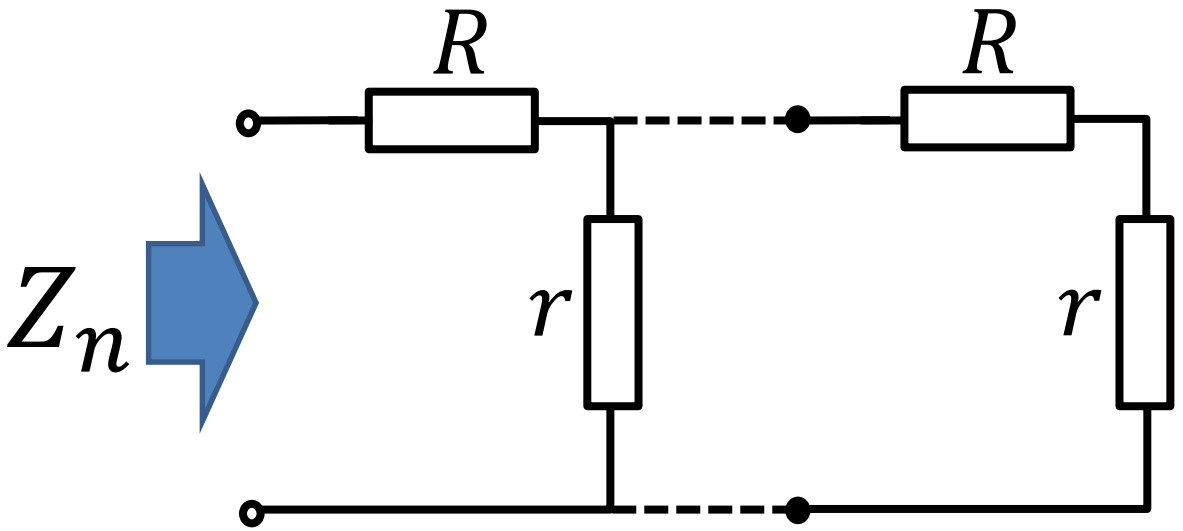
$$Z_n = \frac{\alpha k^n - \beta}{k^n - 1}$$

ただし、

$$\alpha = \frac{1}{2}\left(R + \sqrt{R^2 + 4rR}\right),$$

$$\beta = \frac{1}{2}\left(R - \sqrt{R^2 + 4rR}\right),$$

$$k = \frac{R + r - \beta}{R + r - \alpha}, 1 < k$$

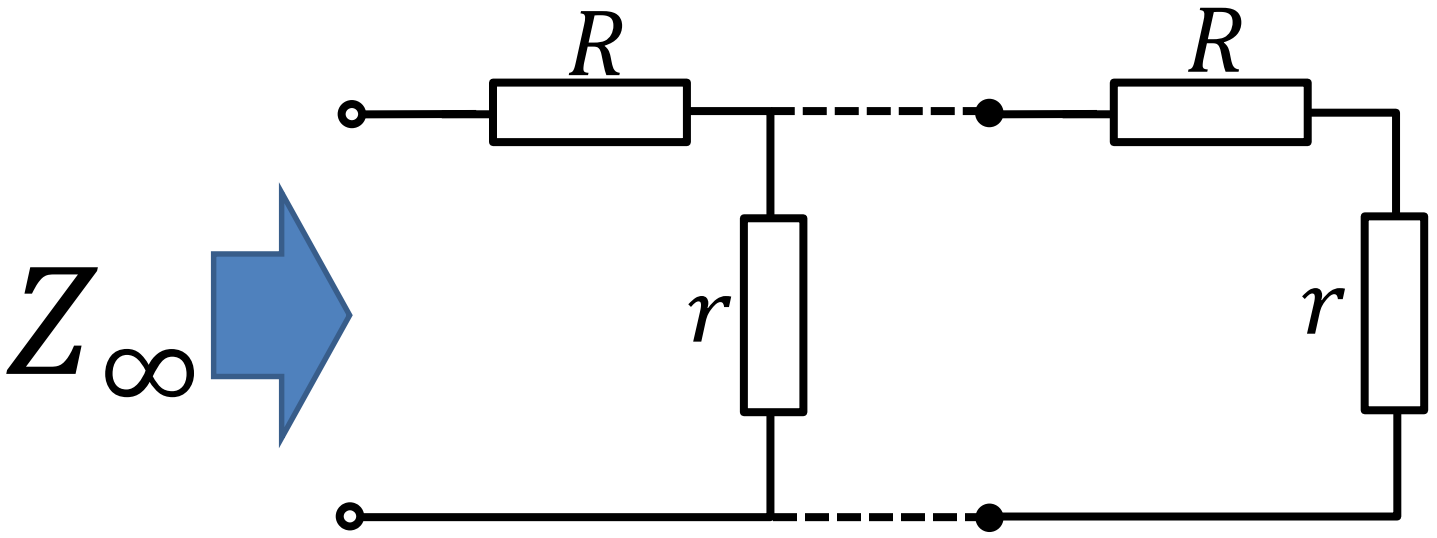


合成抵抗の収束値

$$Z_\infty = \lim_{n\to\infty} Z_n = \lim_{n\to\infty} \frac{\alpha - \beta k^{-n}}{1 - k^{-n}}$$

$$\rightarrow \alpha = \frac{1}{2}\left(R + \sqrt{R^2 + 4rR}\right)$$

$$Z_\infty = \frac{R}{2} + \frac{\sqrt{R(R+4r)}}{2}$$



Outline

- 研究背景･目的
- R-r ラダーネットワーク
  - 合成抵抗の収束値
  - 貴金属数比のラダー , シミュレーション検証
  - $\sqrt{2}$の比のラダー, シミュレーション検証
- 各段の抵抗値が異なるラダー
  - 合成抵抗と連分数展開との対応
  - ネイピア数e,円周率π
- まとめ

貴金属数 λ

- $x^2 - nx - 1 = 0$ の正の解

$$\lambda_n = \frac{n}{2} + \frac{\sqrt{n^2+4}}{2}$$

- nについて「第n貴金属数」
- $n = 1$: 黄金数ϕ

$$\phi = \frac{1+\sqrt{5}}{2}$$

- $n = 2$: 白銀数τ

$$\tau = 1 + \sqrt{2}$$

- $n = 3$: 青銅数ξ

$$\xi = \frac{3+\sqrt{13}}{2}$$

- **連分数形式**

$$\boldsymbol{\lambda_n = n + \cfrac{1}{n + \cfrac{1}{n + \cfrac{1}{n + \ddots}}}}$$

- $1 : (\lambda_n - 1)$を「第n 貴金属比」

参考: 岩本誠一、江口将生、吉良知文、「黄金・白銀・青銅 : 数と比と形と率と」(2008)
https://catalog.lib.kyushu-u.ac.jp/opac_download_md/15758/KJ00005471244.pdf

R-rラダーと貴金属数

R-r ラダーの抵抗

$$Z_\infty = \frac{R}{2} + \frac{\sqrt{R(R+4r)}}{2}$$

$$Z_{n+1} = R + \frac{rZ_n}{r+Z_n}$$

$$= \frac{R}{k}\left(k + \frac{1}{\frac{R}{kr} + \frac{R}{kZ_n}}\right)$$

$$= \frac{R}{k}\left(k + \cfrac{1}{\frac{R}{kr} + \cfrac{1}{k + \cfrac{1}{\frac{R}{kr} + \frac{1}{\ddots}}}}\right)$$

貴金属数

$$\lambda_n = \frac{n}{2} + \frac{\sqrt{n^2+4}}{2}$$

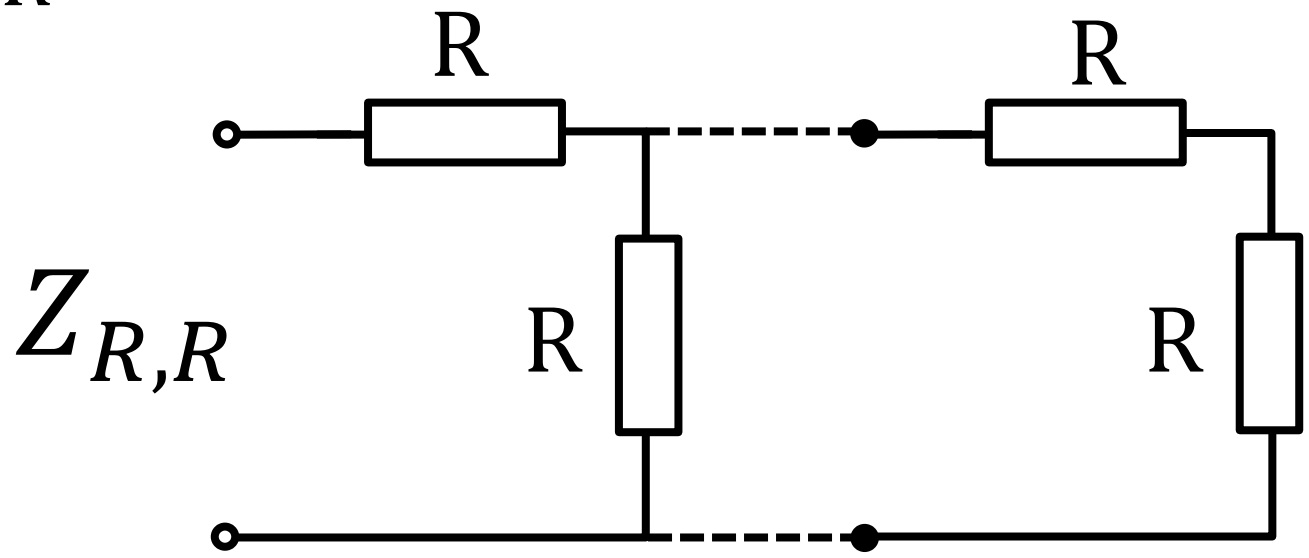
$$\lambda_n = n + \cfrac{1}{n + \cfrac{1}{n + \cfrac{1}{n + \frac{1}{\ddots}}}}$$

R-r ラダーの合成抵抗 ⇒ 貴金属数(無理数)比になりうる

R–Rラダー

$$Z_{R,R} = \frac{R}{2} + \frac{\sqrt{R(R+4r)}}{2}$$
$$= \frac{R}{2} + \frac{\sqrt{R(R+4R)}}{2}$$
$$= \frac{1+\sqrt{5}}{2} R$$

黄金数ϕラダー



2R–0.5Rラダー

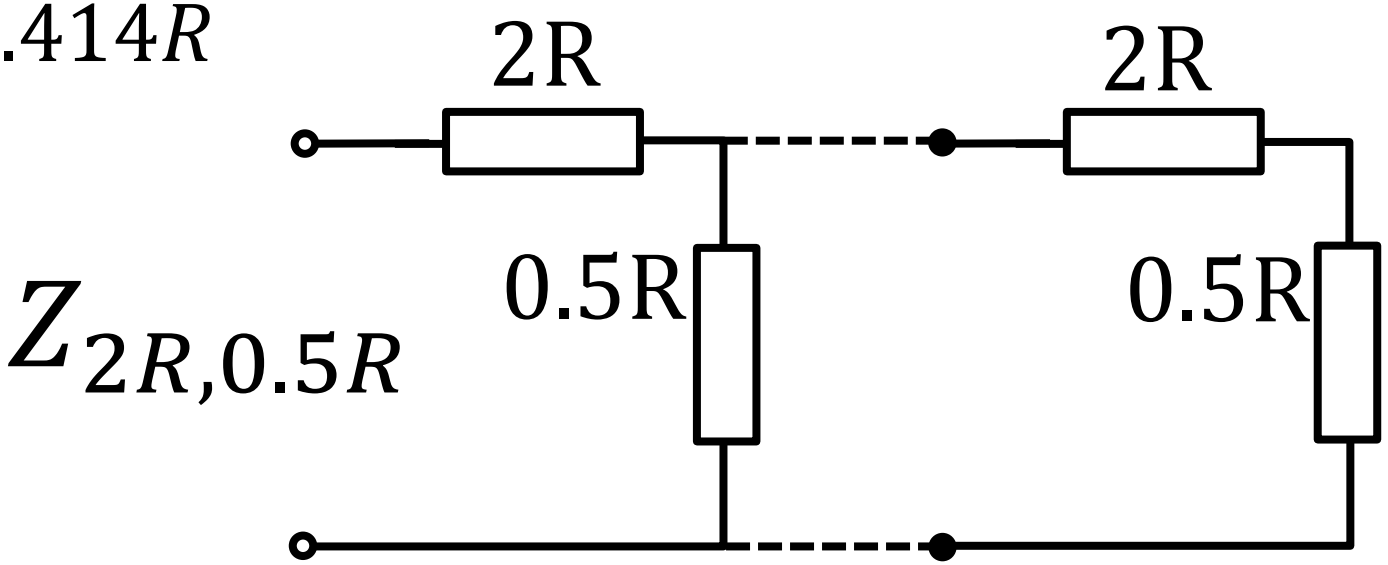
$$Z_{2R,0.5R} = \frac{2R}{2} + \frac{\sqrt{2R(2R + 4 \cdot 0.5R)}}{2}$$

$$= R + \frac{2\sqrt{2R^2}}{2}$$

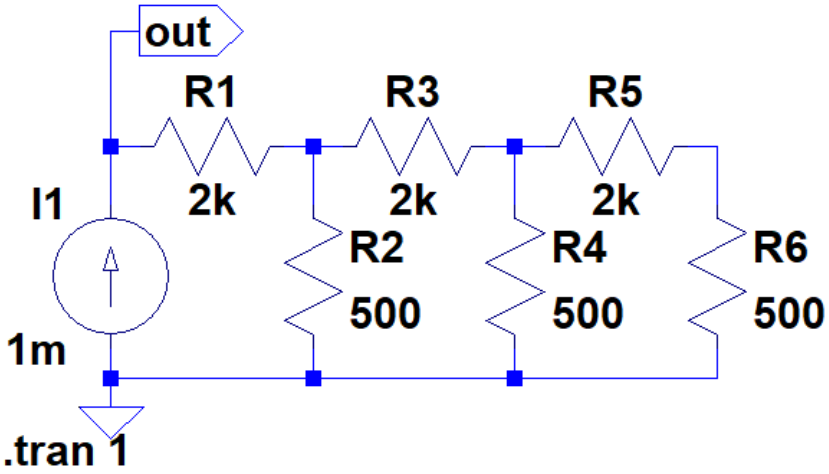
$$= \left(1 + \sqrt{2}\right)R$$

$$\approx 2.414R$$

白銀数τラダー



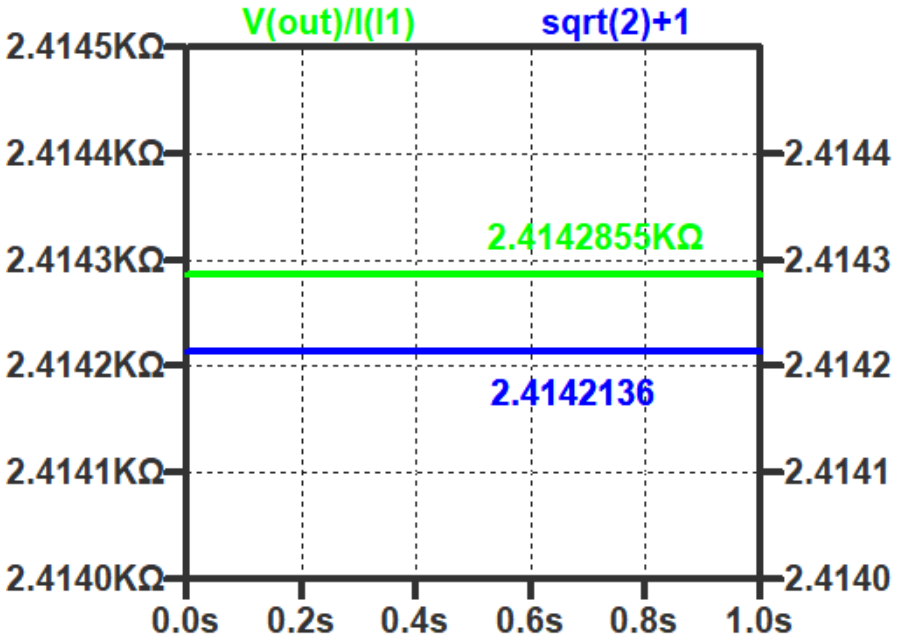
白銀数 $1+\sqrt{2}$ ラダー（3段）



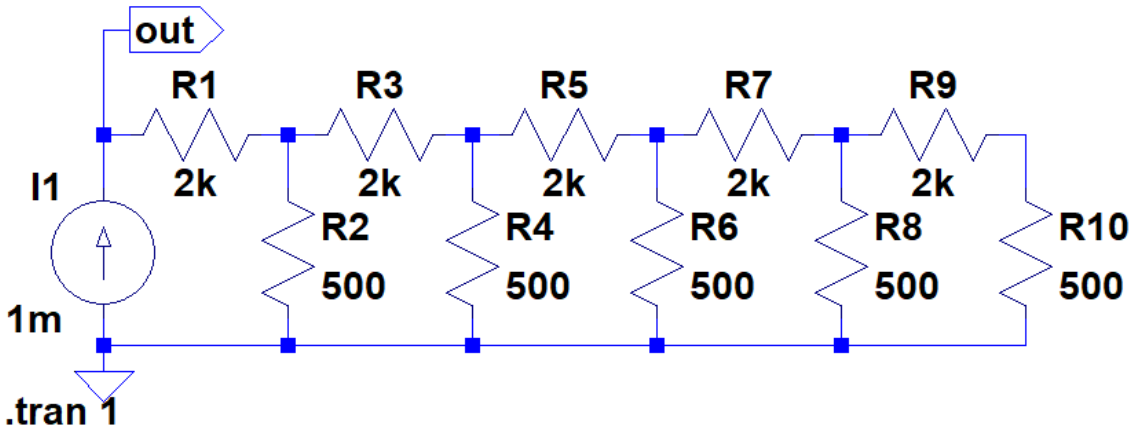
- 条件
  - R は 1 kΩ
  - 電流 1 mAを流して、電圧$V(\mathrm{out})$から抵抗を計算
- 結果

  抵抗値 2.4142855 kΩ

  $1+\sqrt{2}=2.4142135623730950\cdots$



白銀数 $1+\sqrt{2}$ ラダー（5段）

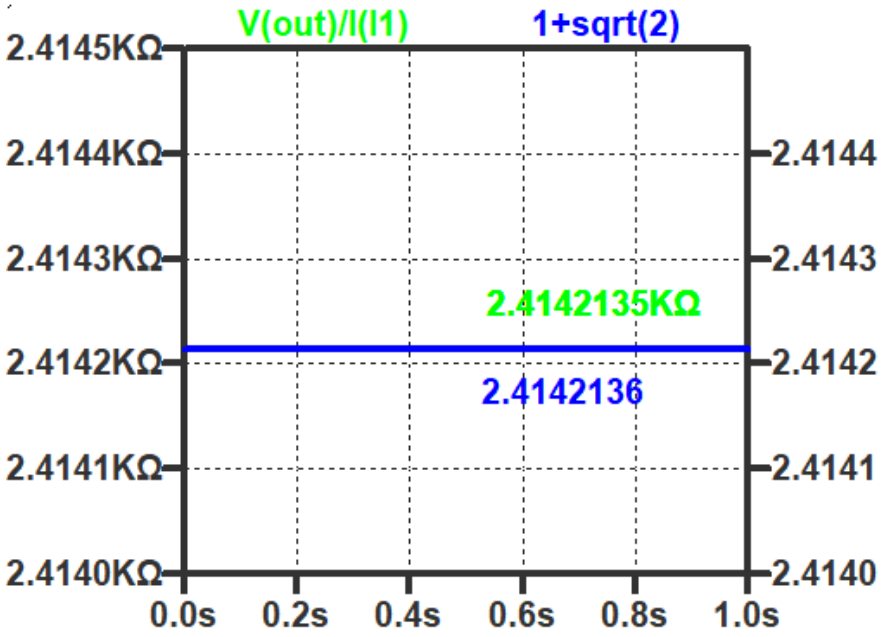


- 条件
  - R は 1 kΩ
  - 電流 1 mAを流して、電圧$V(\mathrm{out})$から抵抗を計算
- 結果

  抵抗値 2.4142135 kΩ

  $1+\sqrt{2}=2.414213562373095\cdots$

段数を増加 $\rightarrow 1+\sqrt{2}$に近づく



Outline

- 研究背景･目的
- R-r ラダーネットワーク
  - 合成抵抗の収束値
  - 貴金属数比のラダー , シミュレーション検証
  - $\sqrt{2}$の比のラダー, シミュレーション検証
- 各段の抵抗値が異なるラダー
  - 合成抵抗と連分数展開との対応
  - ネイピア数e,円周率π
- まとめ

$\sqrt{2}$ラダー

・合成抵抗の比を$\sqrt{2}$にしたい

$$\sqrt{2} = (1+\sqrt{2}) - 1$$

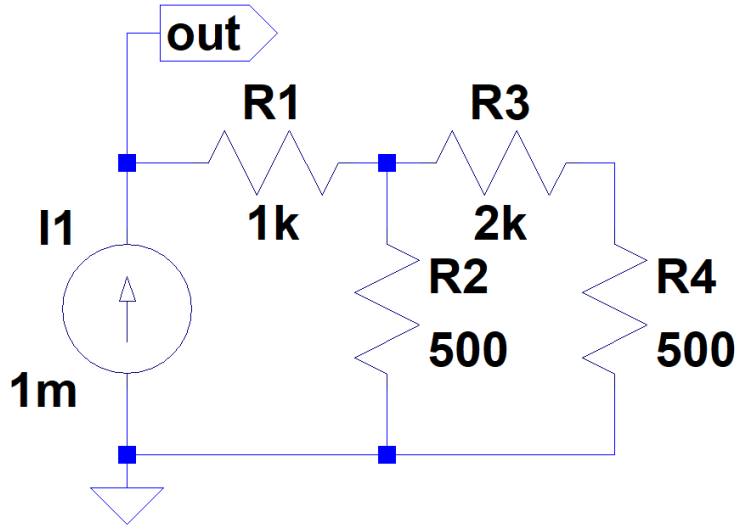
$$= 1 + \cfrac{1}{2 + \cfrac{1}{2 + \cfrac{1}{2 + \cfrac{1}{\ddots}}}}$$

$$Z_{2R,0.5R} = (1+\sqrt{2})R$$

$$Z_{2R,0.5R} - R = \sqrt{2}R$$

2R-0.5Rラダーの先頭の2RをRにする！

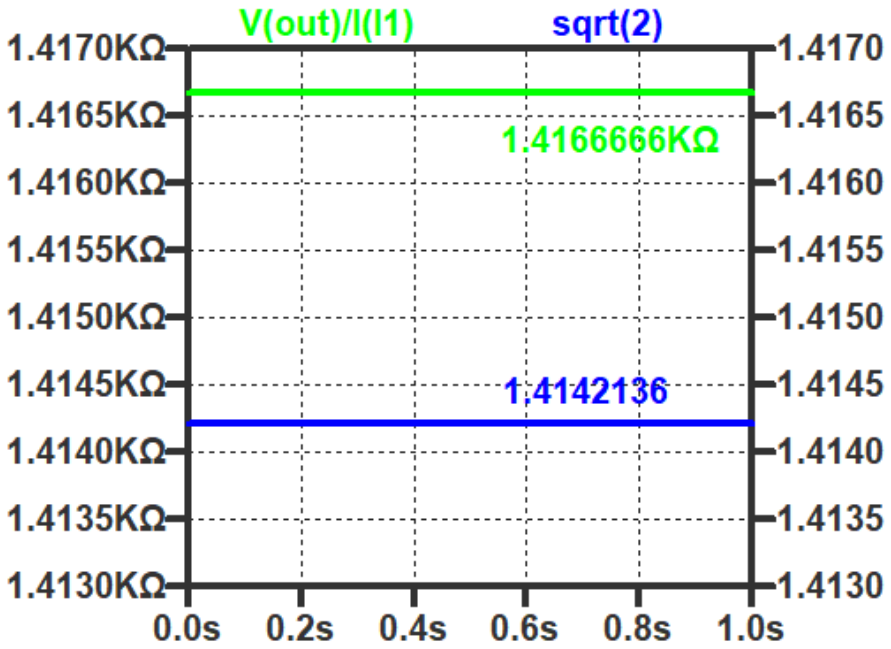
$\sqrt{2}$ラダー（2段）シミュレーション



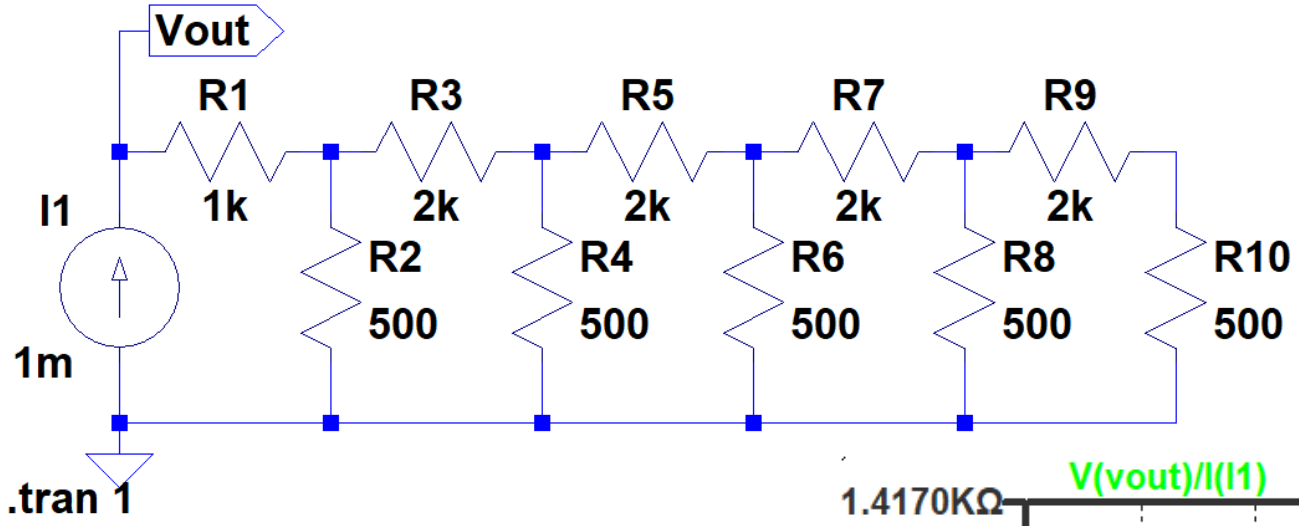
- 条件
  - R は 1 kΩ
  - 電流 1 mAを流して、電圧$V(\mathrm{out})$から抵抗を計算
- 結果

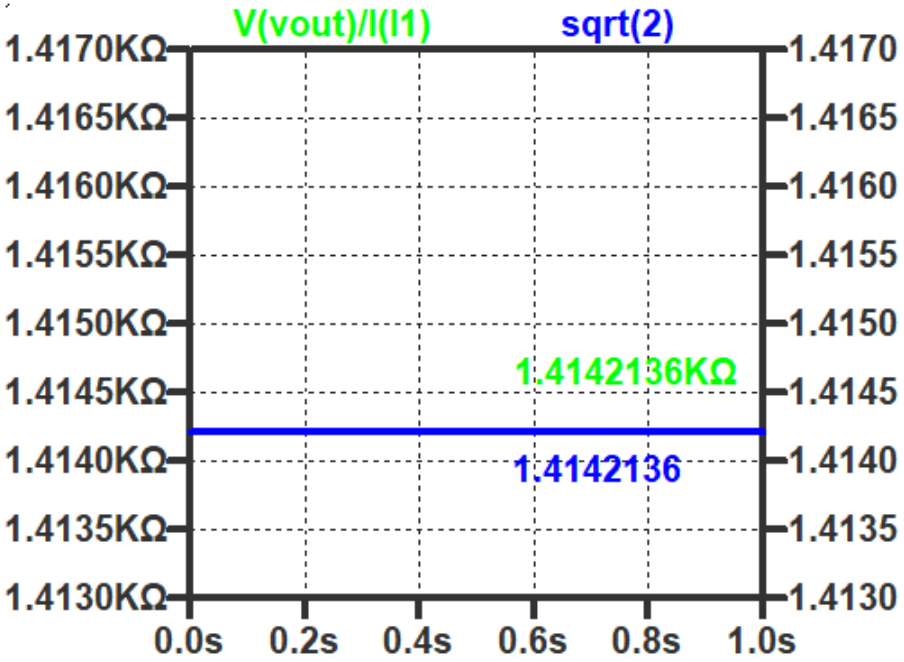
  抵抗値 1.41666666 kΩ

  $\sqrt{2} = 1.41421356237309\cdots$



$\sqrt{2}$ラダー（5段）シミュレーション





- 条件
  - R は 1 kΩ
  - 電流 1 mAを流して、電圧$V(\mathrm{out})$から抵抗を計算
- 結果

  抵抗値 1.4142136 kΩ

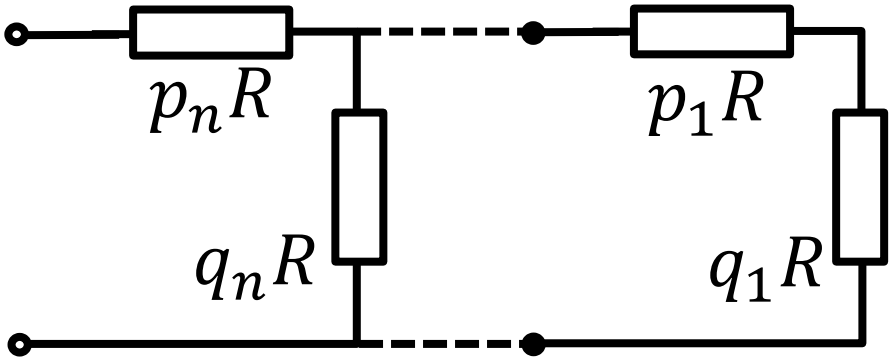
  $\sqrt{2} = 1.41421356237309\cdots$

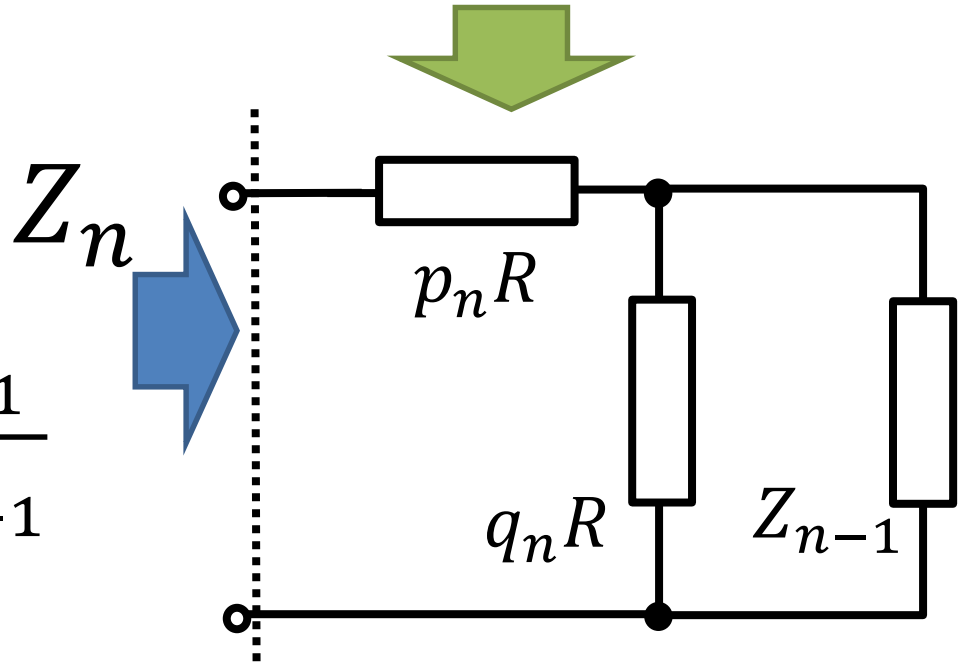
Outline

- 研究背景･目的
- R-r ラダーネットワーク
  - 合成抵抗の収束値
  - 貴金属数比のラダー , シミュレーション検証
  - $\sqrt{2}$の比のラダー, シミュレーション検証
- 各段の抵抗値が異なるラダー
  - 合成抵抗と連分数展開との対応
  - ネイピア数e,円周率π
- まとめ

各段の抵抗値が異なるラダー

n段目の抵抗を
p_n, q_nで重みづけ





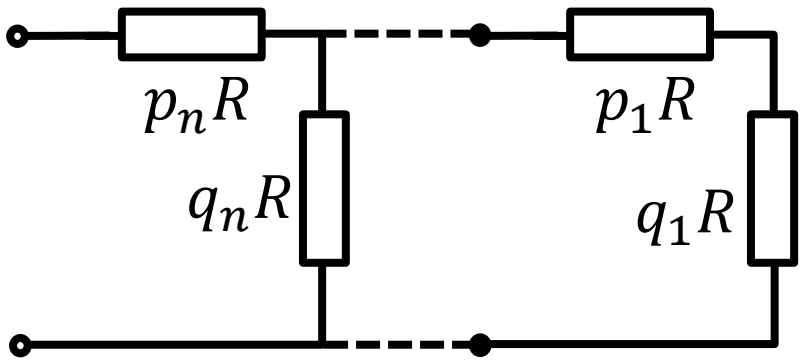
$$Z_n = p_nR + \frac{q_nR \cdot Z_{n-1}}{q_nR + Z_{n-1}}$$

各段の抵抗値が異なるラダー

$$Z_n = p_n R + \frac{q_n R \cdot Z_{n-1}}{q_n R + Z_{n-1}}$$

$$= R\left(p_n + \frac{1}{\frac{1}{q_n} + \frac{R}{Z_{n-1}}}\right)$$

$$= R\left(p_n + \frac{1}{\frac{1}{q_n} + \frac{1}{p_{n-1} + \frac{1}{\frac{1}{q_{n-1}} + \frac{1}{\ddots}}}}\right)$$



任意の数の連分数展開からp_nとq_nを決定
→抵抗の比は任意の数に

Outline

- 研究背景･目的
- R-r ラダーネットワーク
  - 合成抵抗の収束値
  - 貴金属数比のラダー , シミュレーション検証
  - $\sqrt{2}$の比のラダー, シミュレーション検証
- 各段の抵抗値が異なるラダー
  - 合成抵抗と連分数展開との対応
  - ネイピア数e,円周率π
- まとめ

ネイピア数eの比を持つ抵抗ラダー

- 無理数
- 自然対数の底

- 連分数展開が規則性を持つ

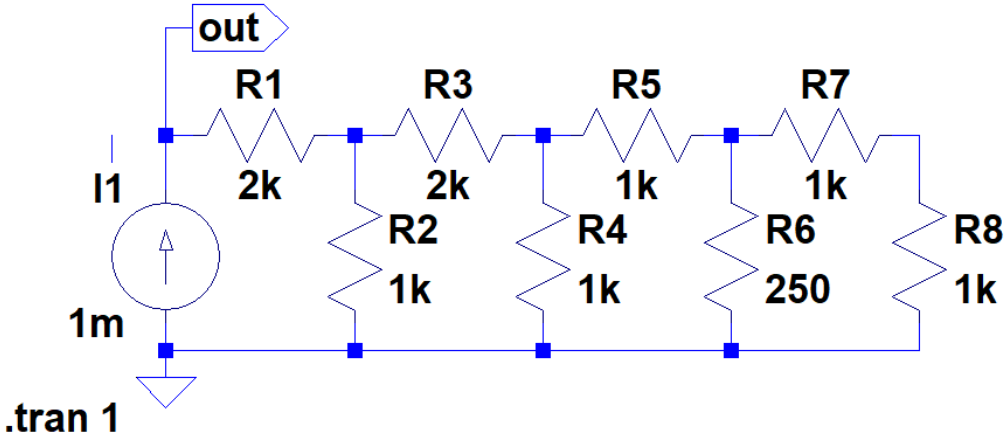
$$e = 2 + \cfrac{1}{1 + \cfrac{1}{2 + \cfrac{1}{1 + \cfrac{1}{1 + \cfrac{1}{1 + \ddots}}}}}$$

$$= [2; 1, 2, 1, 1, 4, 1, 1, 6, 1, 1, 8, 1, \cdots]$$

p_n → 連分数展開 整数部分の奇数番目
2, 2, 1, 1, 6, …

q_n → 連分数展開 整数部分の偶数番目の逆数
1, 1, 1/4, 1, 1 , …

ネイピア数eラダー（4段）

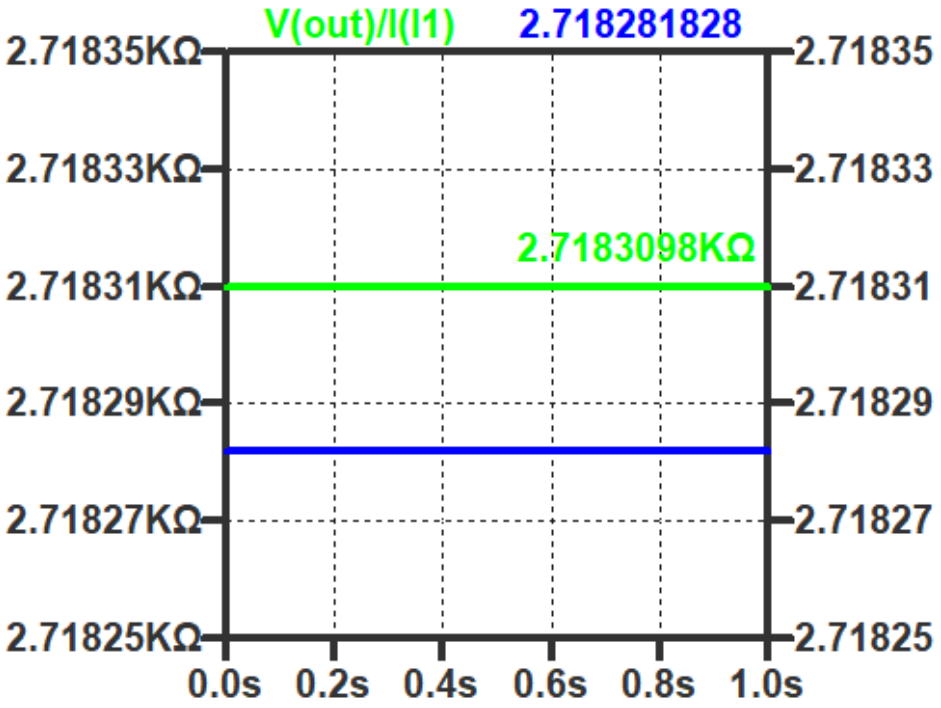


$$e \approx [2; 1, 2, 1, 1, 4, 1, 1]$$

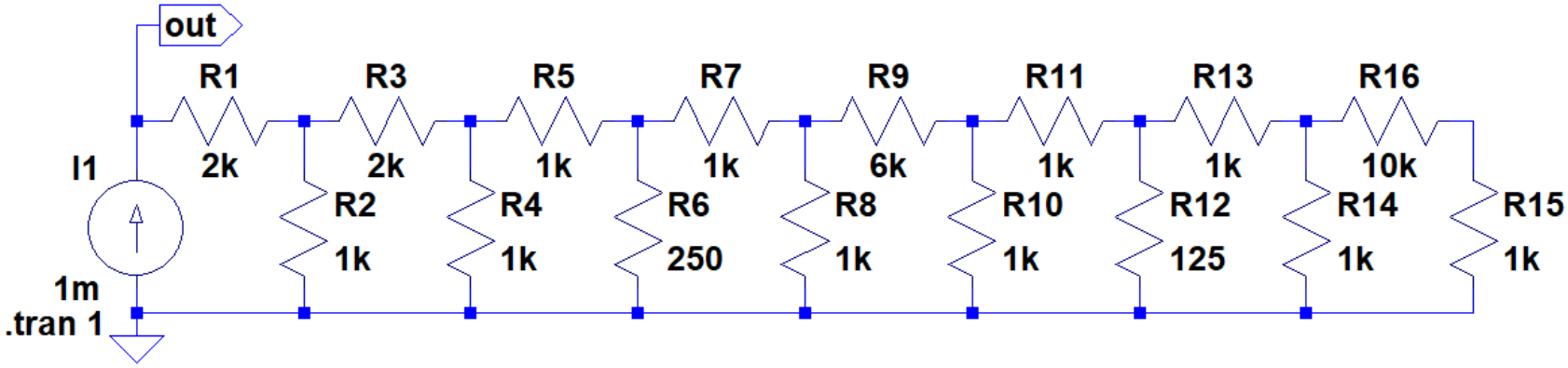
- 条件
  - R は 1 kΩ
  - 電流 1 mAを流して、電圧$V(\mathrm{out})$から抵抗を計算
- 結果

  抵抗値 2.7183098 kΩ

  (e = 2.7182818284595 36…)

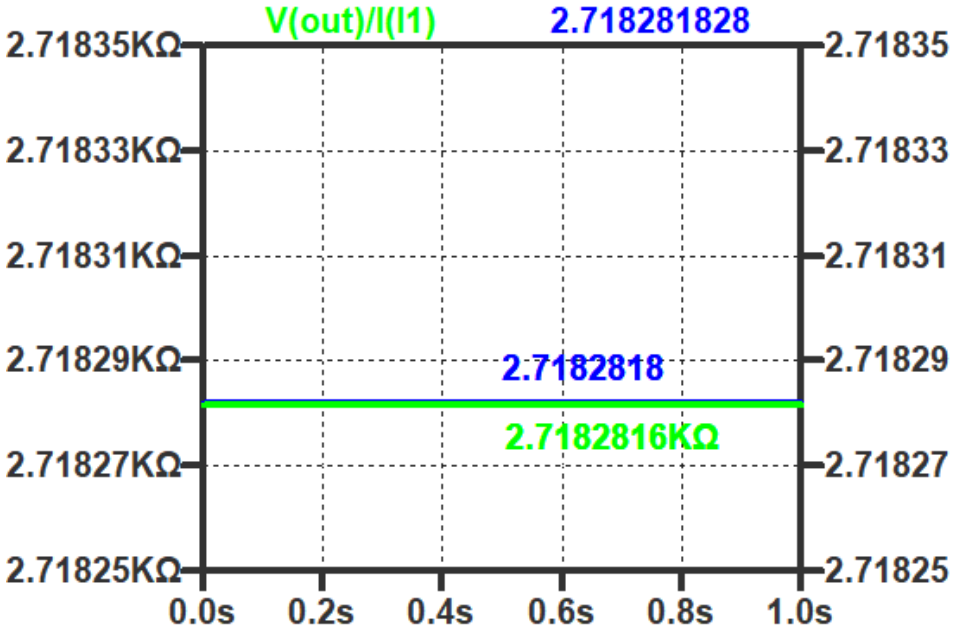


ネイピア数eラダー（8段）



$$e \approx [2; 1, 2, 1, 1, 4, 1, 1, 6, 1, 1, 8, 1, 1, 10, 1]$$

- 条件
  - R は 1 kΩ
  - 電流 1 mAを流して、電圧$V(\mathrm{out})$から抵抗を計算
- 結果
  抵抗値 2.7182816kΩ
  (e = 2.718281828459536…)



円周率 πの比を持つ抵抗ラダー

- 無理数
- 円の周と直径との比

- 規則性をもたない連分数表示

（少数第5位までで近似した連分数）

$$\pi \approx 3.14159$$

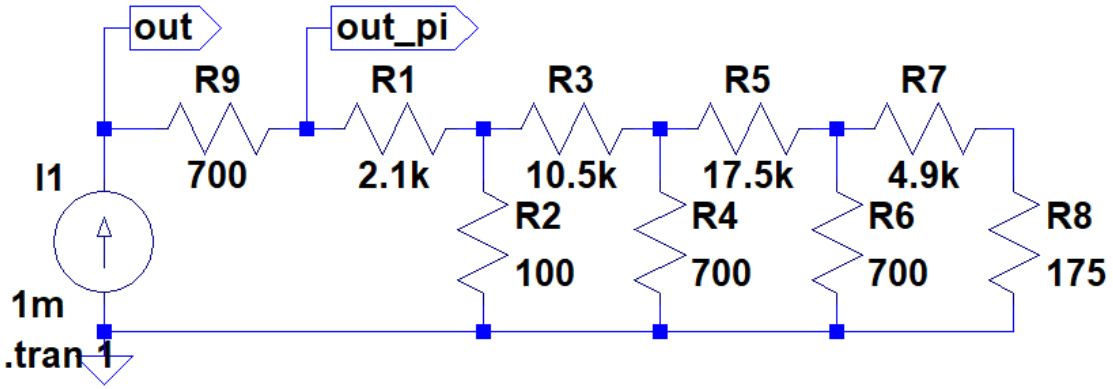
$$= 3 + \cfrac{1}{7 + \cfrac{1}{15 + \cfrac{1}{1 + \cfrac{1}{\ddots}}}}$$

$$= [3; 7, 15, 1, 25, 1, 7, 4]$$

p_n → 連分数展開 整数部分の奇数番目
3, 15, 25, 7

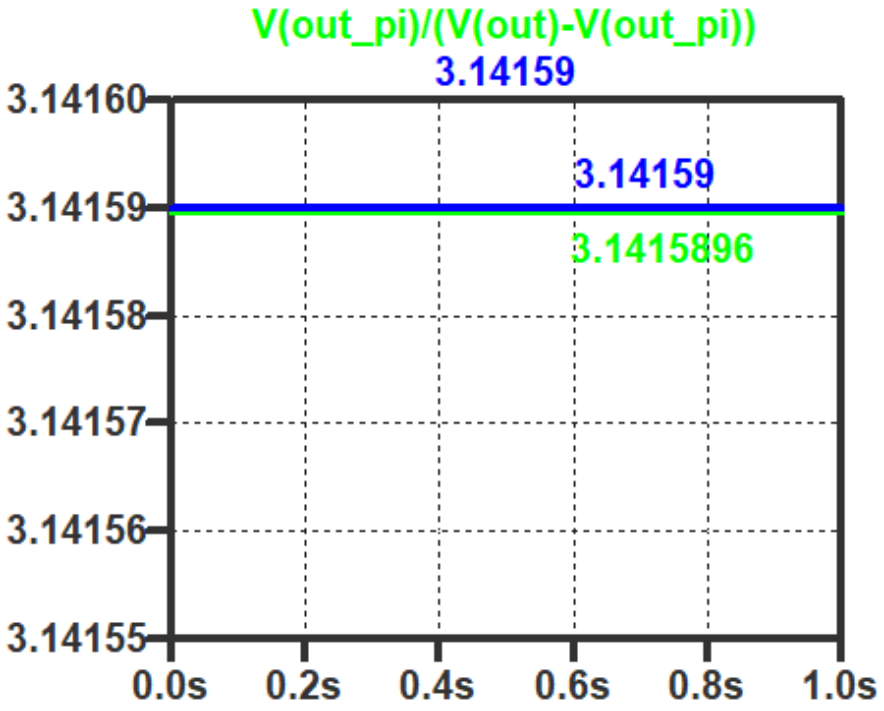
q_n → 連分数展開 整数部分の偶数番目の逆数
1/7, 1, 1, 1/4

円周率πのラダー（シミュレーション）



$$\pi \approx [3; 7, 15, 1, 25, 1, 7, 4]$$

- 条件
  - R は 700 Ω
  - Rと抵抗ラダーに電流 1 mAを流す
    抵抗にかかる電圧の比から抵抗比を計算
- 結果
  Rに対して3.1415896 倍の値
  （設計値 ：3.14159 倍）

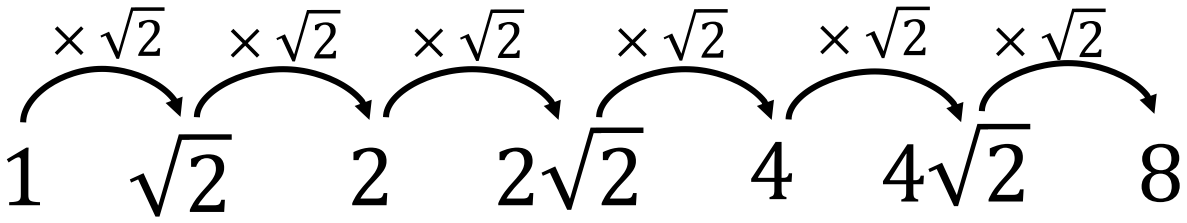


アナログ信号処理への応用

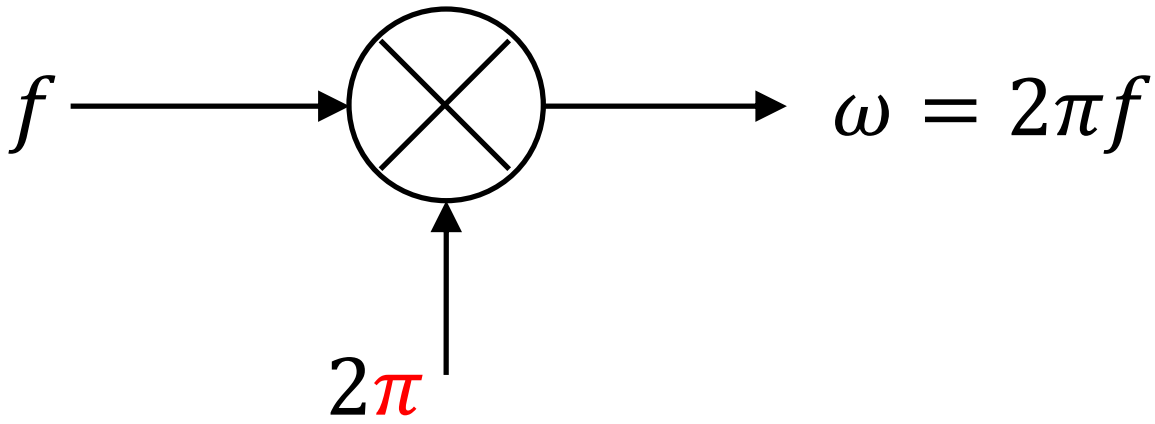
- ADCの冗長設計
  - 黄金比 $1:\phi-1$

  →フィボナッチ重み付け SAR（逐次比較近似）ADC
  - 白銀比 $1:\tau-1$

  →$\sqrt{2}$（白銀比）重み付け SAR ADC



-



Outline

- 研究背景･目的
- R-r ラダーネットワーク
  - 合成抵抗の収束値
  - 貴金属数比のラダー , シミュレーション検証
  - $\sqrt{2}$の比のラダー, シミュレーション検証
- 各段の抵抗値が異なるラダー
  - 合成抵抗と連分数展開との対応
  - ネイピア数e,円周率π
- まとめ

まとめ

＜まとめ＞

次を明らかにした。

- R-r ラダーの合成抵抗は「貴金属数」比を持つ
- 任意の数の連分数展開を用いて、
  その数の比を持つ抵抗を作ることができる
- 近似精度は用いる抵抗の数を増やすことでよくなる

数学によるアナログ信号処理の新機軸

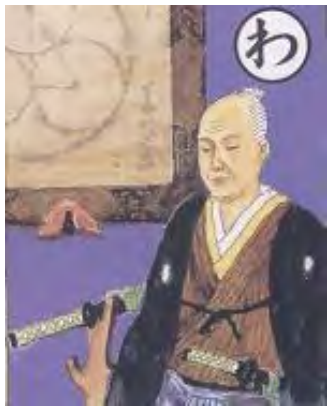
Ludolph van Ceulen

円周率 π

Leonhard Euler

ネイピア数 e

John Napier

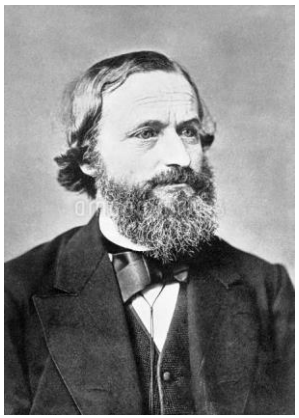


関孝和

連分数展開

Georg Simon Ohm

オームの法則



Gustav Robert Kirchhoff

キリヒホッフの法則

Carver Mead

抵抗ネットワーク

Z_nに関する漸化式

特性方程式

$$x = \frac{(R+r)x + rR}{x+r}$$
$$\leftrightarrow x^2 - Rx - rR = 0$$
$$\leftrightarrow x = \frac{R \pm \sqrt{(-R)^2 + 4rR}}{2}$$

重解があるとき

$$(-R)^2 + 4rR = 0$$
$$\leftrightarrow R(R+4r) = 0$$

$R > 0, r > 0$より重解なし

二項間一般分数系漸化式で、重解のない場合

$$a_{n+1} = \frac{pa_n + q}{ta_n + s}, (\text{特性方程式 } x = \frac{px+q}{tx+s})$$

→特性方程式の二つの解をα, βとして、

$$a_n = \frac{\alpha b_1 k^{n-1} - \beta}{b_1 k^{n-1} - 1}, b_n = \frac{a_n - \beta}{a_n - \alpha}, k = \frac{p - t\beta}{p - t\alpha}.$$

Z_nの一般式は

$$p = R + r, q = rR, t = 1, s = r,$$
$$\alpha = \frac{1}{2}\left(R + \sqrt{R^2 + 4rR}\right),$$
$$\beta = \frac{1}{2}\left(R - \sqrt{R^2 + 4rR}\right),$$
$$k = \frac{R + r - \beta}{R + r - \alpha}.$$

質疑・コメント

Q. Rとrの比だけでいいのか？
Rには絶対値が必要ではないか？

A. Rには具体的な値が必要。その値Rに対し、合成抵抗がスライド14・15のようにある比（$\frac{1+\sqrt{5}}{2}$や$1+\sqrt{2}$）を持つ。

Q. 今回はシミュレーションで検証をしているが実験もできそう。実際の誤差に対して、結果が強いのか弱いのかは検証が必要だと思う。

A. 素子の誤差に関しては、検証が必要だと考えている。ただ、図中左の抵抗ほど合成抵抗に与える影響が大きいことは確認している。

平成 30 年度　第 9 回 電気学会東京支部栃木・群馬支所 合同研究発表会　　ETG-19-58, ETT-19-58
於　小山高専　2019 年 3 月 4 日(月), 5 日(火)

整数論に基づく無理数近似値アナログ信号生成回路

平井　愛統*, 桑名 杏奈, 小林 春夫 (群馬大学)

Analog Signal Generator for Irrational Number Approximation Based on Number Theory

Manato Hirai* , Anna Kuwana, Haruo Kobayashi (Gunma University)

キーワード： 無理数値, 抵抗ネットワーク、連分数展開、整数論、信号生成回路
Keywords: Irrational Number, Resistor Network, Number Theory, Signal Generation Circuit

1. はじめに

抵抗ラダーネットワークは，電流や電圧をある比で分割するために使われ，アナログデジタル変換器/デジタルアナログ変換器(ADC/DAC)の内部回路やアナログ空間フィルタとして回路構成に組み込まれている[1, 2, 3]．例えば R-2R ラダーは，多段に接続した抵抗が，各ノードから見て等価的に一定の抵抗値を持つことを用いて，各ノードで電流を分割する。この性質を用いた R-2R DA 変換器は広く使われている[1].

R-R ラダーの各ノードから右側を見た合成抵抗は，フィボナッチ数に基づいた抵抗値になる [2]．また，隣り合う二つのフィボナッチ数の比は黄金数$\phi = (1+\sqrt{5})/2$に収束する [2, 4]．ここから，R-R ラダーの抵抗は黄金数に収束すると考え，この性質を用いて，二種類の抵抗R, rからなる抵抗ラダーの合成抵抗と，各段の抵抗値が異なる抵抗ラダーの合成抵抗を求め，それらの合成抵抗が収束する値を求めた．そして，それらの合成抵抗が連分数としてあらわせることを示した．集積回路内では比精度が良く実現できるので，ラダーを構成する抵抗値が整数比の場合には，設計値に対して精度のよい信号が発生できる．

本論文は，抵抗ラダーと無理数の関係について明らかにし，抵抗ラダーを用いて無理数近似アナログ信号を出力することを検討した内容を報告する．

2. R-r 抵抗ラダー

図 1 は，二種類の抵抗Rとrからなる抵抗ラダーである．この抵抗ラダーは図 2 のような単位に分解できる．

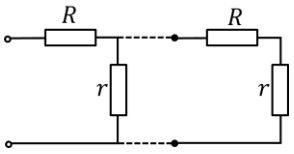


図 1 R-r 抵抗ラダー
Fig. 1 R-r resistor ladder

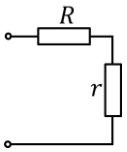


図 2 抵抗ラダーの単位
Fig. 2 Resistor ladder unit

抵抗ラダーを図 2 に示すような回路に分割すると，端子から見た抵抗値Z_1は次式であらわされる．

$$Z_1 = R + r \tag{1}$$

図 2 の抵抗の左に同じ抵抗を接続すると，図 3 のようになる．

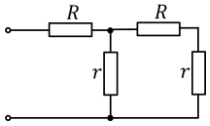


図 3　2 段抵抗ラダー
Fig. 3　2-stage resistor ladder

この抵抗値Z_2は，次式であらわされる．

$$Z_2 = R + \frac{r(R+r)}{r+(R+r)} \tag{2}$$

以上のようにして，Rとrの抵抗を多段に接続した場合，n段接続したときの抵抗Z_nと$n+1$段接続した場合の抵抗Z_{n+1}の間には次式のような関係が成り立つ．

$$\begin{aligned} Z_{n+1} &= R + \frac{rZ_n}{r+Z_n} \\ &= \frac{(r+R)Z_n + rR}{Z_n + r} \end{aligned} \tag{3}$$

これはZ_nに関する漸化式であり，これを解くことで二種類の抵抗R, rとラダーの段数nを定めた場合の合成抵抗値を求めた．Z_nは次式で与えられる．

$$
\begin{aligned}
Z_n &= \frac{\alpha k^n - \beta}{k^n - 1} \\
\alpha &= \frac{1}{2}\left(R + \sqrt{R^2 + 4rR}\right) \\
\beta &= \frac{1}{2}\left(R - \sqrt{R^2 + 4rR}\right) \\
k &= \frac{R + r - \beta}{R + r - \alpha}
\end{aligned} \tag{4}
$$

(4)式において，$1 < k$であるため，抵抗を接続する段数を増やしnの値が大きくなると，この抵抗値は次式の値Z_∞に収束する．

$$
\begin{aligned}
Z_\infty = \lim_{n\to\infty} Z_n &= \lim_{n\to\infty} \frac{\alpha - \beta k^{-n}}{1 - k^{-n}} \\
&\to \alpha = \frac{1}{2}\left(R + \sqrt{R^2 + 4rR}\right)
\end{aligned} \tag{5}
$$

$$
Z_\infty = \frac{R}{2} + \frac{\sqrt{R(R + 4r)}}{2} \tag{6}
$$

また，(3)式から R-r ラダーの合成抵抗は，次式のように連分数で表示することができる．kは整数である．

$$
\begin{aligned}
Z_{n+1} &= \frac{R}{k}\left(k + \frac{1}{\frac{R}{kr} + \frac{R}{kZ_n}}\right) \\
&= \frac{R}{k}\left(k + \cfrac{1}{\frac{R}{kr} + \cfrac{1}{k + \cfrac{1}{\frac{R}{kr} + \frac{1}{\ddots}}}}\right)
\end{aligned} \tag{7}
$$

図 4 のように，各段の抵抗の値が異なる抵抗ラダーを考える．抵抗ラダーのn段目の抵抗値は，ある抵抗値Rに対して，p_nRとq_nRとして，p_nとq_nによって重みづけられているとする．

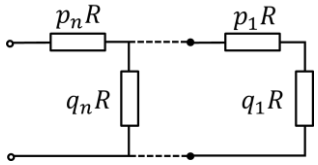


図 4 各段の抵抗値が異なる抵抗ラダー
Fig. 4 Resistor ladder
(different resistance value of each stage)

この時のn段抵抗ラダーの合成抵抗は次式のようになる．

$$
\begin{aligned}
Z_n &= p_nR + \frac{1}{\frac{1}{q_nR} + \frac{1}{Z_{n-1}}} \\
&= R\left(p_n + \frac{1}{\frac{1}{q_n} + \frac{R}{Z_{n-1}}}\right) \\
&= R\left(p_n + \cfrac{1}{\frac{1}{q_n} + \cfrac{1}{p_{n-1} + \cfrac{1}{\frac{1}{q_{n-1}} + \frac{1}{\ddots}}}}\right)
\end{aligned} \tag{8}
$$

(8)式から，任意の数について，その連分数表示に従って整数比の抵抗を接続することで，Rに対してその数の比を持つ抵抗を作ることができる．

3. 抵抗ラダーを用いた無理数（近似）信号の出力

〈3･1〉貴金属数

貴金属数は二次方程式$x^2 - nx - 1 = 0$の正の解であり，正の整数nの値に対して順に，第n貴金属数と呼ばれる[5]．二次方程式$x^2 - nx - 1 = 0$の正の解λ_nとその連分数表示は次式のようにあらわすことができる．

$$
\lambda_n = \frac{n + \sqrt{n^2 + 4}}{2} \tag{9}
$$

$$
\lambda_n = n + \cfrac{1}{n + \cfrac{1}{n + \cfrac{1}{n + \frac{1}{\ddots}}}} \tag{10}
$$

しばしば，$n = 1$の場合を黄金数ϕ，$n = 2$の場合を白銀数τ，$n = 3$の場合を青銅数ξと呼ぶ[5]．ϕ, τ, ξそれぞれの数値は，

$$
\begin{aligned}
\phi &= \frac{1 + \sqrt{5}}{2} \approx 1.618 \\
\tau &= 1 + \sqrt{2} \approx 2.414 \\
\xi &= \frac{3 + \sqrt{13}}{2} \approx 3.303
\end{aligned}
$$

である．

貴金属数が(10)式のように連分数展開できることから，(7)式においてkと$R/(kr)$をある整数値nとすることで，抵抗値が，無理数である貴金属数の比になることが予想される．

(6)式から，R-R ラダーの合成抵抗の収束する値$Z_{R,R}$は，次式の黄金数になる．

$$
Z_{R,R} = \frac{1 + \sqrt{5}}{2}R \tag{11}
$$

また，2R-0.5R ラダーの合成抵抗の収束する値$Z_{2R,0.5R}$は，(6)式と(7)式から次式の白銀数になる．

$$
\begin{aligned}
Z_{2R,0.5R} &= \left(1 + \sqrt{2}\right)R \\
&= R \cdot \left(2 + \cfrac{1}{2 + \cfrac{1}{2 + \cfrac{1}{2 + \frac{1}{\ddots}}}}\right)
\end{aligned} \tag{12}
$$

〈3･2〉$\sqrt{2}$の近似を出力するラダー

$\sqrt{2}$は連分数として次式のようにもあらわすことができる

[3].

$$\sqrt{2} = 1 + \cfrac{1}{2 + \cfrac{1}{2 + \cfrac{1}{2 + \cfrac{1}{\ddots}}}} \tag{13}$$

(13)式と(12)式の連分数部分を比較すると，(12)式の整数部分を 1 にしたものが(13)式である．これは，白銀数τから 1 を減じたものが$\sqrt{2}$であることによるものである．

2R-0.5R ラダーの先頭の抵抗を 2R から R に変えた場合を考える．この抵抗ラダーの合成抵抗の収束する値は次式であらわされる．

$$\begin{aligned} Z_{\sqrt{2}} &= R + \cfrac{1}{2 + \cfrac{1}{2 + \cfrac{1}{2 + \cfrac{1}{\ddots}}}} \cdot R \\ &= \sqrt{2}R \end{aligned} \tag{14}$$

〈3･2〉ネイピア数e，円周率πの近似を出力するラダー

ネイピア数eは連分数として次式のようにあらわすことができる[6].

$$\mathrm{e} = 2 + \cfrac{1}{1 + \cfrac{1}{2 + \cfrac{1}{1 + \cfrac{1}{\ddots}}}} \tag{15}$$

(15)式では省略したが，分数部分の分子をすべて 1 とした場合の整数部分を並べて表示すると，以下のように規則性を持つ[6].

$$\begin{aligned} \mathrm{e} &= 2.71828\cdots \\ &= [2,1,2,1,1,4,1,1,6,1,1,8,1,1,10,\cdots] \end{aligned} \tag{16}$$

図 4 の，各段の抵抗値が異なる抵抗ラダーにおいて，(8)式のp_nを連分数展開整数部分の奇数番目，q_nを偶数番目の逆数とすると，抵抗ラダーの合成抵抗のRに対する比は，近似的にネイピア数eになる．

円周率πを連分数展開すると，その整数部分は規則性を持たない[6]．円周率πを次式のように近似すると，連分数として次式のようにあらわされる[6].

$$\begin{aligned} \pi &\approx 3.14159 \\ &= [3,7,15,1,25,1,7,4] \end{aligned} \tag{17}$$

これを用いて，ネイピア数の場合と同様に抵抗を接続して抵抗ラダーの合成抵抗のRに対する比は，近似的に円周率πになる．

4. 回路シミュレータによる検証

〈4･1〉貴金属数

(12)式を用いて，Rを 1 kΩ とし，2R-0.5R ラダーに 1 mA の電流を流し，その時の出力電圧から，抵抗ラダーの合成抵抗を求めた．

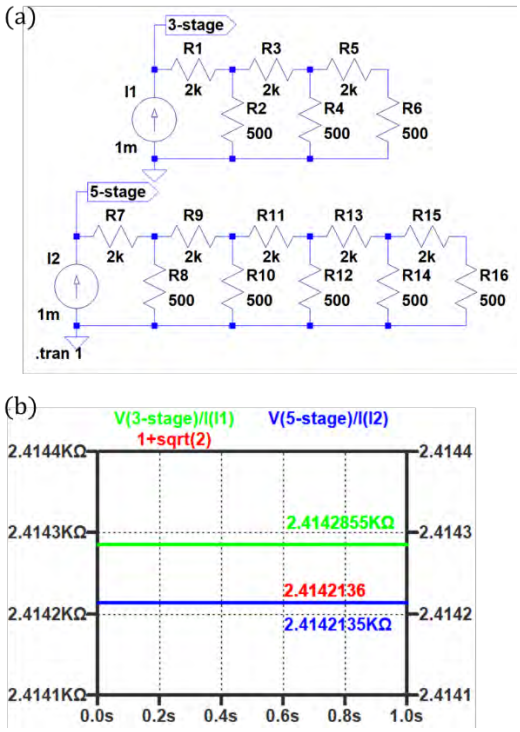
段数を 3 段･5 段にしてシミュレーションを行った．シミュレーションに用いた抵抗ラダーを図 5 (a)に，この時のシミュレーション結果を図 5(b)に示す．3 段の場合の合成抵抗の値は，2.4142855 kΩ であり，5 段の場合の合成抵抗の値は，2.4142135 kΩ であった．

段数の増加に伴い，合成抵抗の値は$\tau = 1 + \sqrt{2}$に近づいた．

〈4･2〉$\sqrt{2}$の近似を出力するラダー

(14)式を用いて，Rを 1 kΩ とし，先頭の抵抗を 2R から R に変えた 2R-0.5R ラダーに 1 mA の電流を流し，その時の出力電圧から，抵抗ラダーの合成抵抗を求めた．

段数を 2 段･5 段にしてシミュレーションを行った．シミュレーションに用いた抵抗ラダーを図 6 (a)に，この時のシミュレーション結果を図 6 (b)に示す．

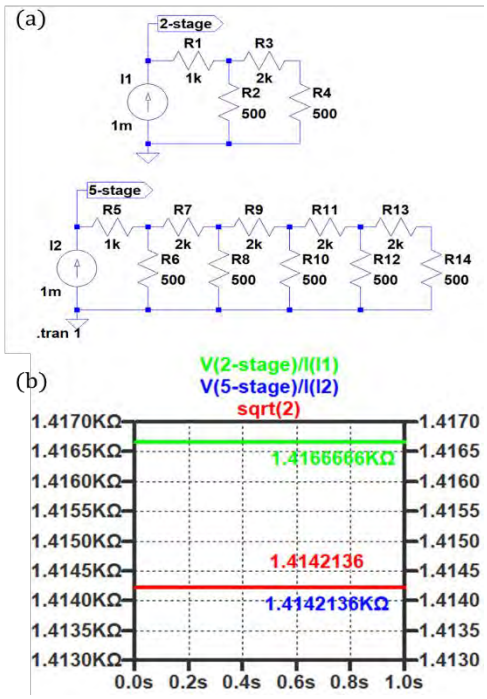


(a)回路図 (b)シミュレーション結果

図 5 2R-0.5R ラダー

(a) Circuit diagram. (b) Simulation result

Fig. 5 2R-0.5R ladder



(a)回路図 (b)シミュレーション結果

図 6 $\sqrt{2}$ラダー

(a) Circuit diagram. (b) Simulation result

Fig. 6 $\sqrt{2}$ resistor ladder

2 段の場合の合成抵抗の値は，1.4166666 kΩ であり，5 段の場合の合成抵抗の値は，1.4142136 kΩ であった．

段数を増やすほど，合成抵抗の値は(14)式の値に近づくことが確認できた．

〈4･3〉ネイピア数eと円周率πの近似を出力するラダー

(16)式を用いて，ネイピア数eを，次式のように近似した．

$$e \approx [2,1,2,1,1,4,1,1] \tag{18}$$

(18)式と(8)式から，(8)式のRを 1 kΩ，p_nをp_4から順に(18)式の奇数番目，q_nをq_4から順に(18)式の偶数番目の逆数として，合成抵抗がネイピア数になる 4 段の抵抗ラダーを作り，シミュレーションで確認した．シミュレーションに用いた回路図を図 7(a)に示す．

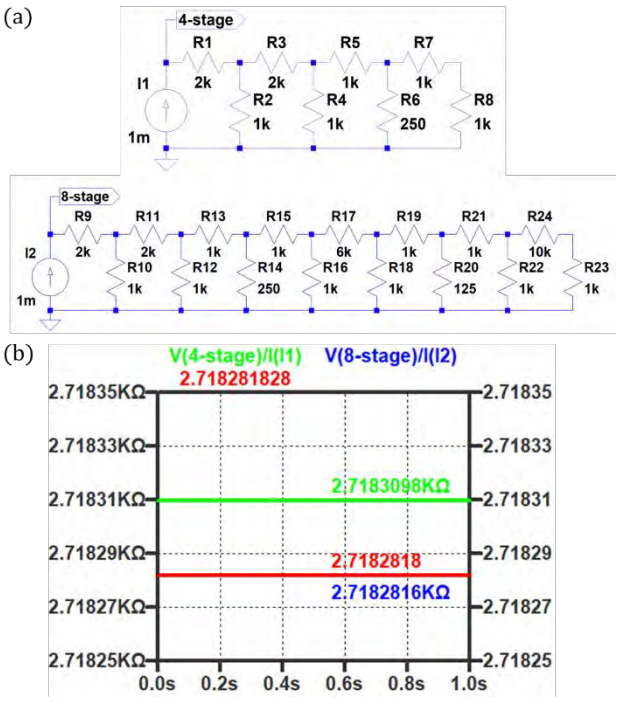
次にネイピア数eを次式のように近似し，ラダーの段数を増やした．

$$e \approx [2,1,2,1,1,4,1,1,6,1,1,8,1,1,10,1] \tag{19}$$

(19)式で近似した値を用いて，合成抵抗がネイピア数になる 8 段の抵抗ラダーを作り，シミュレーションで確認した．シミュレーションに用いた回路図を図 7 (a)に示す．

4 段と 8 段のシミュレーション結果を図 7(b)に示す．4 段のラダーでの合成抵抗Z_eは，2.7183098 kΩ になり，8 段のラダーでの合成抵抗は 2.7182816 kΩ になった．

段数を増やすほど，合成抵抗の値がネイピア数eに近づくことが確認できた．



(a)回路図 (b)シミュレーション結果

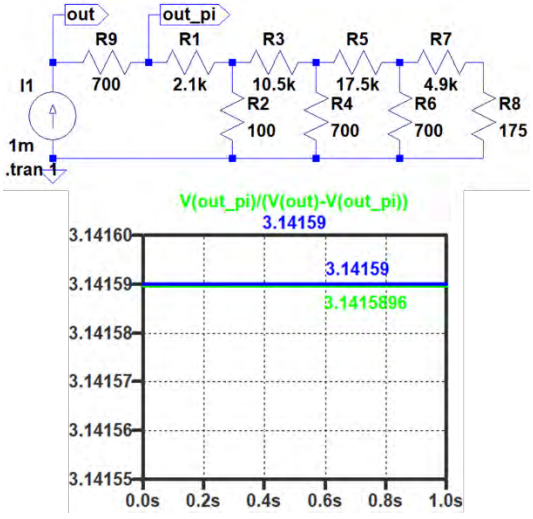
図 7 ネイピア数 e 近似出力ラダー

(a) Circuit diagram. (b) Simulation result

Fig. 7 Napier's constant approximation resistor ladder

(8)式と(17)式から，(8)式においてRを 700 Ω，p_nをp_4から順に(17)式整数部分の奇数番目，q_nをq_4から偶数番目の逆数として，抵抗Rに対して円周率πの比を持つ抵抗ラダーを作り，シミュレーションで確認した．シミュレーションに用いた回路図を図 8 (a)に，シミュレーション結果を図 8 (b)に示す．

図 8 (b)の出力電圧の比から，後段の抵抗ラダーと 700 Ω の抵抗 R9 の比は，(17)式で近似した円周率πに一致した．



(a)回路図 (b)シミュレーション結果

図 8 円周率π 近似出力ラダー

(a) Circuit diagram. (b) Simulation result

Fig. 8 π approximation resistor ladder

5. まとめ

抵抗ラダーを用いて無理数比の電圧を生成する回路を検討した．対象無理数を連分解することでその抵抗ラダーの構成することができる．例として黄金比，貴金属数，$\sqrt{2}$，ネイピア数e, 円周率πの場合の構成を示し，いくつかをSPICEシミュレーションで動作を検証した．

文　　献

[1] F. Maloberti, Data Converters, Springer (2007).
[2] Y. Kobayashi, S. Shibuya, T. Arafune, S. Sasaki, H. Kobayashi, “SAR ADC Design Using Golden Ratio Weight Algorithm”, The 15th International Symposium on Communications and Information Technologies, Nara, Japan (Oct. 2015).
[3] H. Kobayashi, J. L. White and A. A. Abidi, “An Active Resistor Network for Gaussian Filtering of Images", IEEE Journal of Solid-State Circuits, vol.26, no.5, pp.738-748 (May, 1991)
[4] 櫻井進、雪月花の数学、祥伝社黄金文庫　(2010)
[5] 岩本誠一、江口将生、吉良知文、「黄金・白銀・青銅：数と比と形と率と」(2008)
https://catalog.lib.kyushu-u.ac.jp/opac_download_md/15758/KJ00005471244.pdf
[6] 芹沢正三、数論入門、講談社ブルーバックス　(2008)

群馬県和算研究会　和算ジャーナル No.3, 2019 年 5 月 1 日発行　ISSSN 2432-8472

連分数展開を用いる近似無理数の抵抗回路

― 零約術の電気回路への応用 ―

Resistance Circuit of Irrational Number Approximation
Based on Continued Fractions

平 井 愛 統 , 桑 名 杏 奈 , 小 林 春 夫 (群馬大学)

Manato Hirai , Anna Kuwana , Haruo Kobayashi (Gunma University)

Abstract

This paper describes the method to generate irrational number ratio signals by using a resistor ladder and the relation between resistor ladders and irrational numbers. Irrational numbers are expressed as simple continued fractions configured by integers. The combined resistance of resistor ladders is expressed as continued fractions, too. We have designed resistor ladder networks whose overall equivalent resistance values are irrational number approximation ratios to a certain resistance value. Our Circuit simulation has verified this method.

１．はじめに

連分数は分数の分母にさらに分数が連なったものであり、桁数の多い少数や無理数の近似値を連分数としてあらわすことができる。零約術は関孝和が考案し、建部賢弘とその兄の建部賢明が発展させたといわれている。賢明が考案した零約術は、長い小数を分数に近似する方法で、互除法を駆使するが、その結果は、今日では連分数展開とも呼ばれている[1]。

本稿は、この連分数展開を用いて、電子回路中で、整数比の抵抗値を持つ抵抗から無理数近似値の比を持つ等価的な抵抗を作り出す方法を調査し回路シミュレーション検証した結果を報告するものである。

２．連分数展開を用いた抵抗ラダーの構成

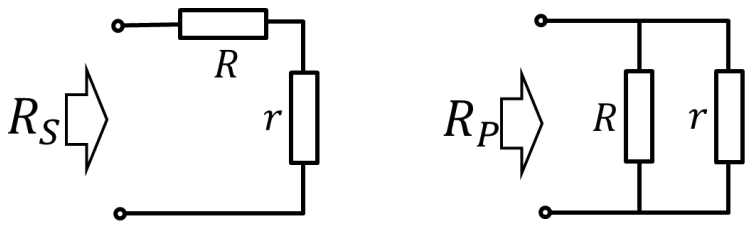
電子回路中の素子の一つである抵抗は、ある電圧を印加したときにその電圧の大きさに比例した電流が流れる、という性質を持つ。この時の電流 I (単位は A、アンペア)と電圧 V(単位は V、ボルト)の比を、電気抵抗R(単位は Ω、オーム)といい、$R = V/I$ である。

いま、二種類の抵抗Rとrがあるとする。この二種類の抵抗を図 1(a)と図 1(b)のようにそれぞれ直列接続、並列接続すると、これらは等価的に

$$R_S = R + r \tag{1}$$

$$\frac{1}{R_P} = \frac{1}{R} + \frac{1}{r} \leftrightarrow R_P = \frac{R \cdot r}{R + r} \tag{2}$$

という抵抗になる。つまり、二つの抵抗を直列に接続した場合の合成抵抗は「それぞれの抵抗の和」になり、並列に接続した場合の合成抵抗の逆数は「それぞれ抵抗の逆数の和」になる。

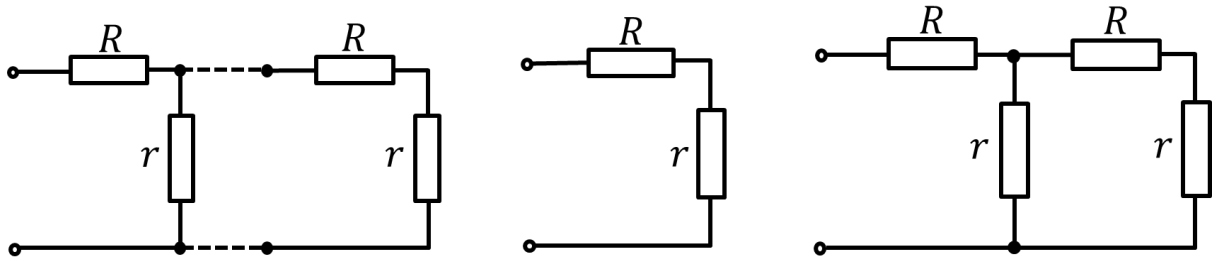


(a) 直列接続のとき　　(b) 並列接続のとき

図 1 抵抗の接続

この二種類の抵抗Rとrで、図 2(a)に示すような“抵抗ラダー”、つまり抵抗のはしごを構成した場合の

合成抵抗を考え、これを R-r ラダーと呼ぶことにする。ラダーが一段の場合の合成抵抗Z_1は、上記の直列接続の場合の合成抵抗から、$Z_1 = R + r$である。ラダーが二段の場合の合成抵抗Z_2は、$Z_2 = R + \frac{r(R+r)}{r+(R+r)}$で



(a) R-rラダー　(b) 抵抗ラダーの単位　(c) 2段のR-rラダー

図 2 R-r 抵抗ラダー

ある。

この作業を繰り返すと、k段接続した場合の合成抵抗Z_kと$k+1$段接続した場合の合成抵抗Z_{k+1}の関係から、次のような漸化式がたつ。

$$Z_{k+1} = R + \frac{rZ_k}{r + Z_k} = \frac{(r+R)Z_k + rR}{Z_k + r} \tag{3}$$

この漸化式から、図のようなRとrからなるk段の抵抗ラダーの合成抵抗は、以下の式で表される。

$$\begin{aligned} Z_k &= \frac{\alpha\gamma^k - \beta}{\gamma^n - 1} \\ \alpha &= \frac{1}{2}\left(R + \sqrt{R^2 + 4rR}\right) \\ \beta &= \frac{1}{2}\left(R - \sqrt{R^2 + 4rR}\right) \\ \gamma &= \frac{R + r - \beta}{R + r - \alpha} \end{aligned} \tag{4}$$

(4)式において、$1 < \gamma$であるため、接続する段数を増やしkが大きくなると、この抵抗値は次式の値Z_∞に収束する。

$$Z_\infty = \lim_{k\to\infty} Z_k \,\&= \lim_{k\to\infty} \frac{\alpha - \beta\gamma^{-k}}{1 - \gamma^{-k}} \to \alpha = \frac{1}{2}\left(R + \sqrt{R^2 + 4rR}\right) \tag{5}$$

また，式(3)から$k+1$段の R-r ラダーの合成抵抗は，次式のように連分数で表示することができる。mは整数である。

$$Z_{k+1} = \frac{R}{m}\left(k + \frac{1}{\frac{R}{mr} + \frac{R}{mZ_n}}\right) = \frac{R}{m}\left(m + \cfrac{1}{\frac{R}{mr} + \cfrac{1}{m + \cfrac{1}{\frac{R}{mr} + \frac{1}{\ddots}}}}\right) \tag{6}$$

次に図 3 のように、各段の抵抗の値が異なる抵抗ラダーを考える。抵抗ラダーのk段目の抵抗値は，ある抵抗値Rに対して，p_kRとq_kRとして，p_kとq_kによって重みづけられているとする。

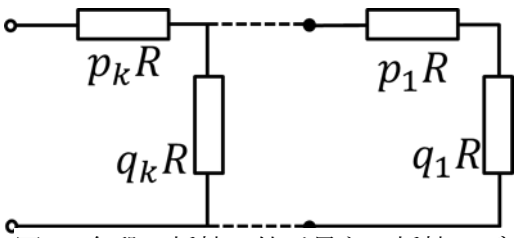


図 3 各段の抵抗の値が異なる抵抗ラダー

この時のk段抵抗ラダーの合成抵抗は、次式であらわされ、p_kとq_kを用いた連分数としてもあらわすことができる。

$$Z_k = p_k R + \frac{1}{\frac{1}{q_k R} + \frac{1}{Z_{k-1}}} = R\left(p_k + \frac{1}{\frac{1}{q_k} + \frac{R}{Z_{k-1}}} \right) = R\left(p_k + \frac{1}{\frac{1}{q_k} + \frac{1}{p_{k-1} + \frac{1}{\frac{1}{q_{k-1}} + \frac{1}{\ddots}}}} \right) \tag{7}$$

この(7)式から、任意の数について、その連分数表示に従ってk段目の抵抗の重みp_kとq_kを決定し、抵抗ラダーを構成することで、Rに対してその数の比を持つ抵抗を作ることができる.

３．抵抗ラダーを用いた無理数（近似）信号の出力

連分数としてあらわされる無理数の例として、貴金属数と呼ばれる数字がある。貴金属数は二次方程式$x^2 - nx - 1 = 0$の正の解であり、正の整数nの値に対して順に、第n貴金属数と呼ばれる[2]. 二次方程式$x^2 - nx - 1 = 0$の正の解λ_nとその連分数表示は次式のようにあらわすことができる。

$$\lambda_n = \frac{n + \sqrt{n^2 + 4}}{2} = n + \frac{1}{n + \frac{1}{n + \frac{1}{n + \frac{1}{\ddots}}}} \tag{8}$$

しばしば、$n = 1$の場合を黄金数ϕ、$n = 2$の場合を白銀数τ、$n = 3$の場合を青銅数ξと呼ぶ[5]。ϕ、τ、ξそれぞれの数値は、

$$\begin{aligned} \phi &= \frac{1 + \sqrt{5}}{2} \approx 1.618 \\ \tau &= 1 + \sqrt{2} \approx 2.414 \\ \xi &= \frac{3 + \sqrt{13}}{2} \approx 3.303 \end{aligned} \tag{9}$$

である。

貴金属数が(8)式のように連分数展開できることから、(6)式においてmと$R/(mr)$を正の整数nとすることで、ラダーの合成抵抗がRに対して貴金属数の比になることが予想された。

例として、R-R ラダーの合成抵抗の収束する値$Z_{R,R}$は、次式のようにRに対して黄金数ϕの比になる。

$$Z_{R,R} = \frac{1 + \sqrt{5}}{2} R \tag{10}$$

また，2R-0.5R ラダーの合成抵抗の収束する値$Z_{2R,0.5R}$は、次式のようにRに対して白銀数$\tau = 1 + \sqrt{2}$の比になる。

$$Z_{2R,0.5R} = \left(1 + \sqrt{2}\right) R = R \cdot \left(2 + \frac{1}{2 + \frac{1}{2 + \frac{1}{2 + \frac{1}{\ddots}}}} \right) \tag{11}$$

$\sqrt{2}$は連分数として次式のようにもあらわすことができる。

$$\sqrt{2} = 1 + \cfrac{1}{2 + \cfrac{1}{2 + \cfrac{1}{2 + \cfrac{1}{\ddots}}}} \tag{12}$$

これは、白銀数τから 1 を減じたものが$\sqrt{2}$であることによるものである。

このことを用いて、2R-0.5R ラダーの先頭の抵抗を 2R から R に変えた場合を考える．この抵抗ラダーの合成抵抗の収束する値は次式であらわされる．

$$Z_{\sqrt{2}} = R + \cfrac{1}{2 + \cfrac{1}{2 + \cfrac{1}{2 + \cfrac{1}{\ddots}}}} \cdot R = \sqrt{2}R \tag{13}$$

ネイピア数eは、自然対数の対数の底であり、無理数である。ネイピア数eは連分数として次式のようにあらわすことができる[3]。

$$\mathrm{e} = \ 2 + \cfrac{1}{1 + \cfrac{1}{2 + \cfrac{1}{1 + \cfrac{1}{\ddots}}}} \tag{14}$$

(14)式では省略したが、この連分数展開で分数部分の分子をすべて 1 とした場合の整数部分を並べて表示すると、以下のように規則性を持つ[3]。

$$\mathrm{e} = 2.71828\cdots = [2, 1, 2, 1, 1, 4, 1, 1, 6, 1, 1, 8, 1, 1, 10, \cdots] \tag{15}$$

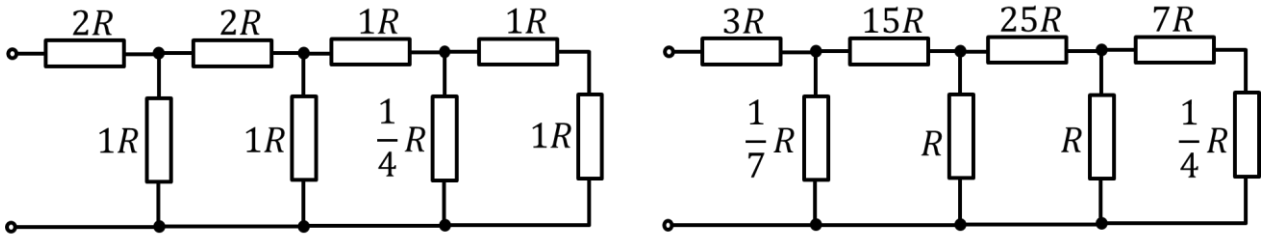
この連分数展開を用いて、図 3 に示した各段の抵抗値が異なる抵抗ラダーにおいて、(7)式のp_kを連分数展開整数部分の奇数番目、q_kを偶数番目の逆数とすることで、抵抗ラダーの合成抵抗のRに対する比は近似的にネイピア数eになる。

円周率πは、円の円周と直径の比であり、無理数である。円周率πを連分数展開すると、その整数部分は規則性を持たない[3]。円周率πを次式のように小数第五位までで近似すると、連分数として次式のようにあらわされる[3]。

$$\pi \approx 3.14159 = [3, 7, 15, 1, 25, 1, 7, 4] \tag{16}$$

これを用いて、ネイピア数の場合と同様に抵抗を接続して抵抗ラダーの合成抵抗のRに対する比は、近似的に円周率πになる。

4 段のネイピア数近似ラダーと円周率近似ラダーの構成を、図 4(a)と図 4(b)にそれぞれ示す。



(a) ネイピア数e近似ラダー(4段)　　(b) 円周率π近似ラダー

図 4 ネイピア数ラダーと円周率ラダー

以上のようにして抵抗ラダーを構成し、その合成抵抗が設計した通りの値を持つことが回路シミュレータ LTspice を用いたシミュレーションからも確認できた。

謝辞：　有意義なコメントをいただきました田部井勝稲先生に感謝します。

参考資料

[1] 小川束, 佐藤健一, 竹之内, 森本光生,「建部賢弘の数学」共立出版（2008）

[2] 岩本誠一、江口将生、吉良知文、「黄金・白銀・青銅：数と比と形と率と」(2008)
https://catalog.lib.kyushu-u.ac.jp/opac_download_md/15758/KJ00005471244.pdf

[3] 芹沢正三、数論入門、講談社ブルーバックス　(2008)

ECT-20-074

抵抗ラダー型デジタルアナログ変換器の微分非直線性の解析

平井 愛統, 谷本 洋, 源代 裕治
山本 修平, 桑名杏奈, 小林 春夫
群馬大学
北見工業大学

アウトライン

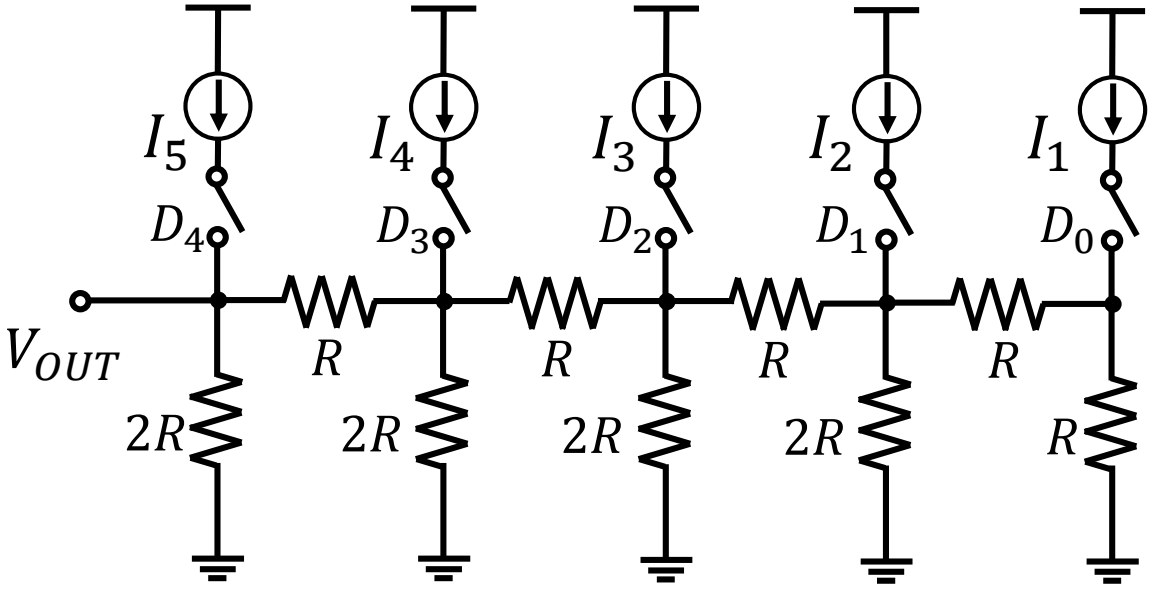
- 背景と目的
- N進抵抗ラダーDAC
  - 構成と例
- 素子ばらつきによるDNL劣化の解析
  - 数式による出力電圧誤差の見積もり
  - N進DACへの適用
- シミュレーションによる検討
- まとめ

アウトライン

- 背景と目的
- N進抵抗ラダーDAC
  - 構成と例
- 素子ばらつきによるDNL劣化の解析
  - 数式による出力電圧誤差の見積もり
  - N進DACへの適用
- シミュレーションによる検討
- まとめ

背景

- 電流モード R-2R DAC
  - 電流源による比較的高速な動作
  - R-2R 抵抗ラダー ・・・ デコーダが不要



5-bit R-2R current-steering DAC

背景・目的

- 電流モード R-2R DACの問題点
  - 分解能の増加 ⇒ 線形性の劣化
  - 原因

    回路中の電流源と抵抗のミスマッチ
- これまでの検討内容
  - 「*N*進抵抗ラダー」を用いたDAC構成
    ⇒ 電流を非2進に分流する抵抗ラダーを用いる

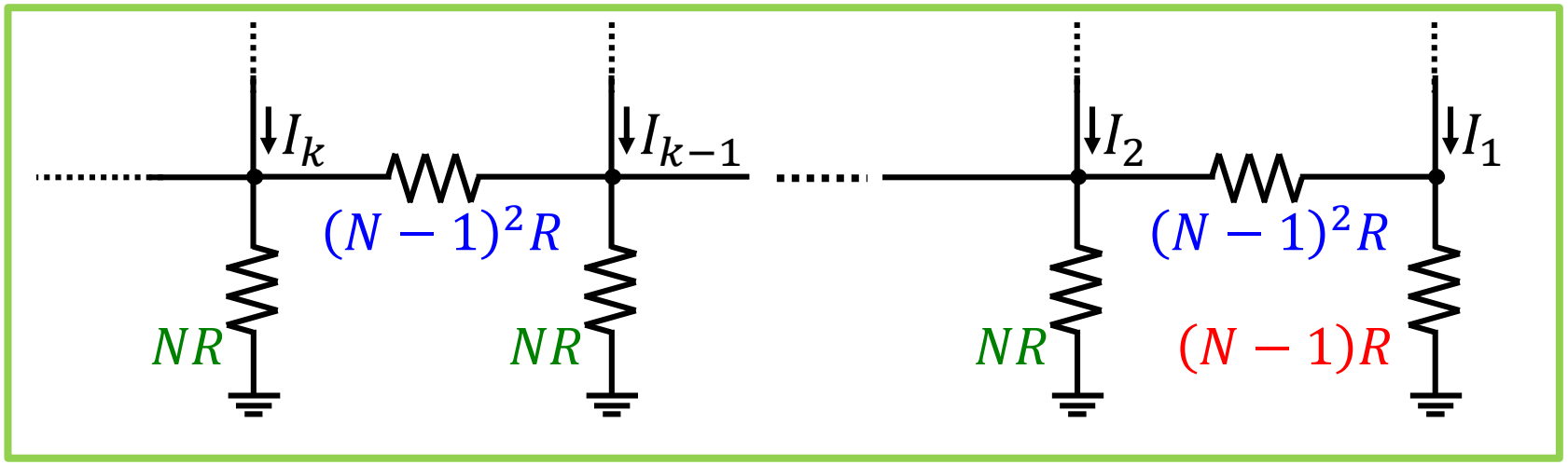
- 目的
  1. N進抵抗ラダーDACにおける線形性劣化傾向を明らかにし、
  2. 自己校正や量産テストの手法開発に役立てる

アウトライン

- 背景と目的
- N進抵抗ラダーDAC
  - 構成と例
- 素子ばらつきによるDNL劣化の解析
  - 数式による出力電圧誤差の見積もり
  - N進DACへの適用
- シミュレーションによる検討
- まとめ

N進抵抗ラダー

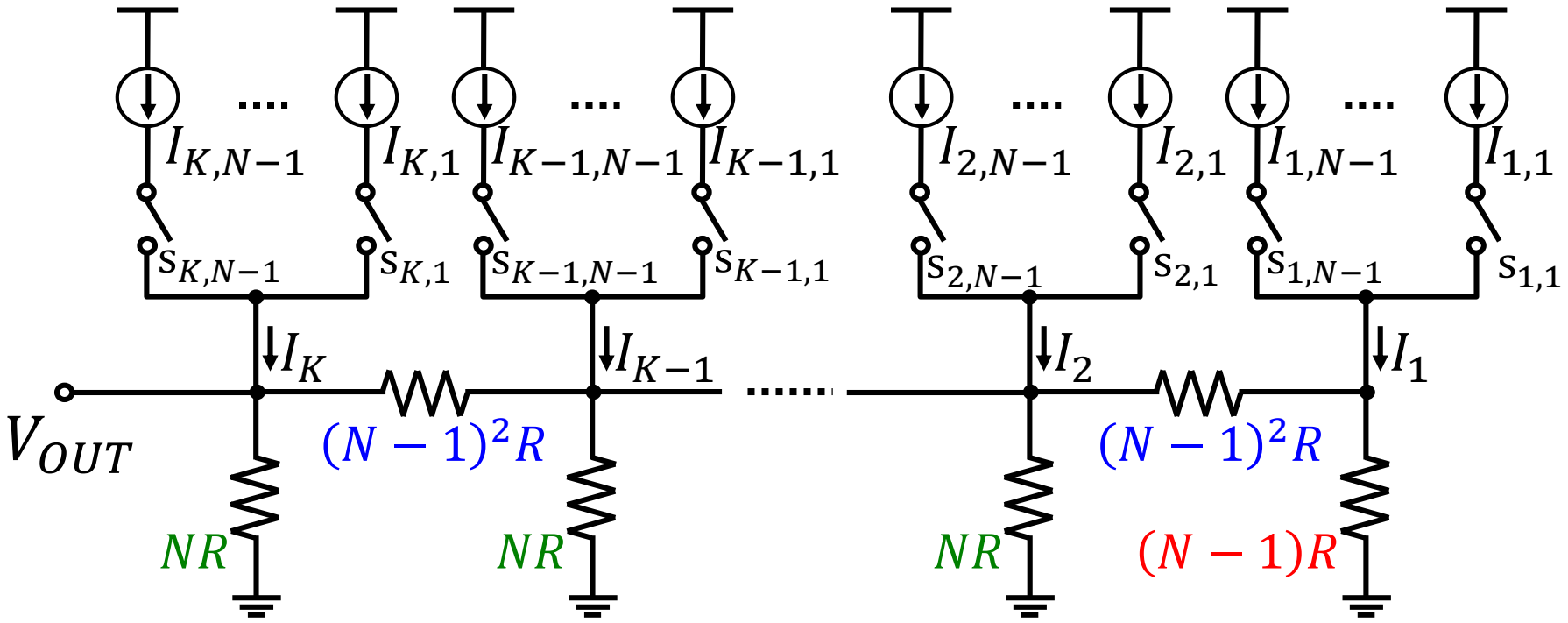
- *N*進抵抗ラダー



  - 電流をR-2Rラダーとは異なる比に分流
  - ラダーの抵抗比は N : $(N-1)^2$
    終端の抵抗 $(N-1)R$
  - N = 2の場合、R-2R ラダー

(4) M. Hirai, S. Yamamoto, H. Arai, A. Kuwana, H. Tanimoto, Y. Gendai, H. Kobayashi, "Systematic Construction of Resistor Ladder Network for N-ary DACs", IEEE ASICON (Oct. 2019)

N進抵抗ラダーDACの構成



N：電流分割比　　K：ラダー段数

I_j：j番目ノードに流し込まれる電流

R：単位抵抗　　I：単位電流

- N = 2の場合 ⇒ K-bit R-2R DAC

出力電圧と出力ステップ数

- 出力電圧

$$V_{\mathrm{OUT}}(I_1,\cdots,I_K,R,N,K)=(N-1)R\sum_{j=1}^{K}\left(\frac{I_j}{N^{K-j}}\right)$$

- 出力電圧最大値

$$V_{\mathrm{MAX}}(I,R,N,K)=RI\cdot N(N-1)\cdot\left(1-\frac{1}{N^K}\right)$$

- 出力電圧最小ステップ

$$V_{\mathrm{MIN}}(I,R,N,K)=(N-1)RI\cdot\frac{1}{N^{K-1}}$$

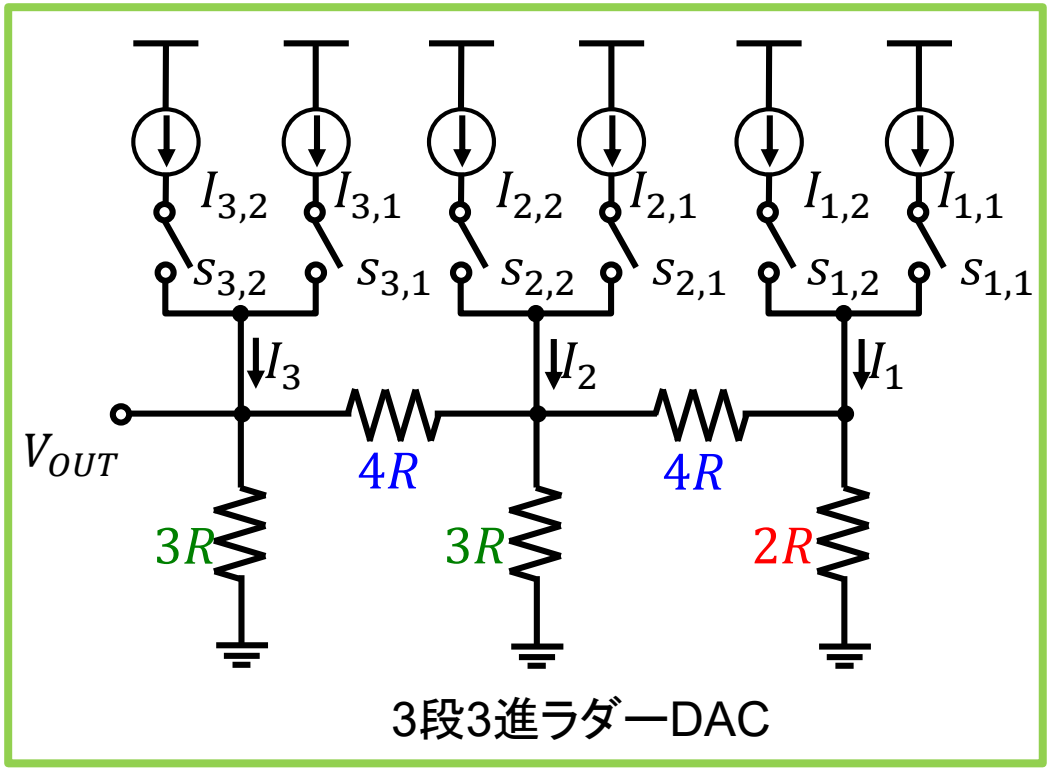
- 出力電圧数　N^K-1

N：電流分割比　　K：ラダー段数
I_j：j番目ノードに流し込まれる電流
R：単位抵抗　　I：単位電流

構成例 $N = 3$, 3進ラダーDAC

- ラダー抵抗比
  $4R : 3R : 2R$

- 出力電圧ステップ数
  $N^K - 1$
  $= 3^3 - 1 = 26$ 段階



3段3進ラダーDAC

- 出力電圧

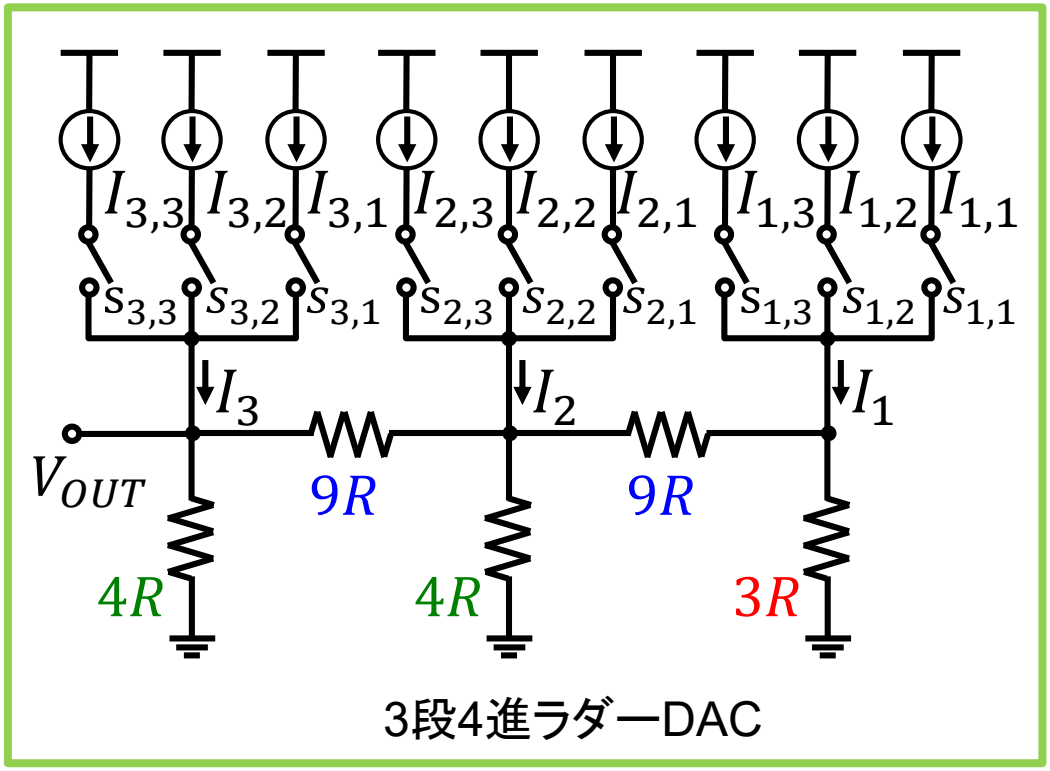
$$V_{OUT}(I_1, I_2, I_3, R) = 2R\left(I_3 + \frac{1}{3^1} I_2 + \frac{1}{3^2} I_1\right)$$

各段のI_j → 出力に対して3倍ずつの重みをもつ

構成例 $N = 4$, 4進ラダーDAC

- ラダー抵抗比
  $9R : 4R : 3R$

- 出力電圧ステップ数
  $N^K - 1$
  $= 4^3 - 1 = 63$ 段階



3段4進ラダーDAC

- 出力電圧

$$V_{OUT}(I_1, I_2, I_3, R) = \underbrace{3R\left(I_3 + \frac{1}{4^1} I_2 + \frac{1}{4^2} I_1\right)}$$

各段のI_j →出力に対して4倍ずつの重みをもつ

アウトライン

- 背景と目的
- N進抵抗ラダーDAC
  - 構成と例
- 素子ばらつきによるDNL劣化の解析
  - 数式による出力電圧誤差の見積もり
  - N進DACへの適用
- シミュレーションによる検討
- まとめ

R-2R DACのDNL解析　先行研究

[先行研究]電流モード R-2R DACのDNL解析

1. 単位抵抗と電流源にばらつきを仮定
2. 出力電圧の誤差を、抵抗ばらつき起因と
   電流ばらつき起因に分割
   （両方のばらつきがかかわる項を無視して近似）
3. それぞれの素子誤差を考慮してjビット目$D_j = 1$の時の
   出力電圧を表し、誤差を含む各コード出力を得る
4. MSB切り替わり時の出力電圧誤差を求め、
   DNL≦0.5 LSBになるためのマッチングを求める
5. シミュレーション結果と比較

[3] C. Chen, N. Lu, “Nonlinearity analysis of R-2R Ladder-Based Current-Steering Digital to Analog Converter,” IEEE International Symposium on Circuits and Systems (May 2013)

R-2R DACのDNL解析　先行研究

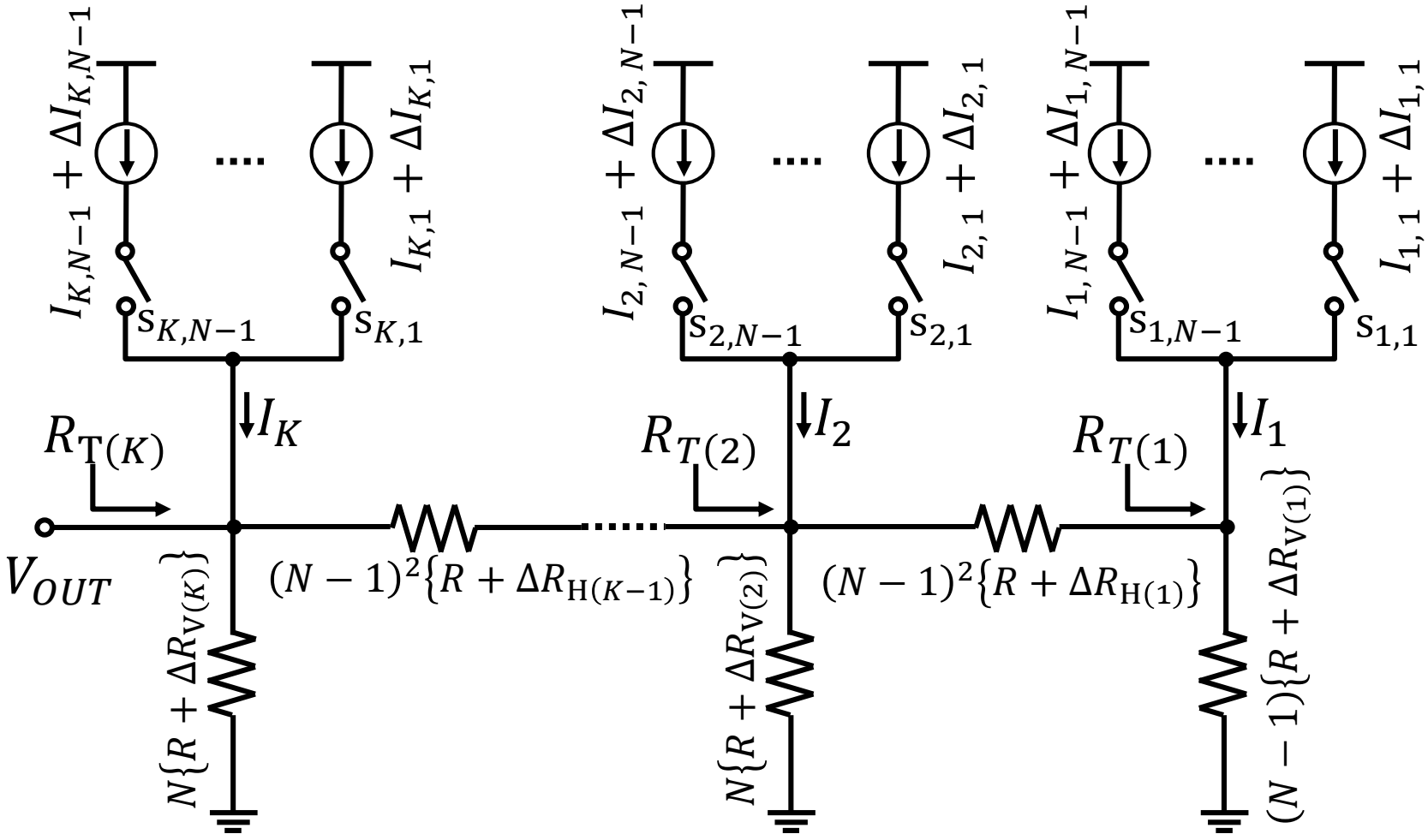
[先行研究]電流モード R-2R DACのDNL解析

1. 単位抵抗と電流源にばらつきを仮定
2. 出力電圧の誤差を、抵抗ばらつき起因と
   電流ばらつき起因に分割
   （両方のばらつきがかかわる項を無視して近似）
3. それぞれの素子誤差を考慮してjビット目$D_j = 1$の時の
   出力電圧を表し、誤差を含む各コード出力を得る
4. MSB切り替わり時の出力電圧誤差を求め、
   DNL≦0.5 LSBになるためのマッチングを求める
5. シミュレーション結果と比較

[3] C. Chen, N. Lu, “Nonlinearity analysis of R-2R Ladder-Based Current-Steering Digital to Analog Converter,” IEEE International Symposium on Circuits and Systems (May 2013)

素子ばらつきを仮定

- 単位抵抗と電流源にばらつきを仮定した回路図



出力電圧誤差の見積もり

- ある入力コードにおける出力電圧
  ⇒ $s_{j,i}$ 導通時の出力電圧を入力コードに応じて加算
- j番目ノードにのみ電流が流されている場合

$$V_{OUT}|_{s_{j,i}=1} = \left(I + \Delta I_{j,i}\right) \cdot \left\{\frac{R(N-1)}{N^{K-i}} + f\left(\Delta R_{s_{j,i}}\right)\right\}$$

$$\cong \frac{N-1}{N^{K-i}} RI + \frac{(N-1)R}{N^{K-i}} \cdot \Delta I_{j,i} + f(\Delta V_R)$$

1. 素子のばらつきがない場合の出力電圧
2. 電流のばらつきに起因する出力電圧誤差
3. 抵抗のばらつきに起因する出力電圧誤差

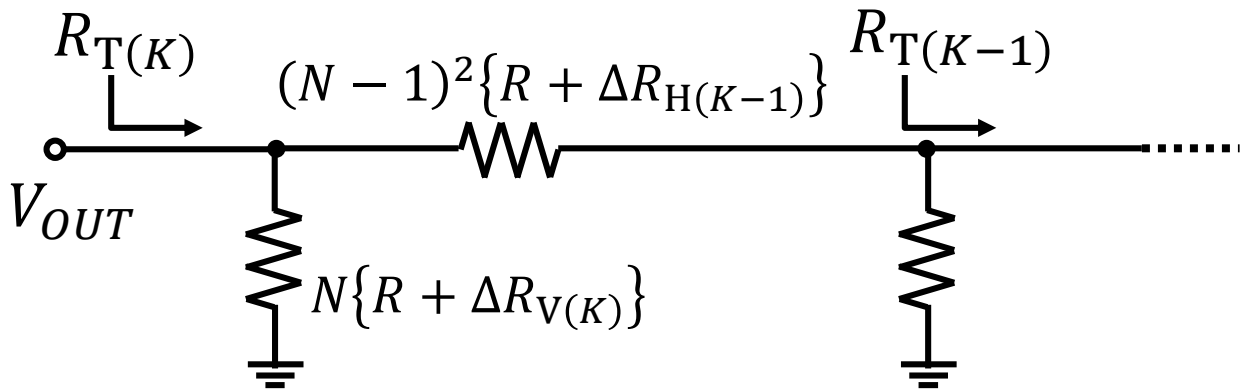
$\Delta I_{j,i}$はN^{K-i}で効く

この項を詳しく求める必要あり

※第4項 $\Delta I_{j,i} \cdot f\left(\Delta R_{s_{j,i}}\right)$ は
出力への寄与が小さいものとして無視

誤差を考慮した$R_{\mathrm{T}(K)}$

- $R_{\mathrm{T}(K)}$を N, R , $\Delta R_{\mathrm{V}(K)}$, $\Delta R_{\mathrm{H}(K-1)}$, $\Delta R_{\mathrm{T}(K-1)}$で表示



$$R_{\mathrm{T}(K)} = \{N(R+\Delta R_{\mathrm{V}(K)})\}//\{(N-1)^2(R+\Delta R_{\mathrm{H}(K-1)})+(N-1)R+\Delta R_{\mathrm{T}(K-1)}\}$$

$$= \frac{N(R+\Delta R_{\mathrm{V}(K)})\{(N-1)^2(R+\Delta R_{\mathrm{H}(K-1)})+(N-1)R+\Delta R_{\mathrm{T}(K-1)}\}}{N^2R+N\Delta R_{\mathrm{V}(K)}+(N-1)^2\Delta R_{\mathrm{H}(K-1)}+\Delta R_{\mathrm{T}(K-1)}}$$

誤差どうしの積を無視

$$\cong \frac{NR\{N(N-1)R+N(N-1)\Delta R_{\mathrm{V}(K)}+(N-1)^2\Delta R_{\mathrm{H}(K-1)}+\Delta R_{\mathrm{T}(K-1)}\}}{N^2R+N\Delta R_{\mathrm{V}(K)}+(N-1)^2\Delta R_{\mathrm{H}(K-1)}+\Delta R_{\mathrm{T}(K-1)}}$$

誤差を考慮した$R_{\mathrm{T}(K)}$

- $x \ll 1$の時のテイラー展開1次項までの近似式
  $1/(1+x) \cong 1-x$ を用いて近似

$$R_{\mathrm{T}(K)} = \frac{(N-1)R + (N-1)\Delta R_{\mathrm{V}(K)} + \dfrac{(N-1)^2}{N}\Delta R_{\mathrm{H}(K-1)} + \dfrac{\Delta R_{\mathrm{T}(K-1)}}{N}}{1 + \underline{\dfrac{N\Delta R_{\mathrm{V}(K)}}{N^2R} + \dfrac{(N-1)^2\Delta R_{\mathrm{H}(K-1)}}{N^2R} + \dfrac{\Delta R_{\mathrm{T}(K-1)}}{N^2R}}}$$

この部分をxとして、テイラー展開で近似

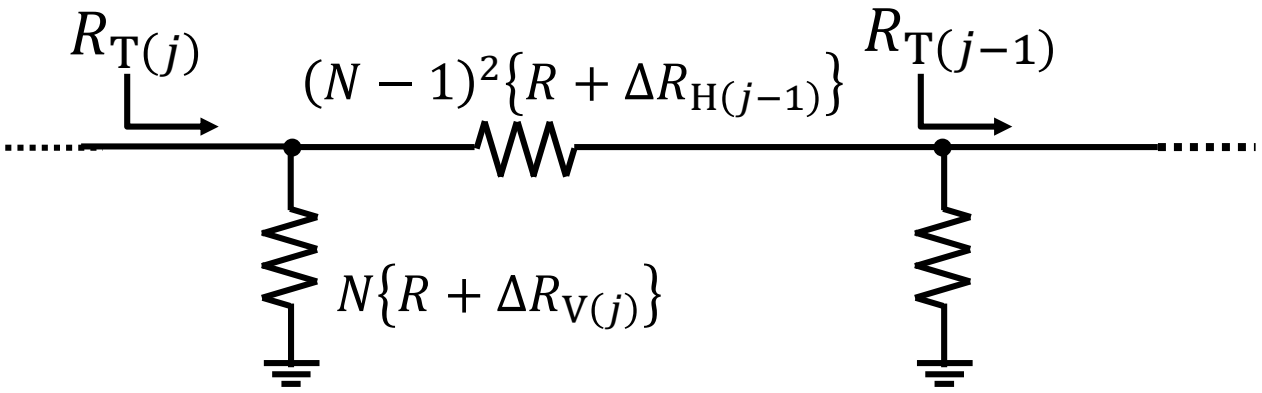
$$\cong \left\{(N-1)R + (N-1)\Delta R_{\mathrm{V}(K)} + \frac{(N-1)^2}{N}\Delta R_{\mathrm{H}(K-1)} + \frac{\Delta R_{\mathrm{T}(K-1)}}{N}\right\}$$

$$\times \left\{1 - \frac{\Delta R_{\mathrm{V}(K)}}{NR} - \frac{(N-1)^2\Delta R_{\mathrm{H}(K-1)}}{N^2R} - \frac{(N-1)\Delta R_{\mathrm{T}(K-1)}}{N^2R}\right\}$$

$$\cong (N-1)R + \frac{(N-1)^2}{N}\cdot\Delta R_{V(K)} + \frac{(N-1)^2}{N^2}\cdot\Delta R_{H(K-1)} + \frac{1}{N^2}\cdot\Delta R_{T(K-1)}$$

抵抗誤差を考慮した$V_{OUT}|_{s_{K,i}=1}$

- 同様の手順で、j番目のノードから見込んだ抵抗$R_{\mathrm{T}(j)}$を近似



$$R_{\mathrm{T}(j)} \cong (N-1)R + \frac{(N-1)^2}{N}\cdot\Delta R_{V(j)} + \frac{(N-1)^2}{N^2}\cdot\Delta R_{H(j-1)} + \frac{1}{N^2}\cdot\Delta R_{T(j-1)}$$

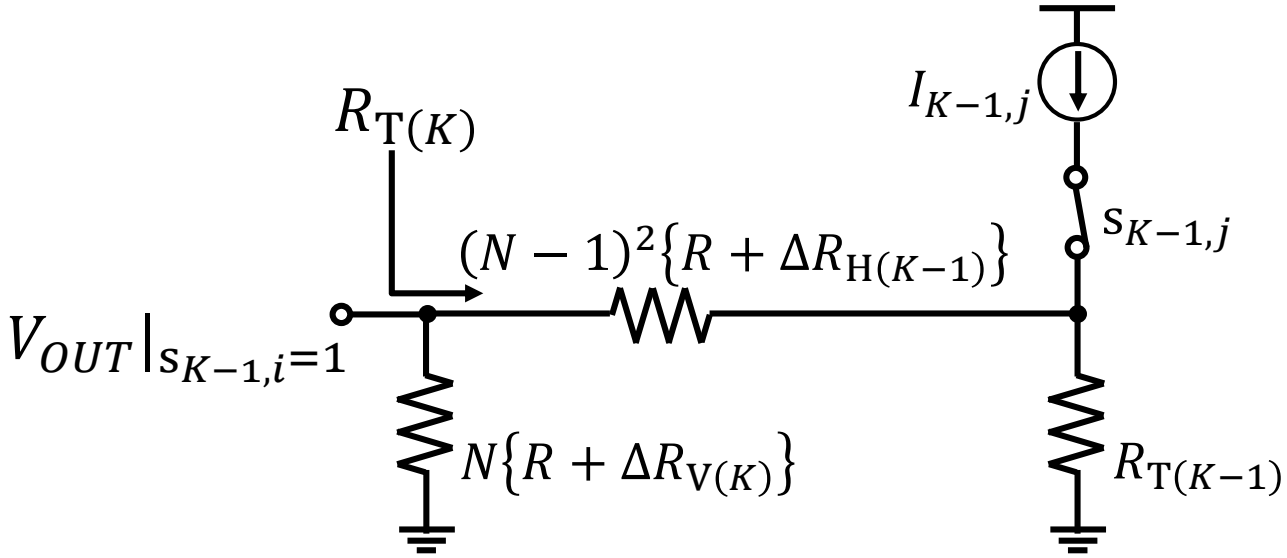
- K番目のノードに電流$I_{K,i}$を流し込んだ時の出力電圧

$$V_{OUT}|_{s_{K,i}=1} = I_{K,i}\cdot R_{\mathrm{T}(K)}$$

- $R_{\mathrm{T}(j)}$を用いて$R_{\mathrm{T}(K)}$を展開
  すべての$\Delta R_{\mathrm{V}(j)}$と$\Delta R_{\mathrm{H}(j)}$を用いて、誤差を含む出力を表せる

$s_{K-1,j} = 1$での出力電圧

- $K-1$番目ノードに電流を流し込んだ時の出力電圧



$$V_{OUT}|_{s_{K-1,i}=1}$$

$$= I_{K-1,i} \cdot \frac{R_{T(K-1)} \cdot R_{V(K)}}{R_{V(K)} + R_{H(K-1)} + R_{T(K-1)}}$$

$$= \frac{I_{K-1,i} \cdot \{(N-1)R + \Delta R_{T(K-1)}\} \cdot \{N(R + \Delta R_{V(K)})\}}{(N-1)R + \Delta R_{T(K-1)} + N(R + \Delta R_{V(K)}) + (N-1)^2(R + \Delta R_{H(K-1)})}$$

$V_{OUT}|_{s_{K-1,i}=1}$の近似

誤差どうしの積を無視

$$V_{OUT}|_{s_{K-1,i}=1} \cong \frac{I_{K-1,i}\{N(N-1)R + NR\Delta R_{\mathrm{T}(K-1)} + N(N-1)R\Delta R_{\mathrm{V}(K)}\}}{N^2R + \Delta R_{\mathrm{T}(K-1)} + N\Delta R_{\mathrm{V}(K)} + (N-1)^2\Delta R_{\mathrm{H}(K-1)}}$$

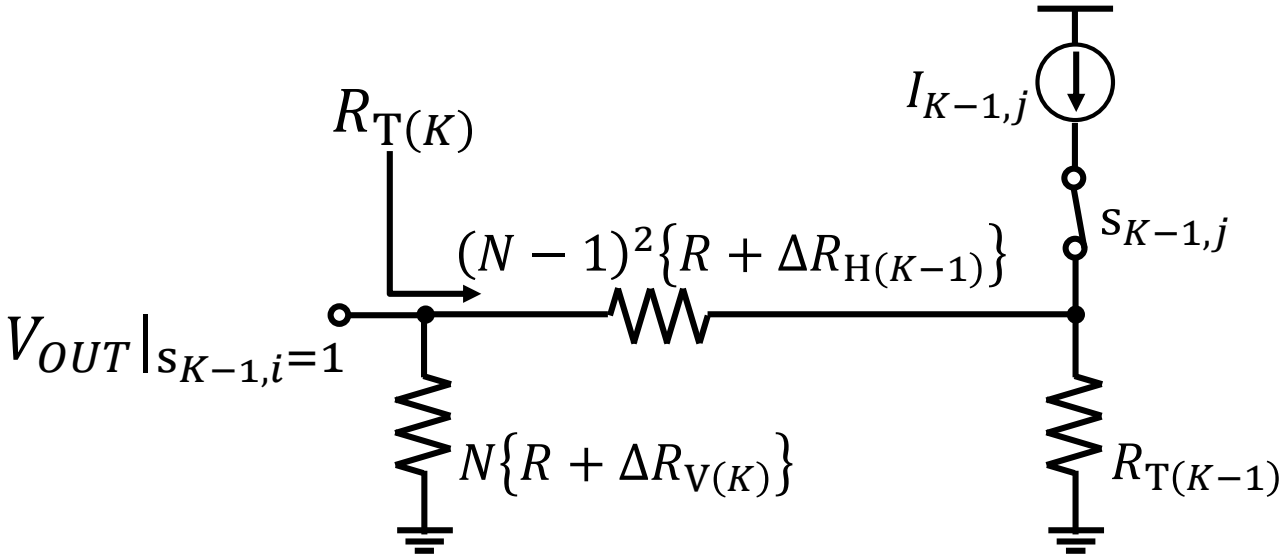
$$= \frac{I_{K-1,i}\left\{\frac{(N-1)R}{N} + \frac{\Delta R_{\mathrm{T}(K-1)}}{N} + \frac{(N-1)\Delta R_{\mathrm{V}(K)}}{N}\right\}}{1 + \frac{\Delta R_{\mathrm{T}(K-1)}}{N^2R} + \frac{N\Delta R_{\mathrm{V}(K)}}{N^2R} + \frac{(N-1)^2\Delta R_{\mathrm{H}(K-1)}}{N^2R}}$$

テイラー展開を用いて近似

$$\cong I_{K-1,i} \cdot \left\{\frac{(N-1)R}{N} + \frac{\Delta R_{\mathrm{T}(K-1)}}{N} + \frac{(N-1)\Delta R_{\mathrm{V}(K)}}{N}\right\}$$
$$\times \left\{1 - \frac{\Delta R_{\mathrm{T}(K-1)}}{N^2R} - \frac{N\Delta R_{\mathrm{V}(K)}}{N^2R} - \frac{(N-1)^2\Delta R_{\mathrm{H}(K-1)}}{N^2R}\right\}$$

抵抗誤差を含んだ$V_{OUT}|_{s_{K-1,i}=1}$

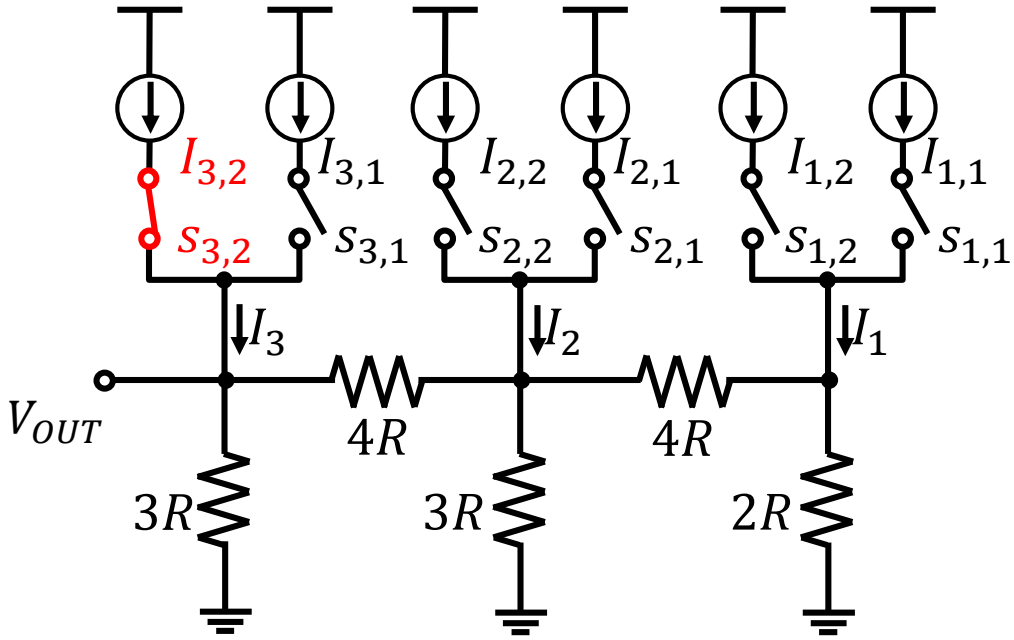
- 抵抗の誤差を考慮して近似した$V_{OUT}|_{s_{K-1,i}=1}$



$$V_{OUT}|_{s_{K-1,i}=1} \cong I_{K-1,j} \cdot \left\{ \frac{N-1}{N} R + \frac{(N-1)^2}{N^2} \Delta R_{V(K)} + \frac{N^2-N+1}{N^3} \Delta R_{T(K-1)} - \frac{(N-1)^3}{N^3} \Delta R_{H(K-1)} \right\}$$

ある$s_{j,i}$導通時の電圧 ⇒ コードごとの誤差を含む出力電圧

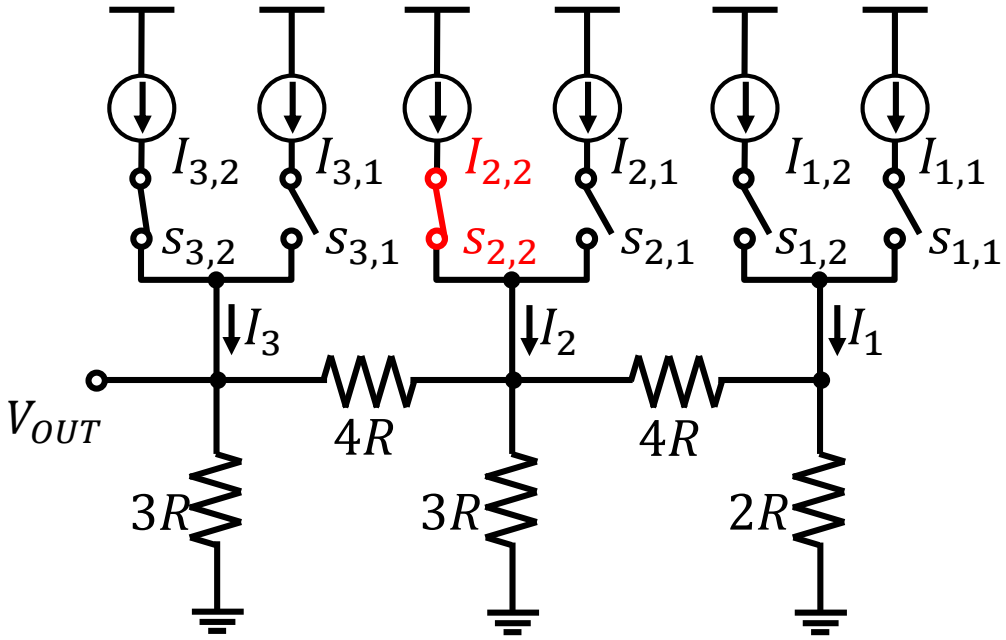
3段3進ラダーDAC, $s_{3,i} = 1$



$$V_{OUT}|_{s_{3,i}=1} \cong I_{3,i} \cdot \left\{ 2R + \frac{4\Delta R_{V(3)}}{3} + \frac{4\Delta R_{H(2)}}{9} + \frac{\Delta R_{T(2)}}{9} \right\}$$

$$\cong I_{3,i} \cdot \left\{ 2R + \frac{4\Delta R_{V(3)}}{3} + \frac{4\Delta R_{H(2)}}{9} + \frac{4\Delta R_{V(2)}}{27} + \frac{4\Delta R_{H(2)}}{81} + \frac{2\Delta R_{V(1)}}{81} \right\}$$

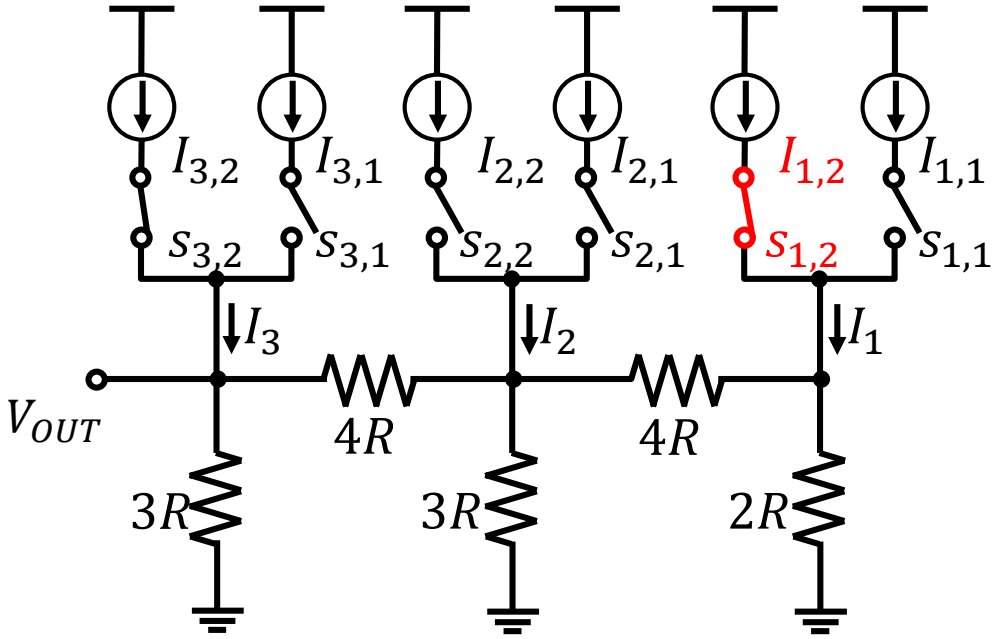
3段3進ラダーDAC, $s_{2,i} = 1$



$$V_{OUT}|_{s_{2,i}=1} \cong I_{2,i} \cdot \left\{ \frac{2R}{3} + \frac{4\Delta R_{V(3)}}{9} - \frac{8\Delta R_{H(2)}}{27} + \frac{7\Delta R_{T(2)}}{27} \right\}$$

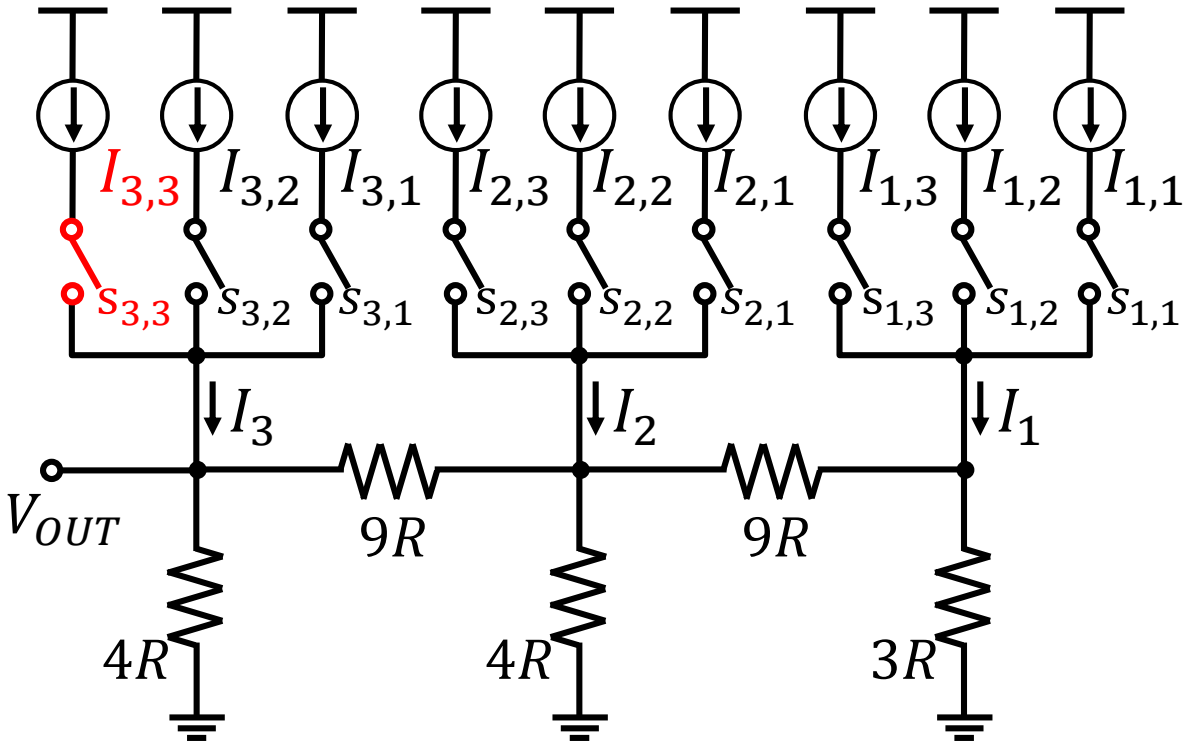
$$\cong I_{2,i} \cdot \left\{ \frac{2R}{3} + \frac{4\Delta R_{V(3)}}{9} - \frac{8\Delta R_{H(2)}}{27} + \frac{28\Delta R_{V(2)}}{81} + \frac{28\Delta R_{H(1)}}{243} + \frac{14\Delta R_{V(1)}}{243} \right\}$$

3段3進ラダーDAC, $s_{1,i} = 1$



$$V_{OUT}|_{s_{1,i}=1} \cong I_{1,j} \cdot \left\{ \frac{2R}{9} + \frac{4\Delta R_{V(3)}}{27} - \frac{8\Delta R_{H(2)}}{81} + \frac{28\Delta R_{V(2)}}{243} - \frac{80\Delta R_{H(1)}}{729} + \frac{122\Delta R_{V(1)}}{729} \right\}$$

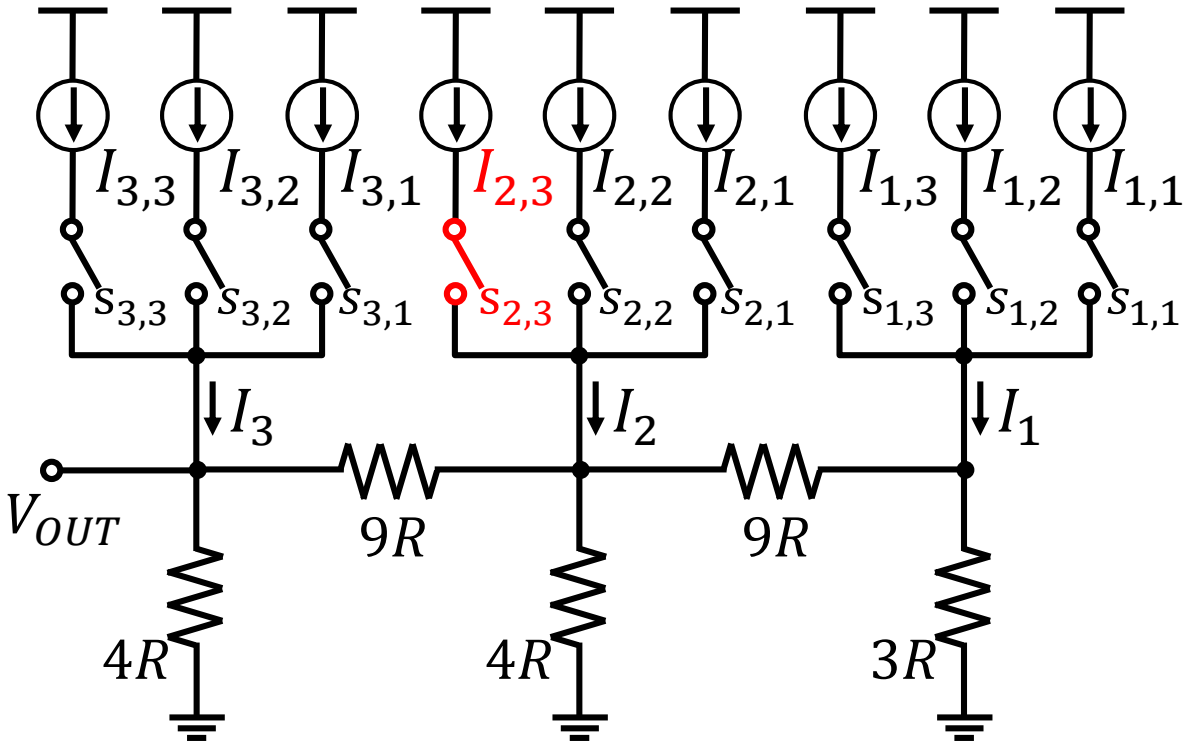
3段4進ラダーDAC, $s_{3,i} = 1$



$$V_{OUT}|_{s_{3,i}=1} \cong I_{3,i} \cdot \left\{ 3R + \frac{9\Delta R_{V(3)}}{4} + \frac{9\Delta R_{H(2)}}{16} + \frac{\Delta R_{T(2)}}{16} \right\}$$

$$\cong I_{3,i} \cdot \left\{ 3R + \frac{9\Delta R_{V(3)}}{4} + \frac{9\Delta R_{H(2)}}{16} + \frac{9\Delta R_{V(2)}}{64} + \frac{9\Delta R_{H(2)}}{256} + \frac{3\Delta R_{V(1)}}{256} \right\}$$

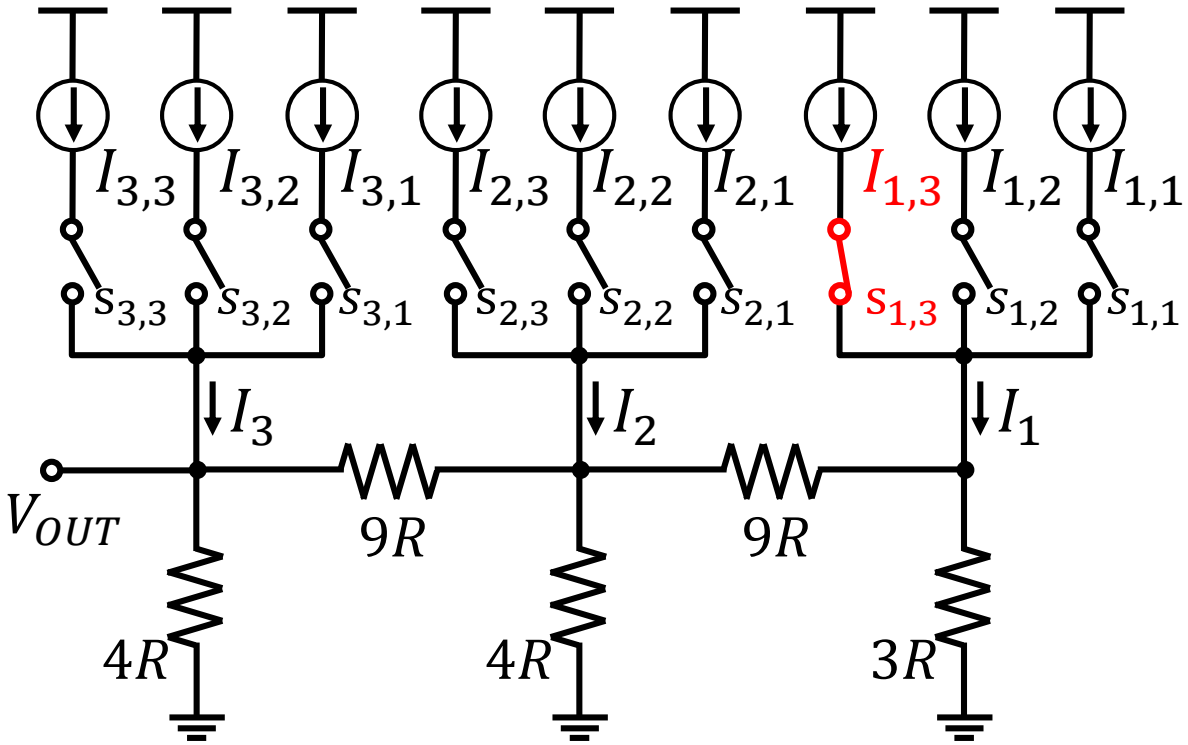
3段4進ラダーDAC, $s_{2,i} = 1$



$$V_{OUT}|_{s_{2,i}=1} \cong I_{2,i} \cdot \left(\frac{3R}{4} + \frac{3\Delta R_{\mathrm{V}(3)}}{9} - \frac{27\Delta R_{\mathrm{H}(2)}}{64} + \frac{13\Delta R_{\mathrm{T}(2)}}{64} \right)$$

$$\cong I_{2,i} \cdot \left\{ \frac{3R}{4} + \frac{3\Delta R_{\mathrm{V}(3)}}{9} - \frac{27\Delta R_{\mathrm{H}(2)}}{64} + \frac{17\Delta R_{V(2)}}{256} + \frac{117\Delta R_{H(1)}}{1024} + \frac{39\Delta R_{V(1)}}{1024} \right\}$$

3段4進ラダーDAC,$s_{1,i}=1$



$$V_{OUT}|_{s_{1,i}=1}$$

$$\cong I_{1,j}\cdot\left\{\frac{3R}{16}+\frac{9\Delta R_{V(3)}}{64}-\frac{27\Delta R_{H(2)}}{256}+\frac{117\Delta R_{V(2)}}{1024}-\frac{459\Delta R_{H(1)}}{4096}+\frac{615\Delta R_{V(1)}}{4096}\right\}$$

アウトライン

- 背景と目的
- N進抵抗ラダーDAC
  - 構成と例
- 素子ばらつきによるDNL劣化の解析
  - 数式による出力電圧誤差の見積もり
  - N進DACへの適用
- シミュレーションによる検討
- まとめ

DNLの計算

- DNLの定義

$$DNL(n) = \frac{V_{\mathrm{OUT}}(n) - V_{\mathrm{OUT}}(n-1)}{V_{\mathrm{LSB}}} - 1$$

V_{LSB}：最小の出力電圧の理想値

- DNL標準偏差 σ_{DNL}
  - スライド16における近似

$$V_{OUT}|_{s_{j.i}=1} \cong \frac{N-1}{N^{K-i}} RI + \frac{(N-1)R}{N^{K-i}} \cdot \Delta I_{j,i} + f(\Delta V_R)$$

  - 抵抗誤差起因のDNL標準偏差 σ_{DNL_R}
    電流誤差起因のDNL標準偏差 σ_{DNL_I}を用いて、

$$\sigma_{DNL}^2 = \sigma_{DNL\_R}^2 + \sigma_{DNL\_I}^2$$

3段3進ラダーDACのDNL

- 3段3進ラダーDACについて

$$DNL(9) = \frac{V_{OUT}|_{s_{3,1}=1} - \sum_{i=1}^{2}\left(V_{OUT}|_{s_{2,i}=1} + V_{OUT}|_{s_{1,i}=1}\right)}{V_{LSB}} - 1$$

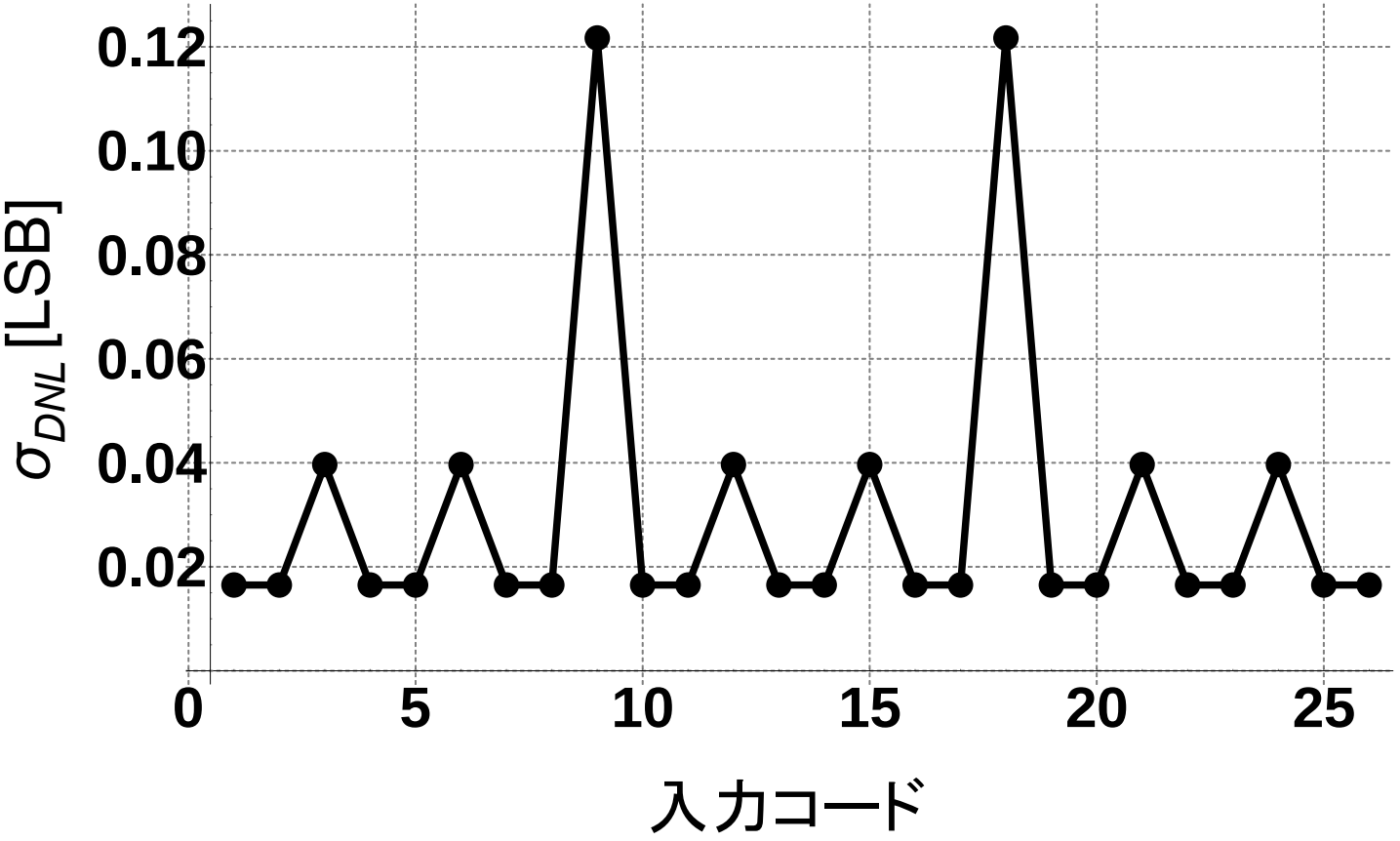
$$DNL(3) = \frac{V_{OUT}|_{s_{2,1}=1} - \left(V_{OUT}|_{s_{1,1}=1} + V_{OUT}|_{s_{1,2}=1}\right)}{V_{LSB}} - 1$$

$$DNL(1) = \frac{V_{OUT}|_{s_{1,1}=1} - 0}{V_{LSB}} - 1$$

- 単位抵抗と単位電流のばらつき
  - 正規分布
  - 標準偏差 1%

3段3進ラダーDAC　σ_{DNL}計算結果

- コード9 と 18でσ_{DNL}が最大
- 最大σ_{DNL}は2番目に大きいσ_{DNL}のおよそ3倍
- 最小σ_{DNL}は2入力について連続で現れる



3段4進ラダーDACのDNL

- 3段4進ラダーDACについて

$$DNL(16) = \frac{V_{OUT}|_{s_{3,1}=1} - \sum_{i=1}^{3}\left(V_{OUT}|_{s_{2,i}=1} + V_{OUT}|_{s_{1,i}=1}\right)}{V_{LSB}} - 1$$

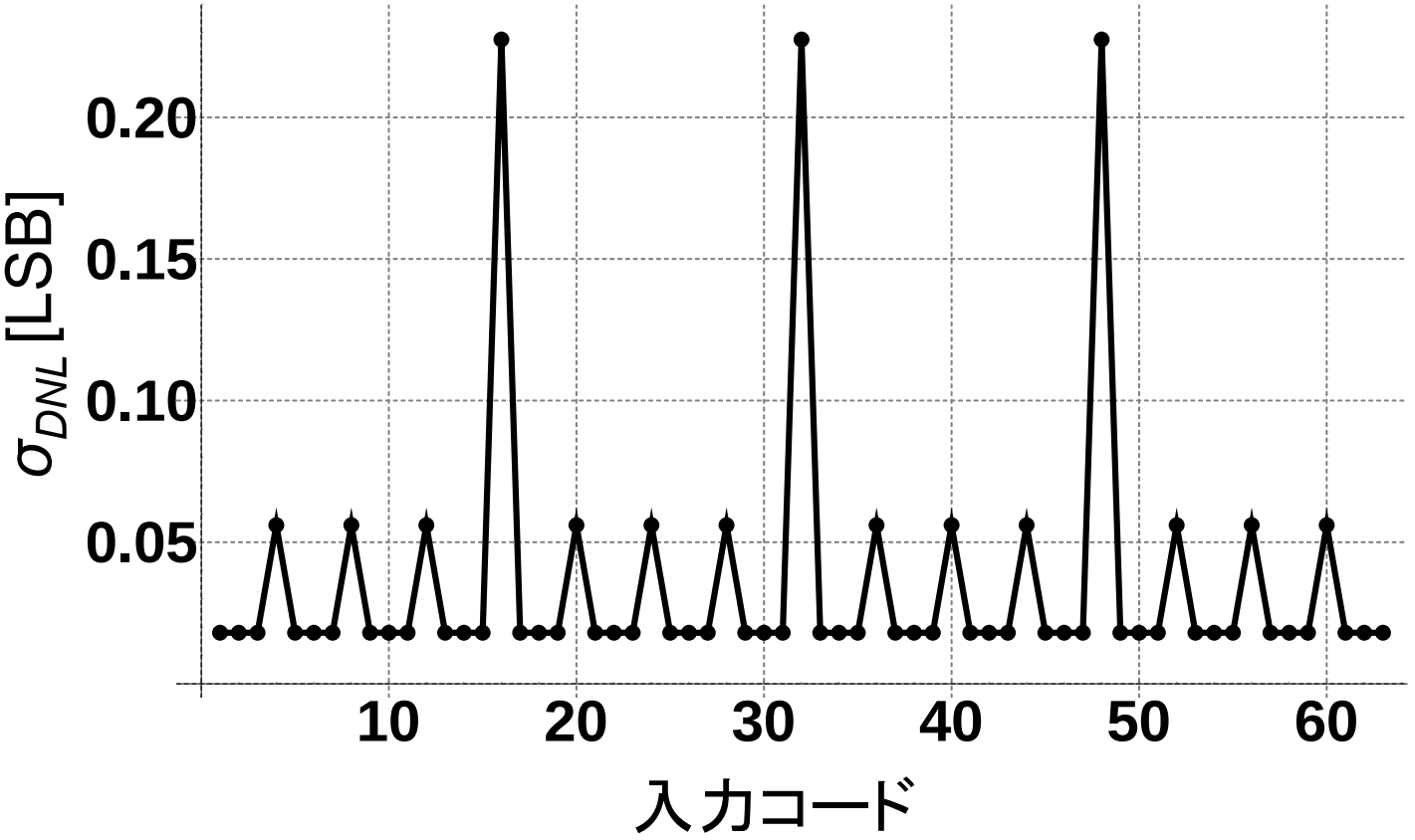
$$DNL(4) = \frac{V_{OUT}|_{s_{2,1}=1} - \sum_{i=1}^{3}\left(V_{OUT}|_{s_{1,i}=1}\right)}{V_{LSB}} - 1$$

$$DNL(1) = \frac{V_{OUT}|_{s_{1,1}=1} - 0}{V_{LSB}} - 1$$

- 単位抵抗と単位電流のばらつき
  - 正規分布
  - 標準偏差 1%

3段4進ラダーDAC　σ_{DNL}計算結果

- コード16, 32, 48でσ_{DNL}が最大
- 最大σ_{DNL}は2番目に大きいσ_{DNL}のおよそ4倍
- 最小σ_{DNL}は3入力について連続で現れる

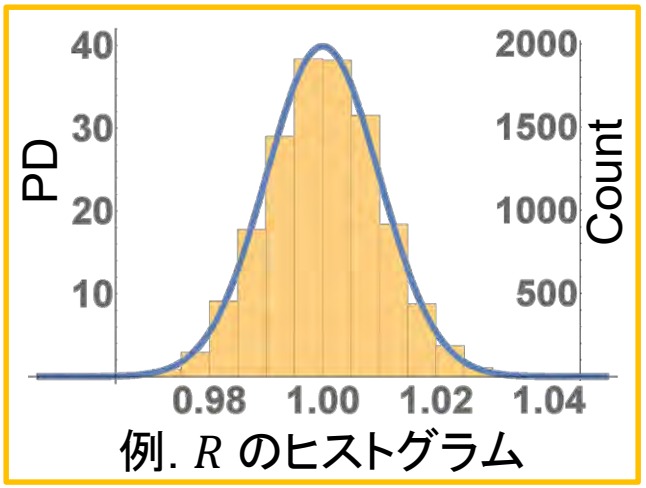


アウトライン

- 背景と目的
- N進抵抗ラダーDAC
  - 構成と例
- 素子ばらつきによるDNL劣化の解析
  - 数式による出力電圧誤差の見積もり
  - N進DACへの適用
- シミュレーションによる検討
- まとめ

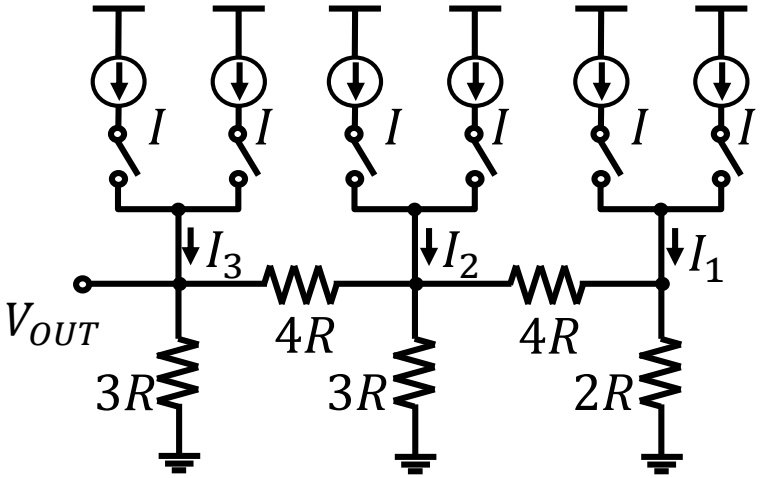
シミュレーション条件

- シミュレーションの条件
  - シミュレーションセット数　3000回
  - 単位抵抗Rと単位電流Iに正規分布のばらつきを仮定
  - 標準偏差σは平均値の1%

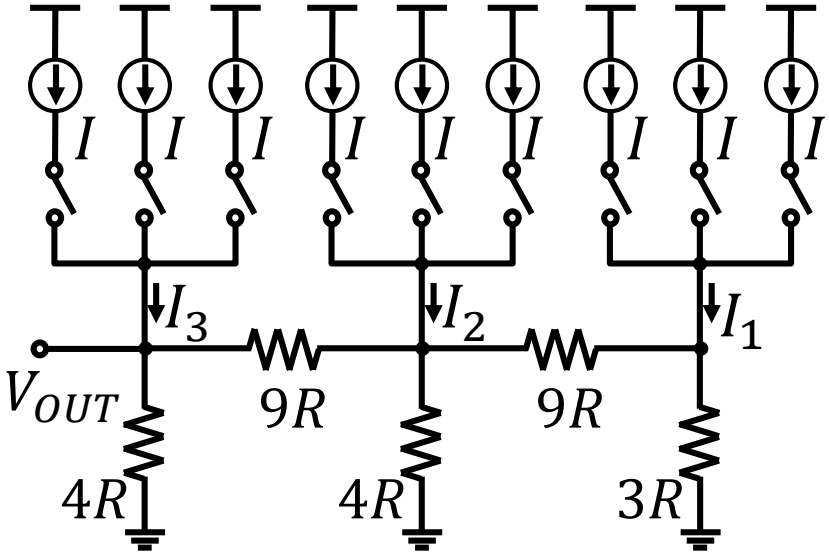


例. R のヒストグラム

- シミュレーションした回路



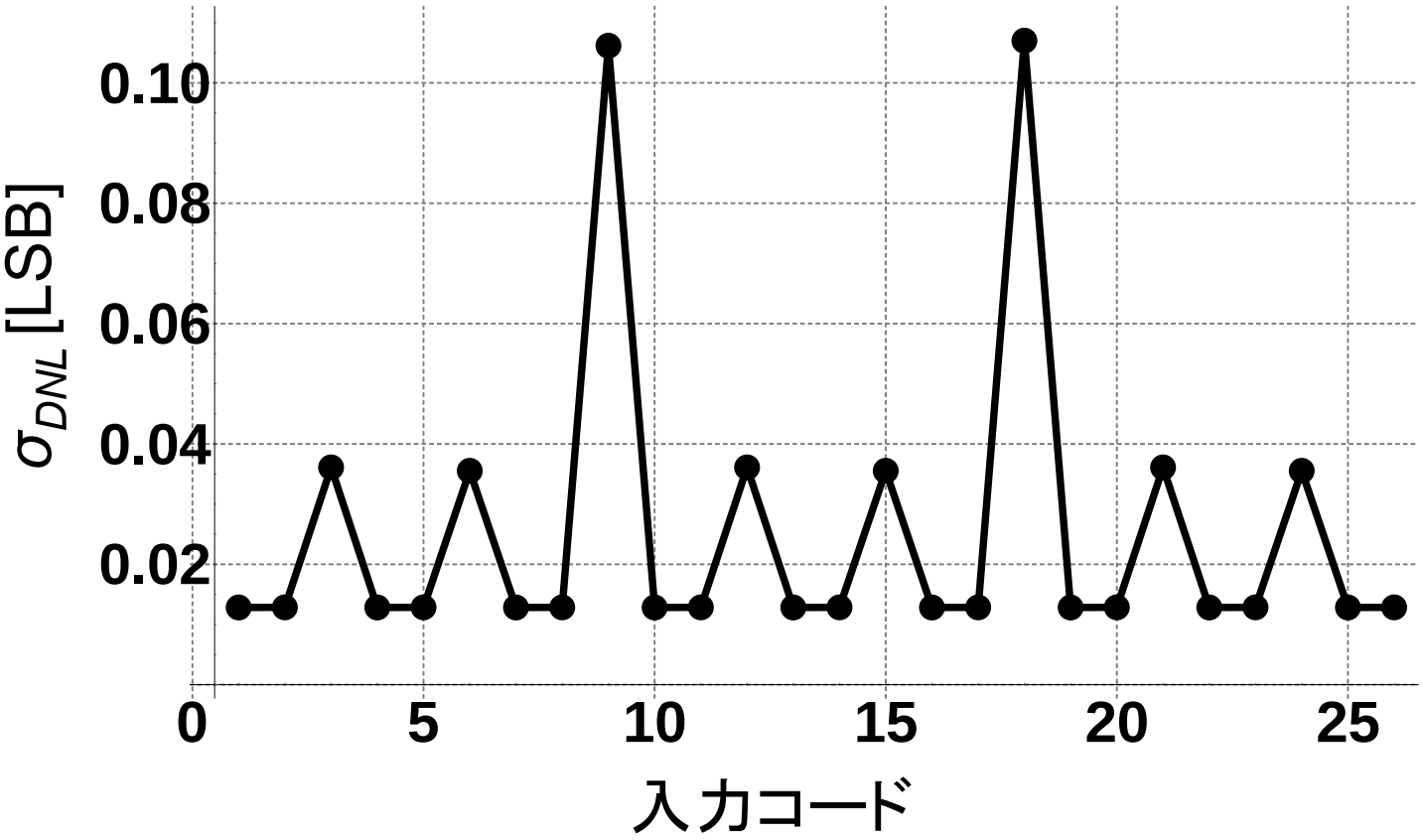
3段3進ラダーDAC



3段4進ラダーDAC

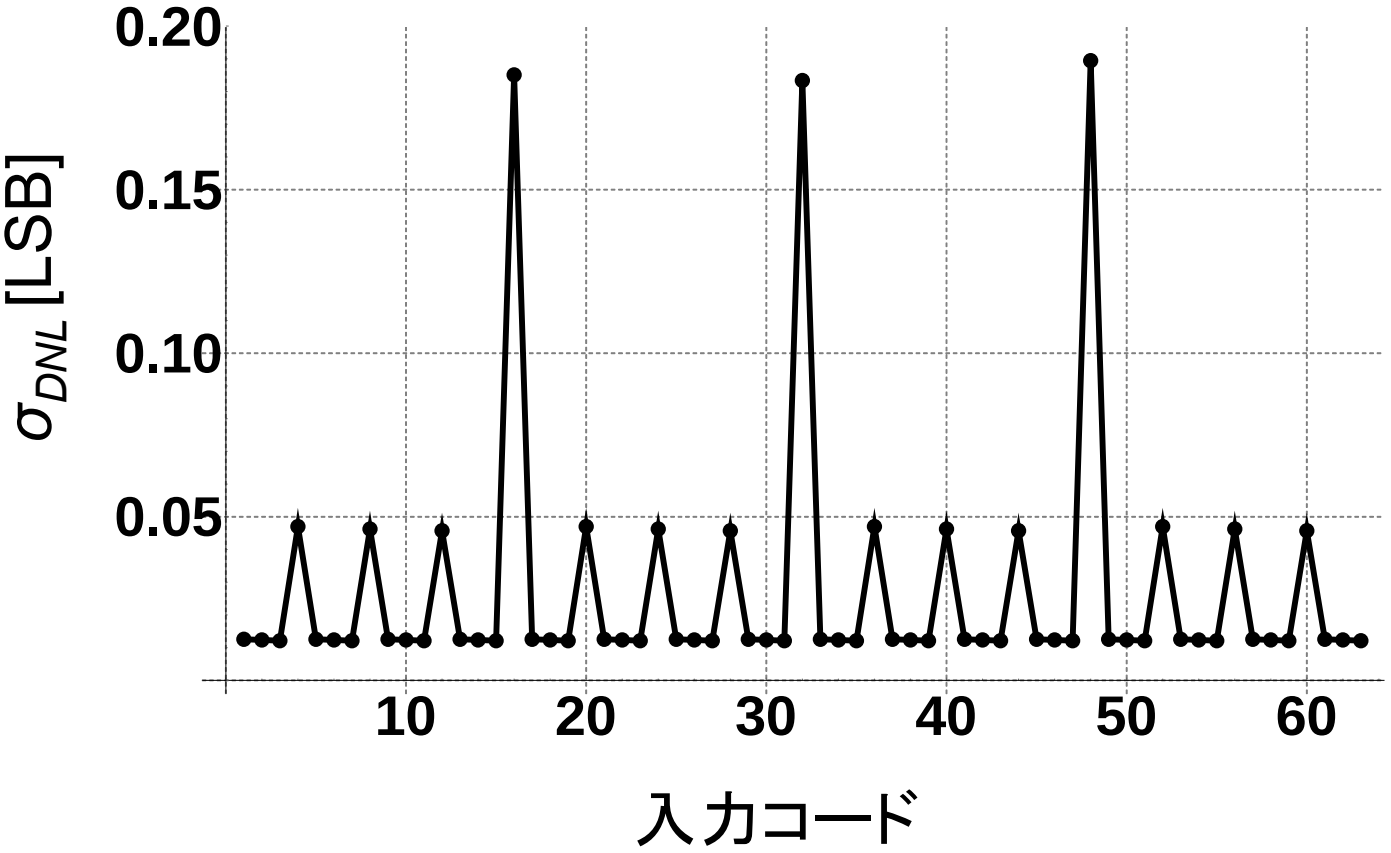
3段3進ラダーDAC シミュレーション

- コード9 と 18でσ_{DNL}が最大
- 最大σ_{DNL}は2番目に大きいσ_{DNL}のおよそ3倍
- 最小σ_{DNL}は2入力について連続で現れる



3段4進ラダーDAC シミュレーション

- コード16, 32, 48でσ_{DNL}が最大
- 最大σ_{DNL}は2番目に大きいσ_{DNL}のおよそ4倍
- 最小σ_{DNL}は3入力について連続で現れる



アウトライン

- 背景と目的
- N進抵抗ラダーDAC
  - 構成と例
- 素子ばらつきによるDNL劣化の解析
  - 数式による出力電圧誤差の見積もり
  - N進DACへの適用
- シミュレーションによる検討
- まとめ

まとめ

- 電流非2進比に分流する抵抗ラダーを用いた
  電流モード DACのDNLの特性について、
  近似した数式を用いた結果と
  モンテカルロシミュレーションの結果を示した
  - 抵抗と電流のばらつきを仮定した
    3進ラダーDACと4進ラダーDAC
  - 3進ラダー: DNLは約1/3ずつ小さくなっていく
    4進ラダー: DNLは約1/4ずつ小さくなっていく
- DNLの標準偏差σ_{DNL}はラダーDACの構成に
  依存した特定のコードで劣化する
  - DNLが劣化する特定コードについて着目し、
    歩留まりの推定、効果的な自己校正、量産試験アルゴリズムの開発に役立てる

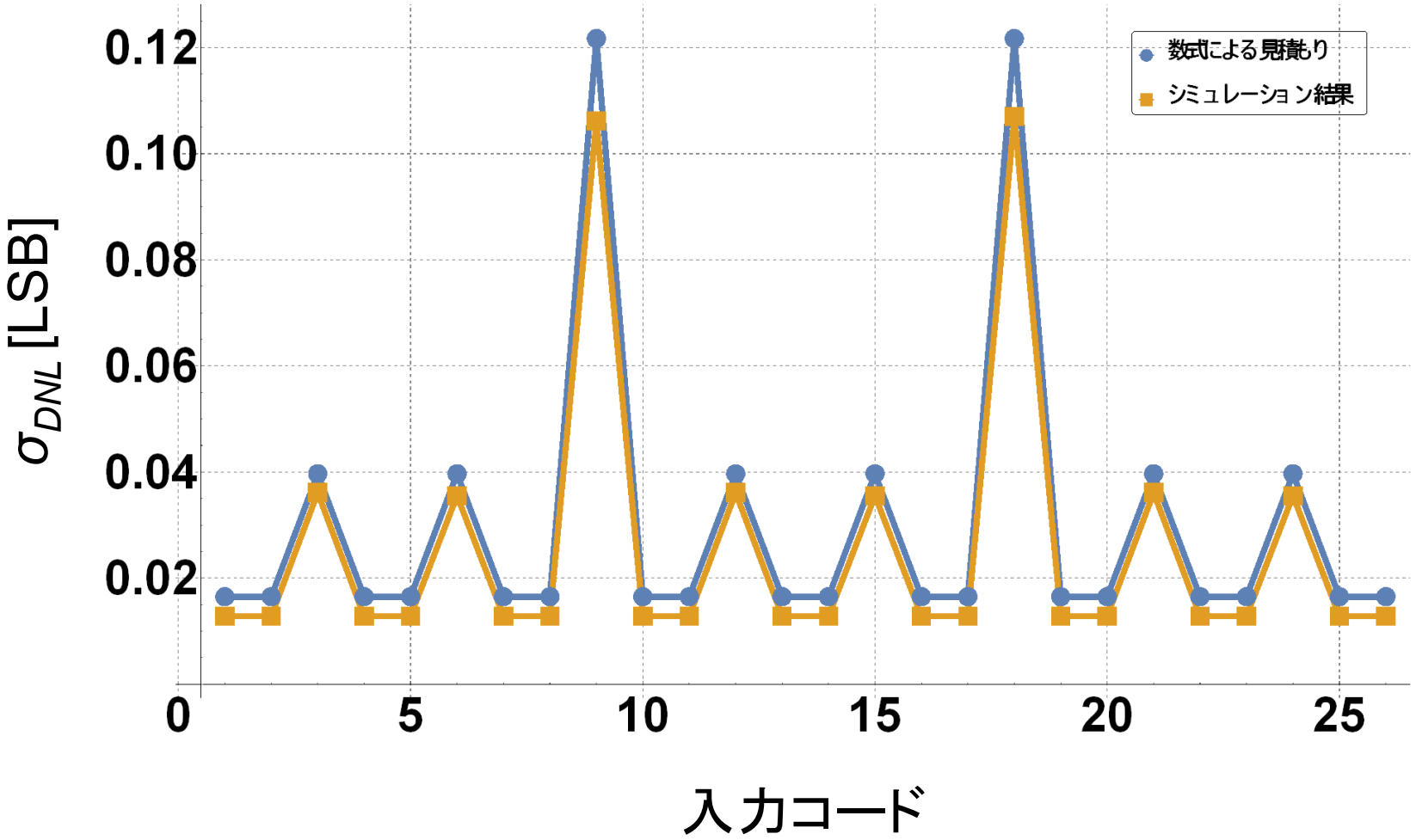
コメント・Q&A

- ラダーの構成を変えて、その構成に応じた境界でDNLが劣化していることを示しているのか。
  - そうです。
  - このままつかうことはあまり現実的ではないとは考えている。（R-2R のデコーダ不要->要デコーダになるなどなど）部分的に構成を変化させて、DNL/コストがよくなるような構成を考えられないかを検討している。
- 「近くの素子のばらつきが小さい」や「素子値が勾配を持つ場合」などを考慮して計算を行うことはできるか？
  - 今の段階ではできない。
    （計算はランダムばらつきのみを仮定している）
    今後の検討課題とします。

以降　資料

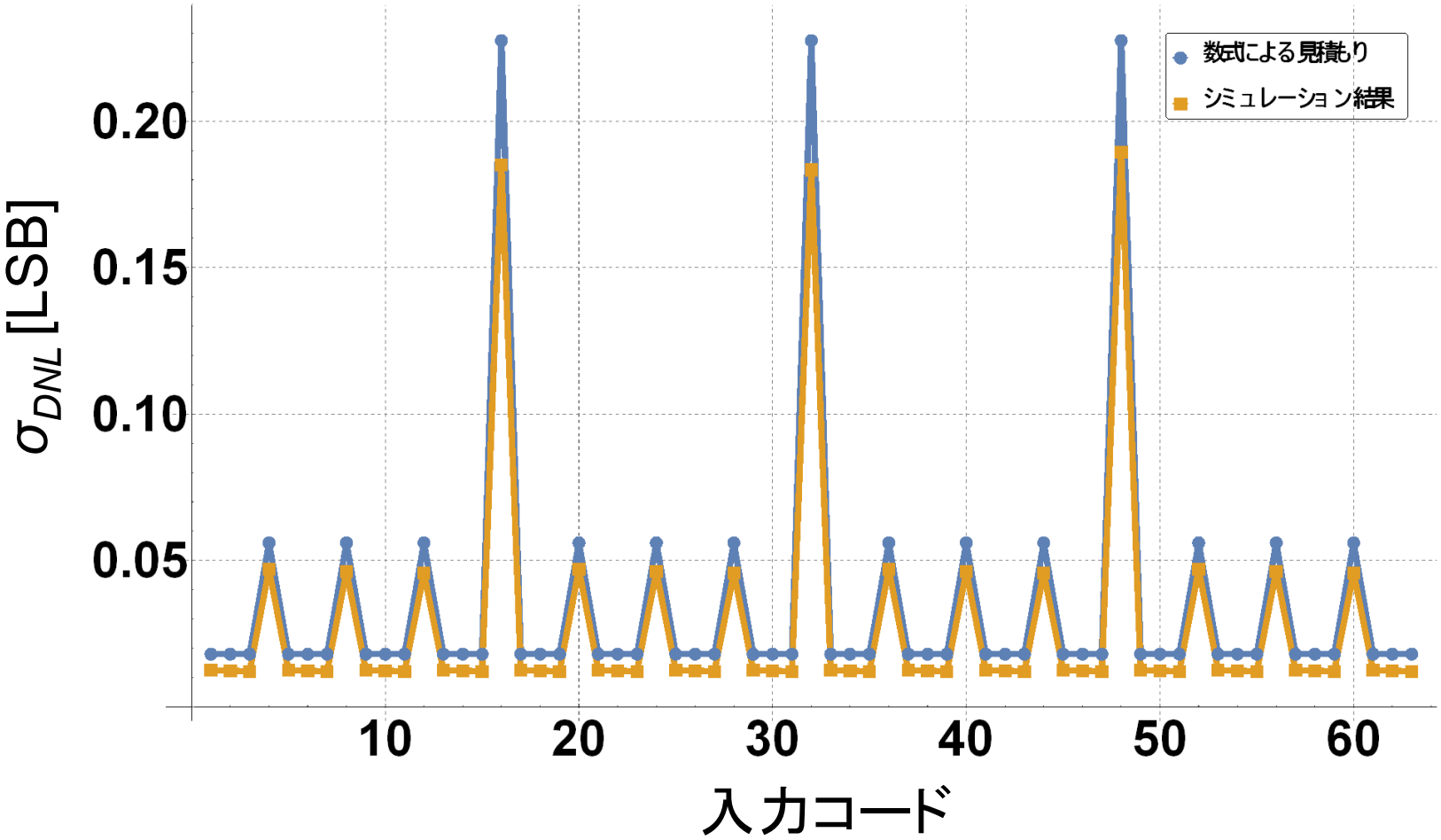
3段3進　比較

- 数式による検討とシミュレーションとの比較



3段4進　比較

- 数式による検討とシミュレーションとの比較



参考文献における図・表

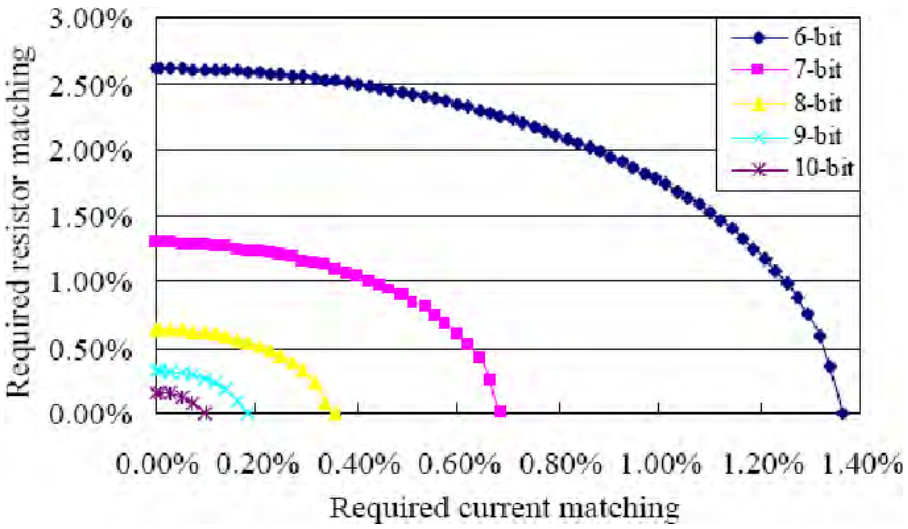


Figure 2. Elliptic curves to demonstrate the theoretical current/resistor matching derived

TABLE I. MONTE-CARLO SIMULATION RESULTS FOR VARIOUS DEVICE MATCHING CALCULATED FROM (17)

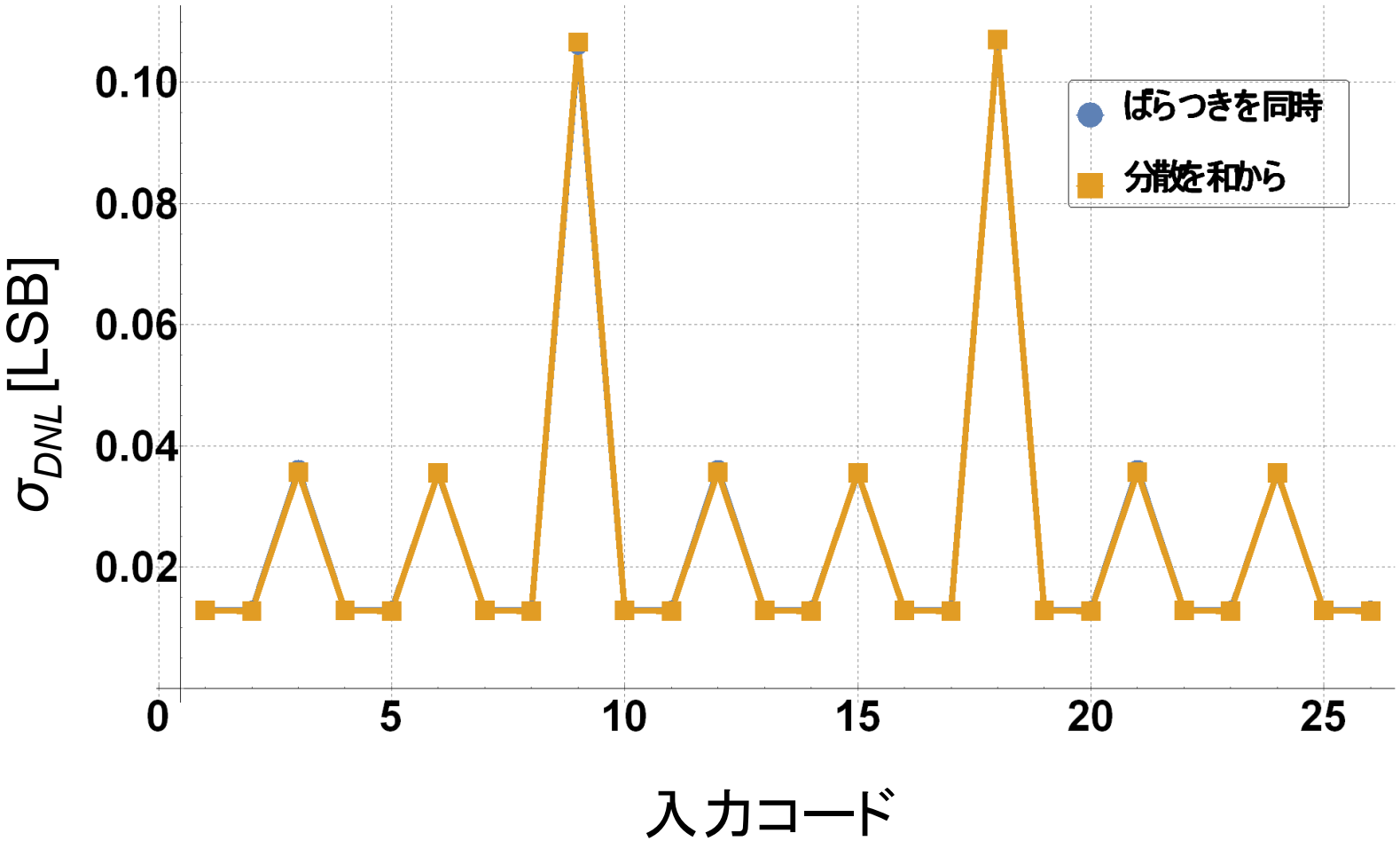
# of Bits	Device matching calculated from (17)		Monte-Carlo Simulated σ_{DNL}
	$\sigma_{\Delta R}$	$\sigma_{\Delta I}$	σ_{DNL} (LSB)
6	2.00 %	0.870 %	0.490
	1.00 %	1.250 %	0.493
	0.50 %	1.328 %	0.494
7	1.00 %	0.426 %	0.485
	0.50 %	0.623 %	0.486
	0.25 %	0.664 %	0.489
8	0.50 %	0.210 %	0.501
	0.25 %	0.311%	0.511
9	0.20 %	0.131 %	0.493
	0.10 %	0.161 %	0.494
10	0.10 %	0.066%	0.489

“Additionally, the slight discrepancy between simulated σDNL and theoretical 0.5 LSB is owing to neglect of ΔI · f(ΔRDi) term in (1).”

[3] C. Chen, N. Lu, “Nonlinearity analysis of R-2R Ladder-Based Current-Steering Digital to Analog Converter,”
IEEE International Symposium on Circuits and Systems (May 2013)

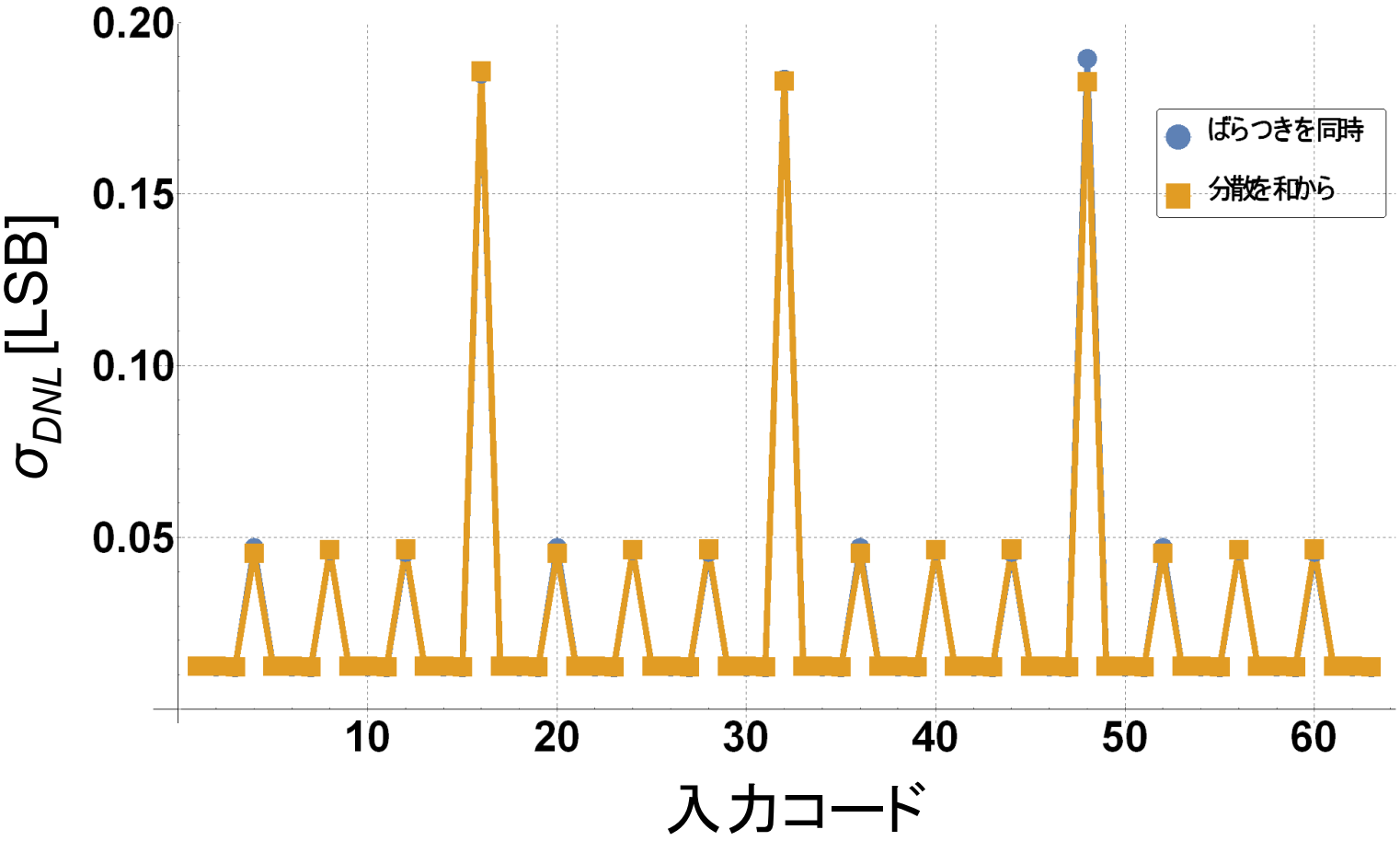
3段3進 DNL標準偏差

- 「電流・抵抗のばらつきを同時に考慮した場合」と、
  「独立にばらつかせた場合の分散の和から求めた場合」



3段4進 DNL標準偏差

- 「電流・抵抗のばらつきを同時に考慮した場合」と、
「独立にばらつかせた場合の分散の和から求めた場合」



抵抗ラダー型デジタルアナログ変換器の微分非直線性の解析

平井　愛統*　（群馬大学）　谷本　洋（北見工業大学）
源代　裕治　　山本　修平　　桑名　杏奈　　小林　春夫（群馬大学）

Differential Nonlineality Analysis for Resistive Ladder-Based Digital-to-Analog Converters

Manato Hirai*(Gunma University) ,Hiroshi Tanimoto, (Kitami Institute of Technology)
Yuji Gendai, Shuhei Yamamoto, Anna Kuwana, Haruo Kobayashi (Gunma University)

This paper presents the differential nonlinearity analysis for several types of resistive ladder-based current-steering digital-to-analog converters by both mathematical technique and Monte-Carlo simulation. We have clarified the trends of DNL for the DAC where a resistor ladder divides the current into the non-binary ratio is used, and shown the difference from those of the R-2R ladder DAC. These results would be useful to estimate their yields and develop their efficient calibration and production testing methods.

キーワード：デジタルアナログ変換器，微分非線形性，抵抗ラダー，モンテカルロ法，非２進重みづけ，
(Digital-to-Analog Converter, Differential Nonlinearity, Resistor Ladder, Monte- Carlo Simulation, Non-binary weighting)

1. はじめに

デジタル/アナログ信号処理において、様々な種類のデジタルアナログ変換器（DAC）が、その特性を生かした用途で用いられている[1, 2]。その中で R-2R 電流モード DAC は、R-2R 抵抗ラダーを用いたシンプルな回路構成と電流モード DAC の特徴である比較的高速な動作という特徴を持つ。しかし、分解能が増加すると、その変換の直線性は抵抗のミスマッチや電流のミスマッチによって劣化する[3]。

本稿では、電流を R-2R ラダーとは異なる比（非 2 進比）に分流する抵抗ラダーを用いて電流モード DAC を構成した場合の DNL の特性について解析する。用いる抵抗ラダーの特性を変化させることで統計的に DNL が劣化するコードの特性が変化することを示す。これらの結果は DAC の歩留まり推定やキャリブレーション、量産時テストの効率的手法の開発の際に役立てることができる。

2. N 進抵抗ラダーDAC

〈2･1〉　N 進抵抗ラダーDAC の構成

これまでに、抵抗ラダーの特性を変化させて抵抗ラダー型電流モード DAC を構成する方法を検討してきた[4-8]。検討の結果として、図 1 にN進抵抗ラダーDAC の回路図を示す。Nは任意に決めることのできる基数、Kはラダー段数、jは抵抗ラダーのノード番号、Rは規格化抵抗値、Iは単位電流である。抵抗ラダーを構成する抵抗比は$N:(N-1)^2$であり、出力と反対の終端は$N-1$の抵抗で終端する。例として、$K=5$、$N=2$の場合、5-bit R-2R ラダー型 電流モード DAC である。I_jはラダーのj番目のノードに流れ込む電流とすると、出力電圧V_{OUT}は(1)式であらわされる。

$$V_{\mathrm{OUT}}(I_1,\cdots,I_K,R,N,K)=(N-1)R\sum_{j=1}^{K}\left(\frac{I_j}{N^{K-j}}\right) \quad \cdots\cdots (1)$$

(1)式で示されるように、ラダーの各段に流れ込む電流I_jは、出力端子に近づくにしたがって出力電圧に対してN倍ずつの重みをもつようになっている。図 1 のように、抵抗ラダーの各段には$N-1$個の電流源が接続されているため、出力電圧範囲を等間隔に分割したN^K-1段階の電圧を得るこ

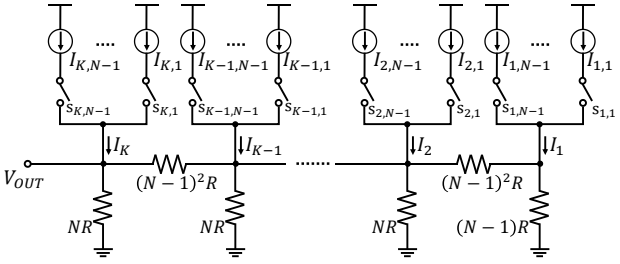


図 1　N 進抵抗ラダー型電流モード DAC

Fig.1 *N*–ary ladder-based current-steering DAC

とができる。

出力端子電圧の最大値V_{MAX}は、(1)式のすべてのI_kについて$(N-1)\cdot I$を代入して求められ、(2)式であらわされる。

$$V_{\mathrm{MAX}}(I,R,N,K) = RI\cdot(N-1)^2\cdot\sum_{j=1}^{K}\left(\frac{1}{N^{K-j}}\right) = RI\cdot N(N-1)\cdot\left(1-\frac{1}{N^K}\right) \quad \text{.....(2)}$$

また、出力電圧の最小ステップは、(1)式においてI_1にIを代入し、それ以外のI_kに 0 を代入することで求められ、 (3)式で表される。

$$V_{\mathrm{MIN}}(I,R,N,K) = (N-1)RI\cdot\frac{1}{N^{K-1}} \quad \text{.....(3)}$$

この回路を DAC として動作させるためにはバイナリコードを N 進コードに変換して各段の電流を操作するデコーダ回路が必要である。また、2 進数と N 進数の桁上がり/桁下がりが起こる値は一般には一致しないため、出力電圧の最大値は DAC として出力電圧の最大値とは限らない。

(1)式を$K=5$、$N=2$とした場合の例として、図 2 の 5-bit R-2R 電流モード DAC の出力電圧は(4)式で表される。ラダーのそれぞれの段に流し込まれる電流は、出力に対して 2 倍ずつの重みをもつ。

$$V_{\mathrm{OUT}}(I_1,I_2,I_3,I_4,I_5,R) = R\left(I_5+\frac{1}{2}I_4+\frac{1}{4}I_3+\frac{1}{8}I_2+\frac{1}{16}I_1\right) \quad \text{.....(4)}$$

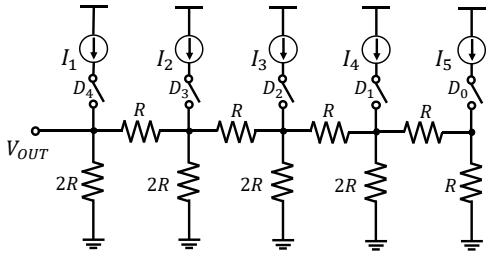


図 2　5 ビット R-2R 電流モード DAC.

Fig.2　5-bit R-2R current-steering DAC.

〈2･2〉　N=3, 3 進 DAC

(1)式において$N=3$とした場合、電流I_jが流し込まれる抵抗ラダーのノードが出力端子に近づくにしたがって、各段に流し込まれる電流の出力電圧に対する重みは、3 倍ずつ大きくなる。各段には 2 つの単位電流源が接続されているため、I_jはIと$2I$の 2 値をとることができ、出力端子では3^K-1段階の電圧を得られる。図 3 に$K=3$、$N=3$とした場合の 3 段 3 進 DAC を示す。(1)式から、このときの出力端子電圧を(5)式で表す。

$$V_{\mathrm{OUT}}(I_1,I_2,I_3R) = 2R\left(I_3+\frac{1}{3}I_2+\frac{1}{3^2}I_1\right) \quad \text{.....(5)}$$

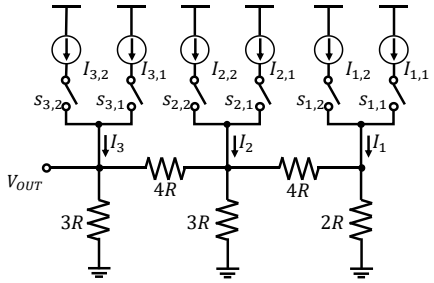


図 3　3 段 3 進抵抗ラダー 電流モード DAC.

Fig.3　3-stage ternary-ladder current-steering DAC.

〈2･3〉　N=4, 4 進 DAC

(1)式において$N=4$とした場合、ラダーの各段に流し込まれる電流I_jは、出力端子に近づくにしたがって出力電圧に対して 4 倍ずつの重みをもつ。各段には 3 つの単位電流源が接続されているため、I_jはI、$2I$、$3I$の 2 値をとることができ、出力端子では4^K-1段階の電圧を得られ、6 ビットの DAC に相当する。

図 4 に$K=3$、$N=4$とした場合の 3 段 4 進 DAC を示す。(1)式から、このときの出力端子電圧を(6)式で表す。

$$V_{\mathrm{OUT}}(I_1,I_2,I_3R) = 3R\left(I_3+\frac{1}{4}I_2+\frac{1}{4^2}I_1\right) \quad \text{..... (6)}$$

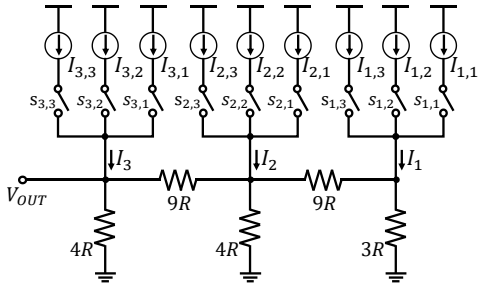


図 4　3 段 4 進抵抗ラダー 電流モード DAC.

Fig.4 3-stage quaternary-ladder current-steering DAC.

3.　素子ばらつきによる DNL 悪化の解析

〈3･1〉　数式による出力電圧誤差の見積もり

R-2R 電流モード DAC を単位電流源と単位抵抗からなる回路モデルで表した場合の、素子ミスマッチを考慮した DNL 特性について詳細な検討がなされており、DNL の標準偏差を 0.5LSB より小さくするために必要とされる電流源と抵抗のマッチングが明らかになっている[3]。この手法を本稿で示した抵抗ラダー型電流モード DAC について適用した結果について述べる。

図 5 に抵抗と電流のばらつき含んだK段N進抵抗ラダー DAC の回路図を示す。$I_{j,i}$と$\Delta I_{j,i}$はj番目のノードに接続されたi個目の電流源とその誤差、$R_{\mathrm{V}(j)}$と$\Delta R_{\mathrm{V}(j)}$はj番目のノードと接地との間の抵抗とその誤差、$R_{\mathrm{H}(j)}$と$\Delta R_{\mathrm{H}(j)}$はj番目のノードと$j+1$番目のノードの間の抵抗とその誤差である。また、$R_{\mathrm{T}(j)}$と$\Delta R_{\mathrm{T}(j)}$は、j番目のノードから右側を見込んだ合成抵抗とそのばらつきであり、抵抗にばらつきがない場合は

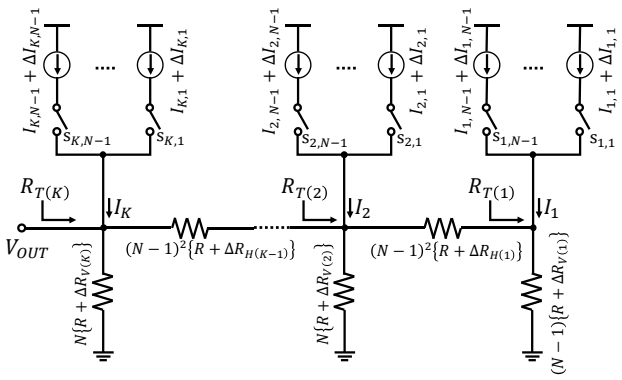


図 5　ばらつき含んだ K 段 N 進抵抗ラダーDAC.

Fig.5　*K*–stage *N*–ary ladder-based DAC with current and resistor mismatches.

$R_{\mathrm{T}(j)}=(N-1)R$である。図 5 から、$R_{\mathrm{T}(1)}$は$R_{\mathrm{V}(1)}$である。

j 番目のノードに流れる電流I_jの状態を示すために、図 5 中のスイッチの状態を$\mathrm{s}_{j,i}$であらわすと、$\mathrm{s}_{j,i}$のみが導通している場合の抵抗と電流の誤差を含んだ出力電圧は、(7)式で表される。

$$V_{OUT}|_{\mathrm{s}_{j,i}=1}=\left(I+\Delta I_{j,i}\right)\cdot\left\{\frac{R(N-1)}{N^{K-i}}+f\left(\Delta R_{\mathrm{s}_{j,i}}\right)\right\}\\ \cong\frac{N-1}{N^{K-i}}RI+\frac{(N-1)R}{N^{K-i}}\cdot\Delta I_{j,i}+f(\Delta V_R) \quad \text{.....(7)}$$

(7)式でjは 1 からKまでの値をとり、iは 1 から$N-1$をとる。第 1 項は、抵抗と電流にばらつきがない時の出力電圧を示す。第 2 項は、電流のばらつき$\Delta I_{j,i}$による出力電圧のばらつきを示す。第 3 項は、すべての抵抗の誤差に起因する出力電圧のばらつきを示す。第 4 項である$\Delta I_{j,i}\cdot f(\Delta R_{\mathrm{s}_{j,i}})$は出力電圧への寄与が小さいとして、無視できるものしている。

図 5 から、$R_{\mathrm{T}(K)}$は、N、R、$\Delta R_{\mathrm{V}(K)}$、$\Delta R_{\mathrm{H}(K-1)}$、$R_{\mathrm{T}(K-1)}$を用いて(8)式で表すことができる。

$$\begin{aligned}R_{\mathrm{T}(K)}&=\{N(R+\Delta R_{\mathrm{V}(K)})\}//\{(N-1)^2\left(R+\Delta R_{\mathrm{H}(K-1)}\right)\\&\quad+(N-1)R+\Delta R_{\mathrm{T}(K-1)}\}\\&=\frac{N\left(R+\Delta R_{\mathrm{V}(K)}\right)\{(N-1)^2\left(R+\Delta R_{\mathrm{H}(K-1)}\right)+(N-1)R+\Delta R_{\mathrm{T}(K-1)}\}}{N^2R+N\Delta R_{\mathrm{V}(K)}+(N-1)^2\Delta R_{\mathrm{H}(K-1)}+\Delta R_{\mathrm{T}(K-1)}}\\&\cong\frac{NR\{N(N-1)R+N(N-1)\Delta R_{\mathrm{V}(K)}+(N-1)^2\Delta R_{\mathrm{H}(K-1)}+\Delta R_{\mathrm{T}(K-1)}\}}{N^2R+N\Delta R_{\mathrm{V}(K)}+(N-1)^2\Delta R_{\mathrm{H}(K-1)}+\Delta R_{\mathrm{T}(K-1)}}\end{aligned} \quad (8)$$

(8)式 3 行目では、微小な値である誤差どうしの積を無視している。

次に、(8)式を$x\ll1$の時のテイラー展開 1 次項までの近似式$1/(1+x)\cong1-x$を用いて近似し、$R_{\mathrm{T}(K)}$を(9)式であらわす。

$$\begin{aligned}&R_{\mathrm{T}(K)}\\&\cong\frac{(N-1)R+(N-1)\Delta R_{\mathrm{V}(K)}+\frac{(N-1)^2}{N}\Delta R_{\mathrm{H}(K-1)}+\frac{\Delta R_{\mathrm{T}(K-1)}}{N}}{1+\frac{N\Delta R_{\mathrm{V}(K)}}{N^2R}+\frac{(N-1)^2\Delta R_{\mathrm{H}(K-1)}}{N^2R}+\frac{\Delta R_{\mathrm{T}(K-1)}}{N^2R}}\\&\cong\left\{(N-1)R+(N-1)\Delta R_{\mathrm{V}(K)}+\frac{(N-1)^2}{N}\Delta R_{\mathrm{H}(K-1)}+\frac{\Delta R_{\mathrm{T}(K-1)}}{N}\right\}\\&\quad\times\left\{1-\frac{\Delta R_{\mathrm{V}(K)}}{NR}-\frac{(N-1)^2\Delta R_{\mathrm{H}(K-1)}}{N^2R}-\frac{(N-1)\Delta R_{\mathrm{T}(K-1)}}{N^2R}\right\}\\&\cong(N-1)\left\{R+\Delta R_{\mathrm{V}(K)}+\frac{(N-1)}{N}\cdot\Delta R_{\mathrm{H}(K-1)}+\frac{\Delta R_{\mathrm{T}(K-1)}}{N}\right\}\\&\cong(N-1)R+\frac{(N-1)^2}{N}\cdot\Delta R_{V(K)}\\&\quad+\frac{(N-1)^2}{N^2}\cdot\Delta R_{H(K-1)}+\frac{1}{N^2}\cdot\Delta R_{T(K-1)}\end{aligned} \quad \text{..... (9)}$$

(9)式の手順で、j番目のノードから見込んだ抵抗$R_{\mathrm{T}(j)}$は(10)式で表すことができる。

$$\begin{aligned}R_{T(j)}&\cong(N-1)R+\frac{(N-1)^2}{N}\cdot\Delta R_{V(j)}\\&\quad+\frac{(N-1)^2}{N^2}\cdot\Delta R_{H(j-1)}+\frac{1}{N^2}\cdot\Delta R_{T(j-1)}\end{aligned} \quad \text{..... (10)}$$

(9)式を用いて、K番目のノードに電流を流し込んだ時の出力電圧$V_{OUT}|_{\mathrm{s}_{K,i}=1}$を(11)式であらわすことができる。このとき、$R_{\mathrm{T}(K)}$は(10)式を用いて、再帰的に展開することができ、出力電圧は$I_{K,i}$、R、$j=1$からKまでの$\Delta R_{\mathrm{V}(j)}$、$j=1$から$K-1$までの$\Delta R_{\mathrm{H}(j)}$を用いて表される。

$$V_{OUT}|_{\mathrm{s}_{K,i}=1}=I_{K,i}\cdot R_{\mathrm{T}(K)} \quad \text{.....(11)}$$

同様の近似手法を用いて、電流の誤差を考慮しない場合に、$I_{K-1,j}$から$K-1$番目のノードに電流を流し込んだ時の出力電圧$V_{OUT}|_{\mathrm{s}_{K-1,i}=1}$を求める。

$$\begin{aligned}V_{OUT}|_{\mathrm{s}_{K-1,i}=1}&=I_{K-1,i}\cdot\frac{R_{\mathrm{T}(K-1)}\cdot R_{\mathrm{V}(K)}}{R_{\mathrm{V}(K)}+R_{\mathrm{H}(K-1)}+R_{\mathrm{T}(K-1)}}\\&=\frac{I_{K-1,i}\cdot\{(N-1)R+\Delta R_{\mathrm{T}(K-1)}\}\cdot\{N\left(R+\Delta R_{\mathrm{V}(K)}\right)\}}{(N-1)R+\Delta R_{\mathrm{T}(K-1)}+N\left(R+\Delta R_{\mathrm{V}(K)}\right)+(N-1)^2\left(R+\Delta R_{\mathrm{H}(K-1)}\right)}\\&\cong\frac{I_{K-1,i}\{N(N-1)R+NR\Delta R_{\mathrm{T}(K-1)}+N(N-1)R\Delta R_{\mathrm{V}(K)}\}}{N^2R+\Delta R_{\mathrm{T}(K-1)}+N\Delta R_{\mathrm{V}(K)}+(N-1)^2\Delta R_{\mathrm{H}(K-1)}}\\&=\frac{I_{K-1,i}\left\{\frac{(N-1)R}{N}+\frac{\Delta R_{\mathrm{T}(K-1)}}{N}+\frac{(N-1)\Delta R_{\mathrm{V}(K)}}{N}\right\}}{1+\frac{\Delta R_{\mathrm{T}(K-1)}}{N^2R}+\frac{N\Delta R_{\mathrm{V}(K)}}{N^2R}+\frac{(N-1)^2\Delta R_{\mathrm{H}(K-1)}}{N^2R}}\\&\cong I_{K-1,i}\cdot\left\{\frac{(N-1)R}{N}+\frac{\Delta R_{\mathrm{T}(K-1)}}{N}+\frac{(N-1)\Delta R_{\mathrm{V}(K)}}{N}\right\}\\&\quad\times\left\{1-\frac{\Delta R_{\mathrm{T}(K-1)}}{N^2R}-\frac{N\Delta R_{\mathrm{V}(K)}}{N^2R}-\frac{(N-1)^2\Delta R_{\mathrm{H}(K-1)}}{N^2R}\right\}\\&\cong I_{K-1,j}\cdot\left\{\frac{N-1}{N}R+\frac{(N-1)^2}{N^2}\Delta R_{\mathrm{V}(K)}+\frac{N^2-N+1}{N^3}\Delta R_{\mathrm{T}(K-1)}-\right.\\&\quad\left.\frac{(N-1)^3}{N^3}\Delta R_{\mathrm{H}(K-1)}\right\}\end{aligned} \quad \text{..... (12)}$$

ある$\mathrm{s}_{j,i}$が導通しているときの抵抗の誤差のみを考慮した出力電圧は、同じ手法で近似して求めることができる。

ある入力コードにおける抵抗の誤差を考慮した出力電圧は、こうして求めた一つの$\mathrm{s}_{j,i}$が導通しているときの出力電圧を入力コードに応じて足し合わせて求められる。

〈3･2〉　3 段 3 進 DAC の場合の DNL 見積もり

DAC の DNL は隣接コードの出力電圧差から計算され、(13)式で定義される[9]。ここで、$V_{\mathrm{OUT}}(n)$はコードnでの出力電圧、V_{LSB}は最小の出力電圧の理想値である。

$$DNL(n)=\frac{V_{\mathrm{OUT}}(n)-V_{\mathrm{OUT}}(n-1)}{V_{\mathrm{LSB}}}-1 \quad \text{..... (13)}$$

抵抗と電流のばらつきを考慮した 3 段 3 進 DAC の場合、3 段目の$\mathrm{s}_{3.i}$、2 段目の$\mathrm{s}_{2.i}$、1 段目の$\mathrm{s}_{1,i}$がそれぞれ導通して

いるときの出力電圧は、(10)式や(11)式の近似手法を用いた場合、(14)式、(15)式、(16)式で表される。これらを各入力コードでの$s_{j,i}$の状態に応じて足し合わせることで、素子のばらつきを考慮した任意の入力コードでの出力電圧を表すことができる。

$$V_{OUT}|_{s_{3,i}=1} \cong (I_{3,i}+\Delta I_{3,i}) \cdot\left\{2R+\frac{4\Delta R_{V(3)}}{3}+\frac{4\Delta R_{H(2)}}{9}+\frac{\Delta R_{T(2)}}{9}\right\}$$

$$=(I_{3,i}+\Delta I_{3,i})\cdot\left[2R+\frac{4\Delta R_{V(3)}}{3}+\frac{4\Delta R_{H(2)}}{9}+\frac{1}{9}\left\{\frac{(3-1)^2}{3}\Delta R_{V(2)}+\frac{(3-1)^2}{3^2}\Delta R_{H(2)}+\frac{3-1}{3^2}\Delta R_{V(1)}\right\}\right]$$

$$=(I_{3,i}+\Delta I_{3,i})\cdot\left\{2R+\frac{4}{3}\Delta R_{V(3)}+\frac{4}{9}\Delta R_{H(2)}+\frac{4}{27}\Delta R_{V(2)}+\frac{4}{81}\Delta R_{H(2)}+\frac{2}{81}\Delta R_{V(1)}\right\} \quad \cdots\cdots(14)$$

$$V_{OUT}|_{s_{2,i}=1}=(I_{2,i}+\Delta I_{2,i})\cdot\left(\frac{2R}{3}+\frac{4\Delta R_{V(3)}}{9}-\frac{8\Delta R_{H(2)}}{27}+\frac{7\Delta R_{T(2)}}{27}\right)$$

$$=(I_{2,i}+\Delta I_{2,i})\cdot\left(\frac{2R}{3}+\frac{4\Delta R_{V(3)}}{9}-\frac{8\Delta R_{H(2)}}{27}+\frac{28\Delta R_{V(2)}}{81}+\frac{28\Delta R_{H(1)}}{243}+\frac{14\Delta R_{V(1)}}{243}\right) \quad \cdots\cdots(15)$$

$$V_{OUT}|_{s_{1,i}=1}=(I_{1,j}+\Delta I_{2,j})\cdot\left(\frac{2R}{9}+\frac{4\Delta R_{V(3)}}{27}-\frac{8\Delta R_{H(2)}}{81}+\frac{28\Delta R_{V(2)}}{243}-\frac{80\Delta R_{H(1)}}{729}+\frac{122\Delta R_{V(1)}}{729}\right) \quad \cdots\cdots(16)$$

(13)式での定義と(14)式、(15)式、(16)式から、3 段 3 進ラダーを用いたときの$DNL(9)$を、(17)式で表す。その他のコードにおける DNL についても、前後の$s_{j,i}$の状態に応じて(14)式、(15)式、(16)式を加減算することで(17)式と同様に表すことができる。

$$DNL(9)=\frac{V_{OUT}|_{s_{3,1}=1}-\sum_{i=1}^{2}\left(V_{OUT}|_{s_{2,i}=1}+V_{OUT}|_{s_{1,i}=1}\right)}{V_{LSB}}-1 \quad (17)$$

(7)式での仮定から、抵抗ラダーを用いた電流モード DAC において、コードごとの出力電圧の誤差は、単位電流の誤差$\Delta I_{j,i}$に起因する成分と、抵抗の誤差に起因する成分$f(\Delta V_R)$とに分けることができる。抵抗のばらつき起因の電圧誤差を含んだ DNL は、(17)式においてすべての$\Delta I_{j,i}$をゼロにすることで得られる。同様に、電流のばらつき起因の電圧誤差がある場合の DNL は、(17)式においてすべての$\Delta R_{V(j)}$、$\Delta R_{H(j)}$をゼロにすることで求められる。

電流のばらつきに起因する DNL の標準偏差σ_{DNL_I}と抵抗ばらつきに起因する DNL 標準偏差σ_{DNL_R}を用いて、電流と抵抗がともにばらついた場合の DNL 標準偏差σ_{DNL}を(18)式で表す。

$$\sigma_{DNL}^2=\sigma_{DNL-R}^2+\sigma_{DNL-I}^2 \quad \cdots\cdots(18)$$

これまでの結果を用いて、単位抵抗と単位電流がばらつき、その標準偏差が平均値の 1%であるとした場合のコードごとの DNL 標準偏差、図 6 に示す。

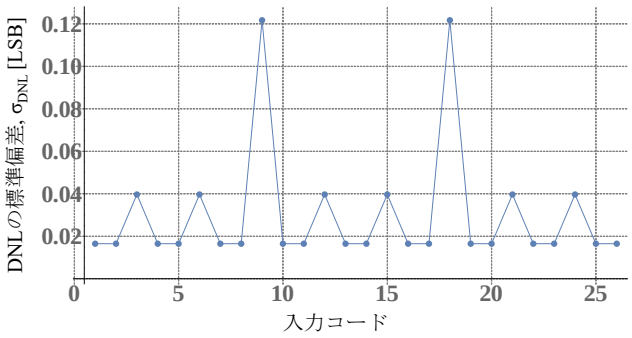


図 6　近似した数式を用いた計算による DNL 標準偏差（3 段 3 進 DAC）.

Fig.6　3-stage ternary DAC, standard deviation of DNL from approximate calculation results.

DNL が最も悪化するコード 9 とコード 18 は、最も出力端子に近い、つまり出力電圧に対する重みの大きい電流が切り替わるコードである。

また、出力から最も離れたノードに流し込まれる電流が切り替わるコードでは DNL が小さくなっている。

〈3･3〉　3 段 4 進 DAC の場合の DNL 見積もり

〈3･2〉における 3 段 3 進 DAC の場合と同様にして、3 段 4 進 DAC で抵抗と電流のばらつきを考慮し、$s_{3,i}$、$s_{2,i}$、$s_{1,i}$のそれぞれが導通しているときの出力電圧は、(19)式、(20)式、(21)式で表される。

$$V_{OUT}|_{s_{3,i}=1} \cong (I_{3,i}+\Delta I_{3,i}) \cdot\left\{3R+\frac{9\Delta R_{V(3)}}{4}+\frac{9\Delta R_{H(2)}}{16}+\frac{\Delta R_{T(2)}}{16}\right\}$$

$$=(I_{3,i}+\Delta I_{3,i})\cdot\left[3R+\frac{9\Delta R_{V(3)}}{4}+\frac{9\Delta R_{H(2)}}{16}+\frac{1}{16}\left\{\frac{(4-1)^2}{4}\Delta R_{V(2)}+\frac{(4-1)^2}{4^2}\Delta R_{H(2)}+\frac{4-1}{4^2}\Delta R_{V(1)}\right\}\right]$$

$$=(I_{3,i}+\Delta I_{3,i})\cdot\left\{3R+\frac{9\Delta R_{V(3)}}{4}+\frac{9\Delta R_{H(2)}}{16}+\frac{9}{64}\Delta R_{V(2)}+\frac{9}{256}\Delta R_{H(2)}+\frac{3}{256}\Delta R_{V(1)}\right\} \quad \cdots\cdots(19)$$

$$V_{OUT}|_{s_{2,i}=1}=(I_{2,i}+\Delta I_{2,i})\cdot\left(\frac{3R}{4}+\frac{3\Delta R_{V(3)}}{9}-\frac{27\Delta R_{H(2)}}{64}+\frac{13\Delta R_{T(2)}}{64}\right)$$

$$=(I_{2,i}+\Delta I_{2,i})\cdot\left(\frac{3R}{4}+\frac{3\Delta R_{V(3)}}{9}-\frac{27\Delta R_{H(2)}}{64}+\frac{17\Delta R_{V(2)}}{256}+\frac{117\Delta R_{H(1)}}{1024}+\frac{39\Delta R_{V(1)}}{1024}\right) \quad \cdots\cdots(20)$$

$$V_{OUT}|_{s_{1,i}=1}=(I_{1,j}+\Delta I_{1,j})\cdot\left(\frac{3R}{16}+\frac{9\Delta R_{V(3)}}{64}-\frac{27\Delta R_{H(2)}}{256}+\right.$$

$$\frac{117\Delta R_{V(2)}}{1024} - \frac{459\Delta R_{H(1)}}{4096} + \frac{615\Delta R_{V(1)}}{4096}\Big) \quad \cdots\cdots (21)$$

3 段 3 進ラダーの場合と同じ方法で、(13)式での定義と(19)式、(20)式、(21)式を用いて、3 段 4 進ラダーを用いたときの$DNL(16)$を(22)式に表す。

$$DNL(16) = \frac{V_{OUT}|_{s_{3,1}=1} - \Sigma_{i=1}^{3}\left(V_{OUT}|_{s_{2,i}=1} + V_{OUT}|_{s_{1,i}=1}\right)}{V_{LSB}} - 1 \quad (22)$$

(7)式での仮定と(18)式を用いて、3 段 4 進ラダーDAC において単位抵抗と単位電流がばらつき、その標準偏差が平均値の 1%であるとした場合の、コードごとの DNL 標準偏差を、図 7 に示す。

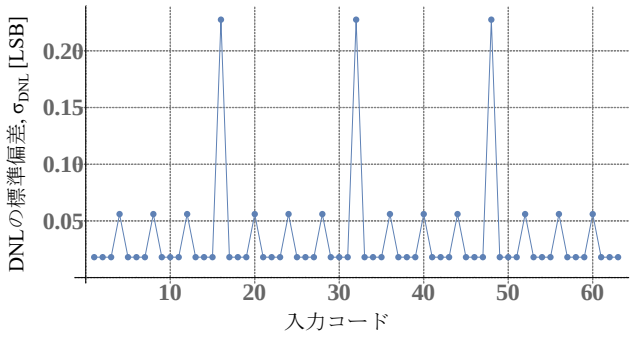


図 7　近似した数式を用いた計算による
DNL 標準偏差（3 段 4 進 DAC）.

Fig.7　3-stage quaternary DAC, standard deviation of DNL from approximate calculation results.

DNL が最も悪化するコード 16、コード 32、コード 48 は、最も出力端子に近い、つまり出力電圧に対する重みが最も大きい電流が切り替わるコードである。

また、出力から最も離れたノードに流し込まれる電流が切り替わるコードでは DNL が小さくなっていて、最も小さい DNL は 3 つの連続した入力について続いて現れる。

4.　モンテカルロシミュレーションによる検討

〈4･1〉　シミュレーション条件

図 3 と図 4 の回路について、モンテカルロシミュレーションによって誤差を持った素子のセットについて出力電圧を求め、DNL の標準偏差を求めた。シミュレーションの条件は以下である。

- 単位抵抗と単位電流は、標準偏差が平均値の 1%である正規分布をする。
- 平均値が定数a倍の抵抗は、標準偏差が$\sqrt{a}$倍された正規分布をする。
- シミュレーションセット数は 3000
- それぞれのセットについて(13)式から DNL を求める。

〈4･2〉　3 段 3 進 DAC シミュレーション結果

図 3 の回路について行ったモンテカルロシミュレーションの結果から、コードごとの DNL の標準偏差を求めた。結果を図 8 に示す。

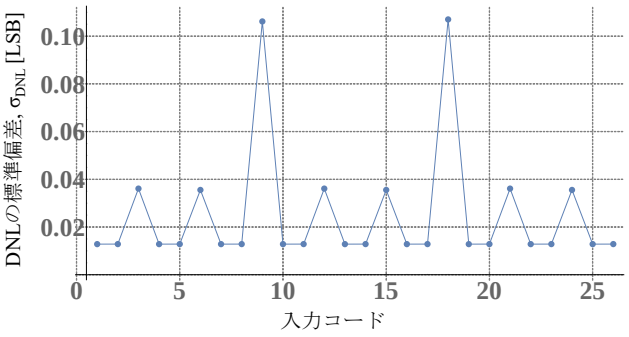


図 8　モンテカルロシミュレーションによる
DNL 標準偏差（3 段 3 進 DAC）.

Fig.8　3-stage ternary DAC, standard deviation of DNL from Monte-Carlo simulation results.

図 8 の DNL 標準偏差の傾向は、近似を用いた数式から求めた DNL 標準偏差である図 6 と似た傾向を示している。3 進ラダーを用いた DAC において、入力レンジの 1/3 と 2/3 での DNL 標準偏差が最大になっており、これは出力端子に最も近い電流が変化するコードである。また、入力レンジを 1/3 に分割した領域においても、同様にその 1/3、2/3 のコードで DNL 標準偏差が大きくなる傾向がみられる。

〈4･3〉　3 段 4 進 DAC シミュレーション結果

図 4 の回路について行ったモンテカルロシミュレーションの結果から、コードごとの DNL の標準偏差を求めた。結果を図 9 に示す。

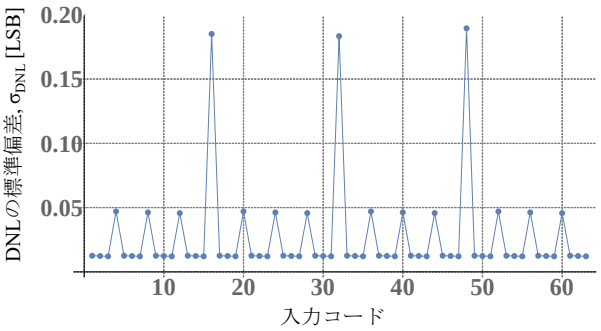


図 9　モンテカルロシミュレーションによる
DNL 標準偏差（3 段 4 進 DAC）.

Fig.9　3-stage quaternary DAC, standard deviation of DNL from Monte-Carlo simulation results.

図 9 から、4 進ラダーを用いた DAC においては、入力レンジの 1/4 と 2/4 と 3/4 での DNL 標準偏差が最大になっている。この傾向は R-2R DAC において上位 2 ビットを温度計コードによる駆動にした場合と同じである。また、入力レンジを 1/4 にずつに分割した領域においても、同様にその 1/4 と 2/4 と 3/4 のコードで DNL 標準偏差が大きくなる傾向がみられる。

5. まとめ

電流を R-2R ラダーとは異なる非 2 進比に分流する抵抗ラダーを用いて、電流モード DAC を構成した場合の DNL の特性について、数式を用いた近似とモンテカルロシミュレーションを用いて解析した。特に 3 進および 4 進抵抗ラダーを用いた電流モード DAC で、構成する抵抗間および電流源間に相対ミスマッチがある場合に、統計的に DNL が劣化するコードの特性を示した。この結果はこれらの DAC の歩留まりの推定、効果的なキャリブレーションおよび量産試験アルゴリズムの開発に役立てることが期待できる。

文　　献

(1) F. Maloberti, Data Converters, Springer (2010).
(2) The Data Conversion Handbook, Analog Devices Inc. (2005)
(3) C. Chen, N. Lu, “Nonlinearity analysis of R-2R Ladder-Based Current-Steering Digital to Analog Converter,” IEEE International Symposium on Circuits and Systems (May 2013)
(4) M. Hirai, S. Yamamoto, H. Arai, A. Kuwana, H. Tanimoto, Y. Gendai, H. Kobayashi, "Systematic Construction of Resistor Ladder Network for N-ary DACs", IEEE ASICON (Oct. 2019)
(5) M. Hirai, H. Tanimoto, Y. Gendai, S. Yamamoto, A. Kuwana, H. Kobayashi,"Nonlineality Analysis of Resistive Ladder-Based Current-Steering Digital-to-Analog Converter"17th International SOC Design Conference Yeosu, Korea (Oct. 2020)
(6) Y. Kobayashi, S. Shibuya, T. Arafune, S. Sasaki, H. Kobayashi, “SAR ADC Design Using Golden Ratio Weight Algorithm”, International Symposium on Communications and Information Technologies, Nara, Japan (Oct. 2015)
(7) S. Yamamoto, M. Hirai, T. Arai, A. Kuwana, H. Kobayashi, K. Kubo, "Proposal of Ternary Resistor Network DACs", 5th Taiwan and Japan Conference on Circuits and Systems, Nikko, Tochigi, Japan (Aug. 2019).
(8) Y. Du, X. Bai, M. Hirai, S. Yamamoto, A. Kuwana, H. Kobayashi, K. Kubo, "Digital-to-Analog Converter Architectures Based on Polygonal and Prime Numbers", 17th International SOC Design Conference Yeosu, Korea (Oct. 2020)

電流非一定分割抵抗ラダーを用いた DA 変換器構成と微分非線形性の解析

平井愛統（Manato Hirai），
谷本 洋，源代裕治，山本修平，
桑名杏奈，小林春夫

群馬大学，北見工業大学

OUTLINE

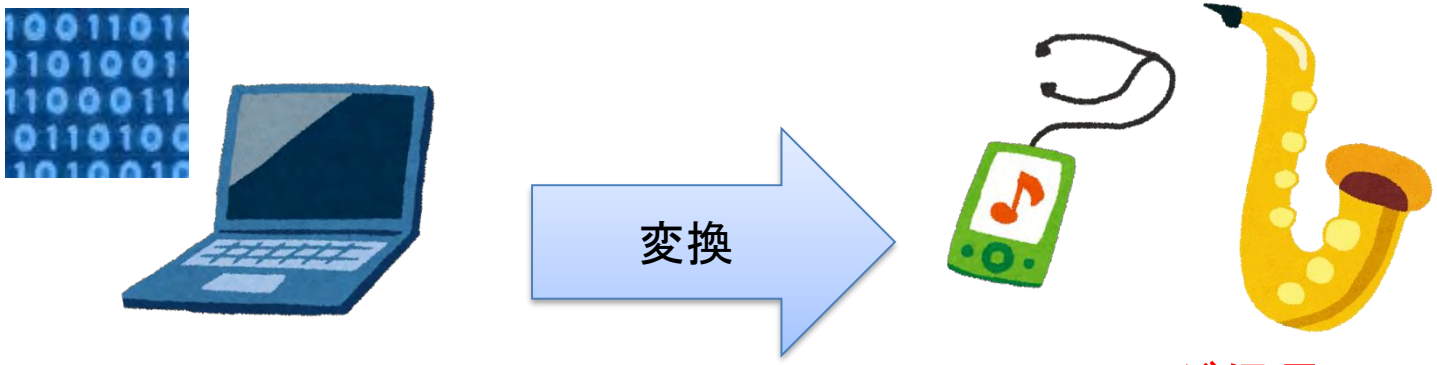
- 研究背景・目的
- DA変換器の構成
  - R-2R 電流源 DAC
- N進抵抗ラダーDACの構成
  - 抵抗ラダーを用いた電流分割
  - N進抵抗ラダーを用いた構成
- 異なる分流比を持つ抵抗ラダーの接続
  - 接続の条件・手順
  - DAC構成案と非線形性シミュレーション
- 結論

OUTLINE

- 研究背景・目的
- DA変換器の構成
  - R-2R 電流源 DAC
- N進抵抗ラダーDACの構成
  - 抵抗ラダーを用いた電流分割
  - N進抵抗ラダーを用いた構成
- 異なる分流比を持つ抵抗ラダーの接続
  - 接続の条件・手順
  - DAC構成案と非線形性シミュレーション
- 結論

研究の背景

- デジタルアナログ変換器（DAC）の用途
  - デジタル信号処理結果の出力



デジタル信号

アナログ信号
（電圧・音・光など）

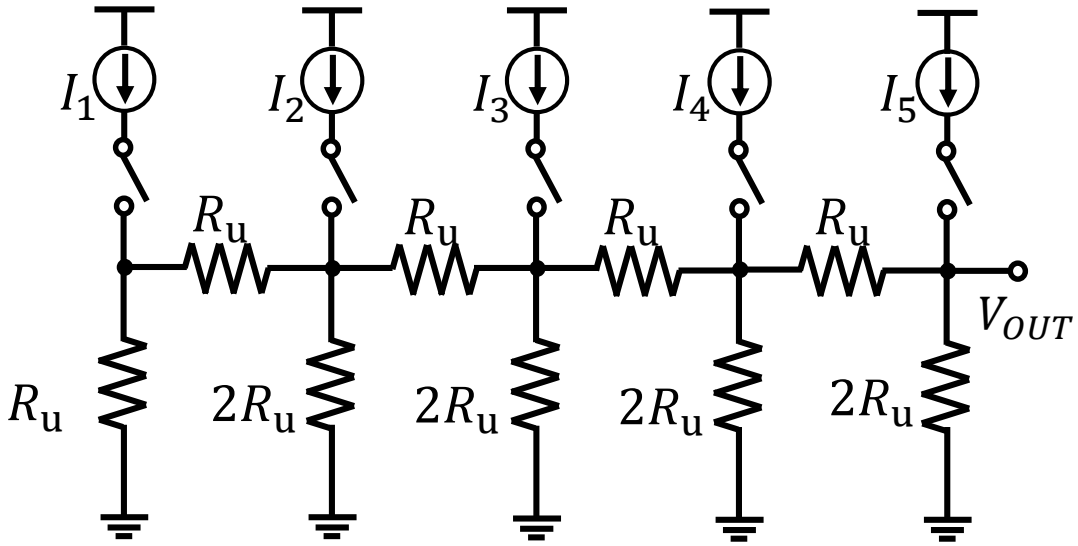
  - アナログデジタル変換器（ADC）の内部回路

目的：抵抗ラダーによる非一定の電流分割を用いた
DA変換器回路構成の検討と性能向上

OUTLINE

- 研究背景・目的
- DA変換器の構成
  - R-2R 電流源 DAC
- N進抵抗ラダーDACの構成
  - 抵抗ラダーを用いた電流分割
  - N進抵抗ラダーを用いた構成
- 異なる分流比を持つ抵抗ラダーの接続
  - 接続の条件・手順
  - DAC構成案と非線形性シミュレーション
- 結論

R-2R 電流源 DAC

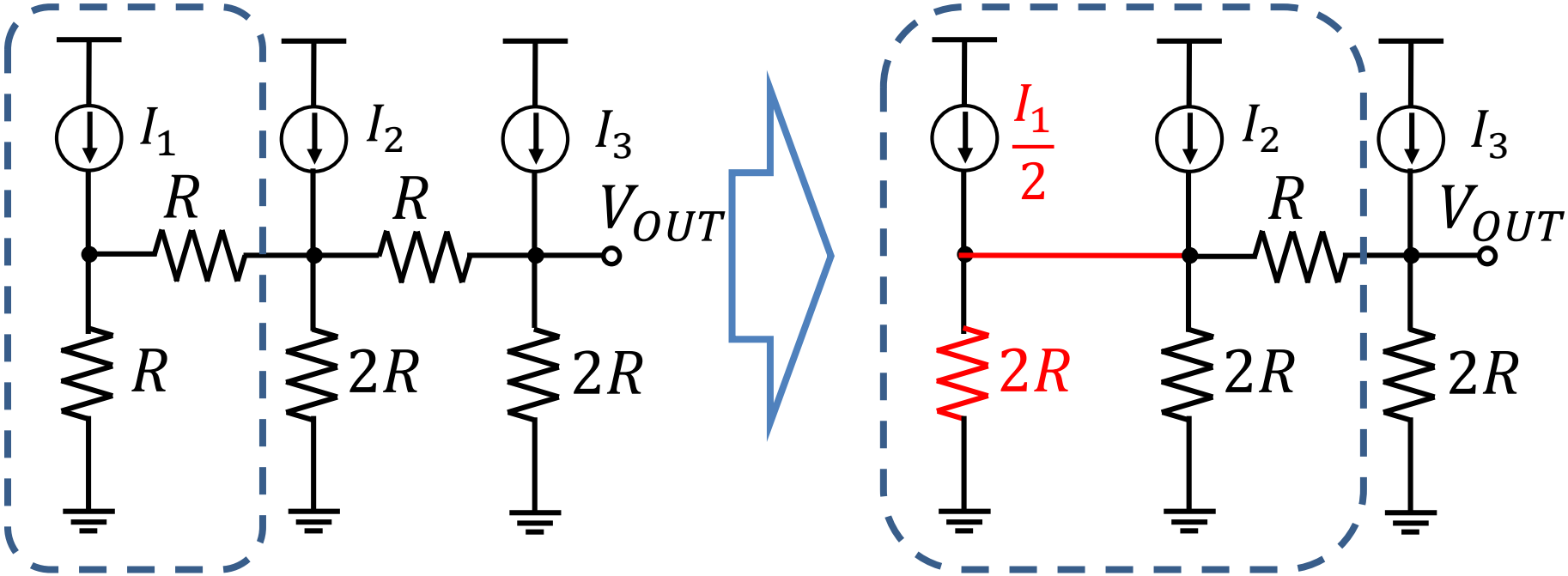


5-bit R-2R Current-steering DAC（線形回路モデル）

- R-2R 抵抗ラダーによる2進重みづけを利用
  - 電流$I_1, \ldots, I_5$は2進重みづけされる
- 利点
  - デコーダ不要
  - 電流源による比較的高速な動作
- 欠点
  - 素子誤差でDNLが劣化
  - 上位ビット変化タイミングでグリッチが発生

R-2R DAC の動作原理

- ノートンの定理を用いて出力から離れた側から等価回路に変換
  - 内部の電流が出力に対して2倍ずつの重みをもつ

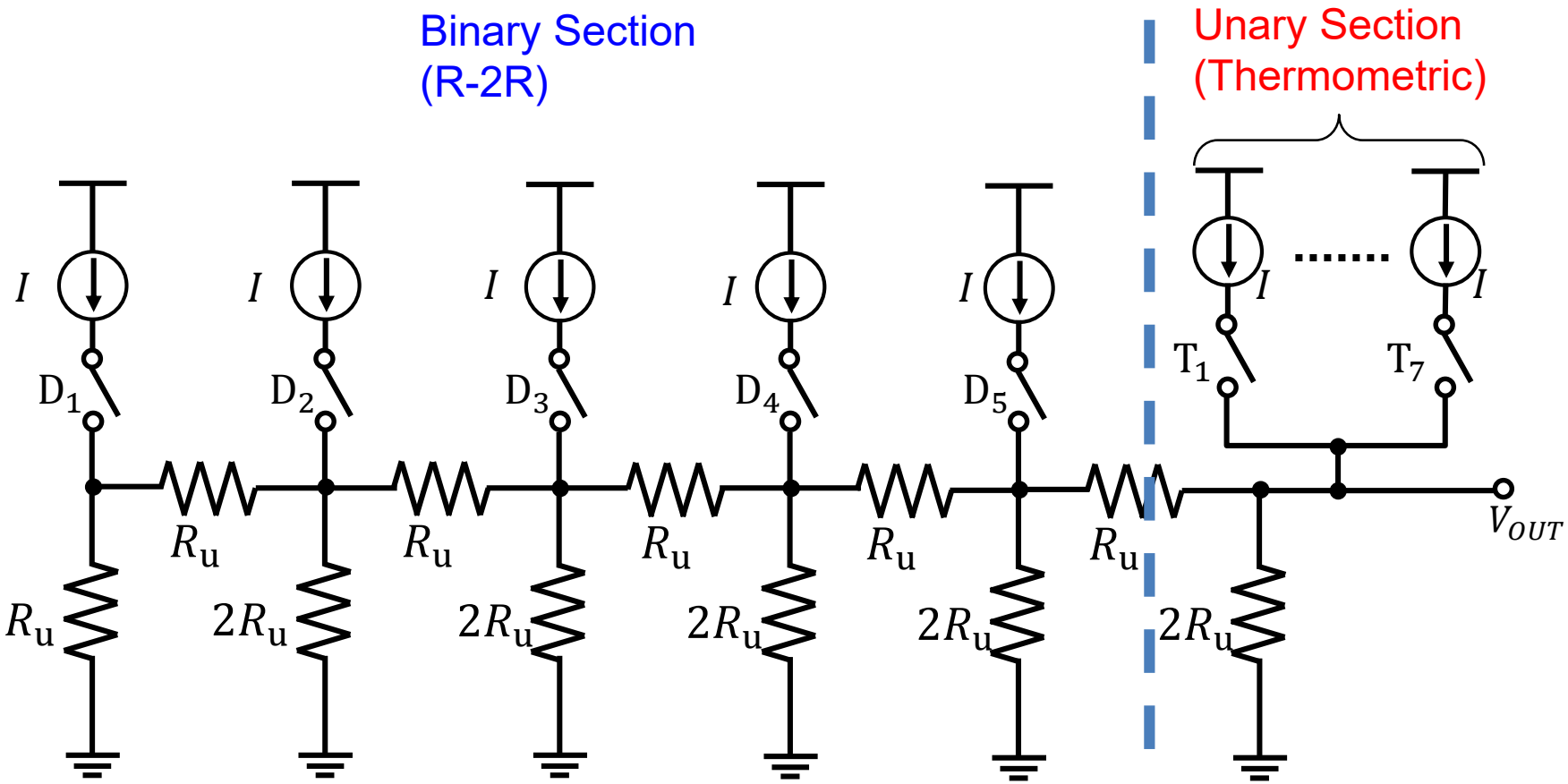


3-bit R-2R DAC

等価回路への変換

セグメント化 R-2R DAC

- 素子ばらつきに起因する線形性劣化・グリッチを軽減
  - Unary部駆動のための温度計デコーダが必要

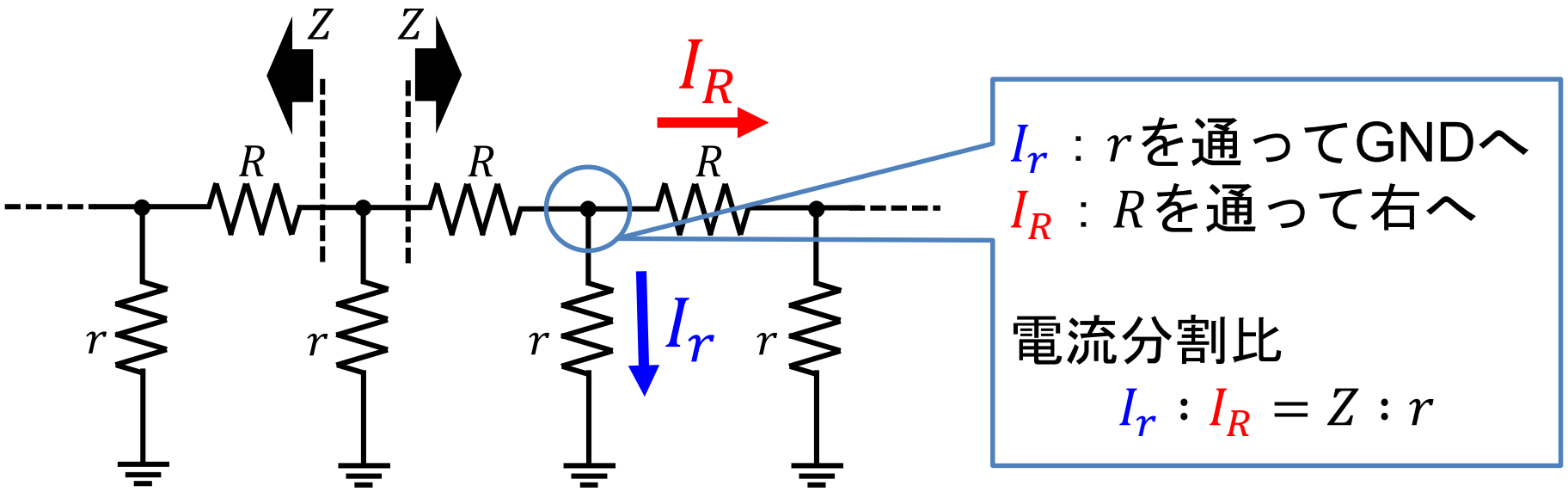


8-bit Segmented R-2R DAC（3-bit Unaryコード駆動）

OUTLINE

- 研究背景・目的
- DA変換器の構成
  - R-2R 電流源 DAC
- N進抵抗ラダーDACの構成
  - 抵抗ラダーを用いた電流分割の検討
  - N進抵抗ラダーを用いた構成
- 異なる分流比を持つ抵抗ラダーの接続
  - 接続の条件・手順
  - DAC構成案と非線形性シミュレーション
- 結論

抵抗ラダーを用いた等比電流分割



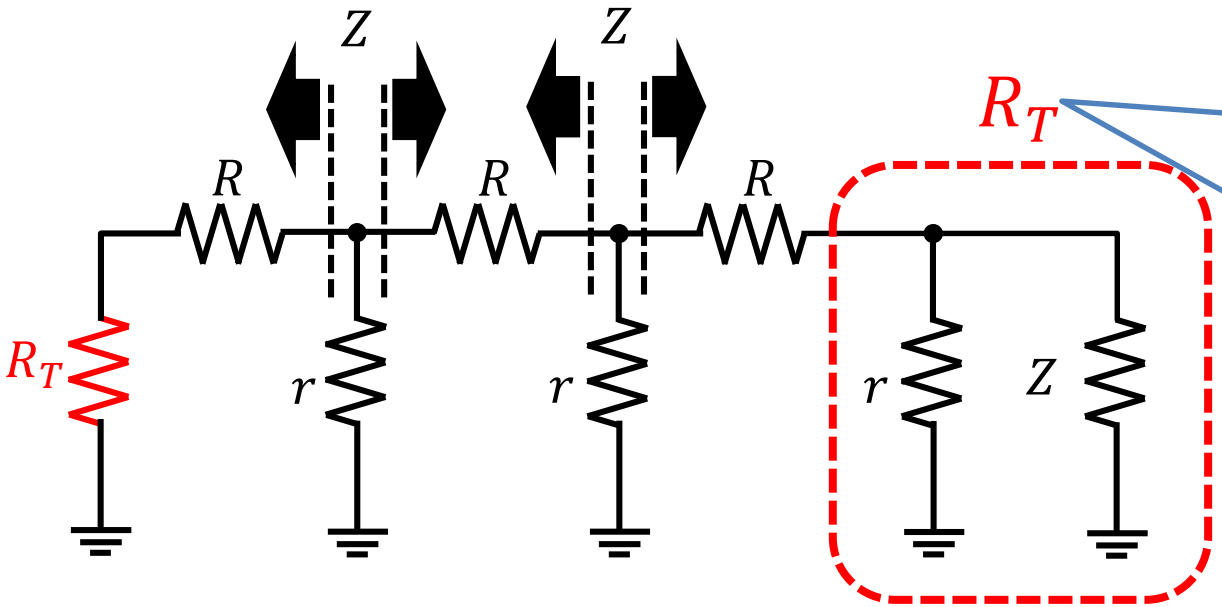
- 無限に続く抵抗ラダーの合成抵抗Z

$$Z = \frac{R}{2} + \frac{\sqrt{R(R+4r)}}{2}$$

- 整数Nについて電流分割比 $N-1:1$ にしたい場合

$$I_r : I_R = Z : r = N-1 : 1$$
$$\leftrightarrow R : r = (N-1)^2 : N$$

無限抵抗ラダーの有限打ち切り



各ノードから左右を見込んだ抵抗がZになるよう、終端抵抗R_Tで終端

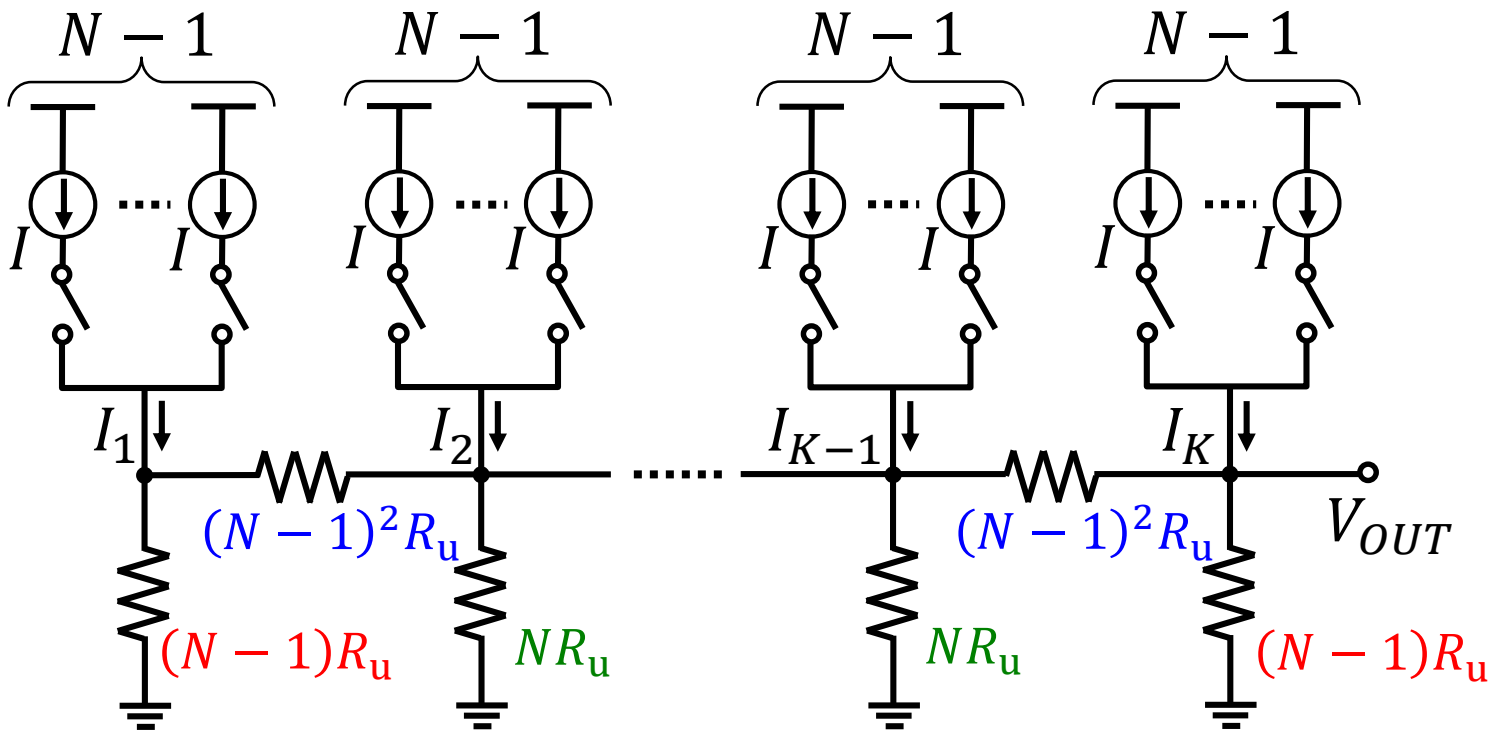
- 電流分割比$N-1:1$を変えない終端抵抗値　R_T

$$R_T = Z - R = \frac{R}{N-1}$$

電流を$N-1:1$に分割する有限抵抗ラダーの R , r , R_T の抵抗比

$$R:r:R_T = (N-1)^2:N:N-1$$

N進抵抗ラダーDACの構成



N：電流分割比　　K：ラダー段数

I_j：j番目ノードに流し込まれる電流

R_u：基準抵抗　　I：単位電流

- N = 2の場合 ⇒ K-bit R-2R DAC

出力電圧と出力ステップ数

- 出力電圧

$$V_{\mathrm{OUT}}(I_1,\cdots,I_K,R_{\mathrm{u}},N,K)=(N-1)R_{\mathrm{u}}\sum_{j=1}^{K}\left(\frac{I_j}{N^{K-j}}\right)$$

- 出力電圧最大値

$$V_{\mathrm{MAX}}(I,R_{\mathrm{u}},N,K)=R_{\mathrm{u}}I\cdot N(N-1)\cdot\left(1-\frac{1}{N^K}\right)$$

- 出力電圧最小ステップ

$$V_{\mathrm{MIN}}(I,R_{\mathrm{u}},N,K)=(N-1)R_{\mathrm{u}}I\cdot\frac{1}{N^{K-1}}$$

- 出力電圧数　N^K-1

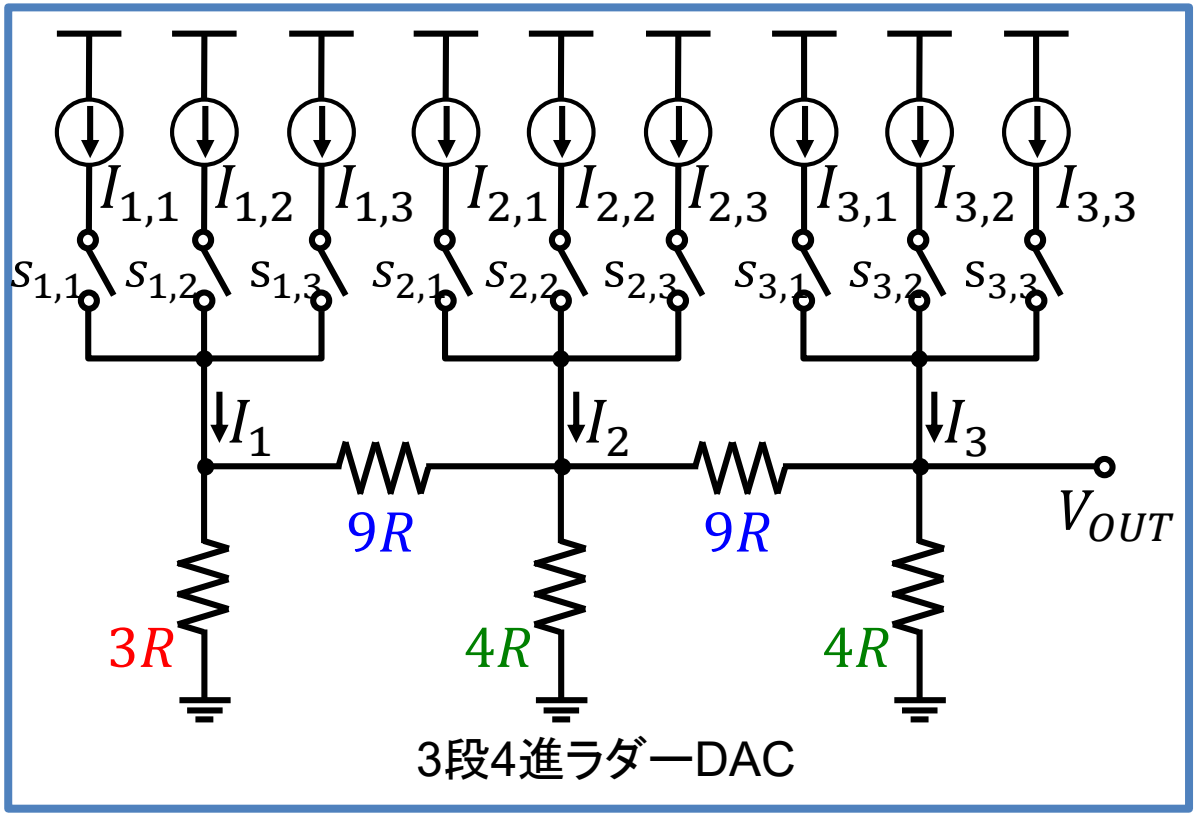
N：電流分割比　　K：ラダー段数
I_j：j番目ノードに流し込まれる電流
R_{u}: 基準抵抗　　I：単位電流

OUTLINE

- 研究背景・目的
- DA変換器の構成
  - R-2R 電流源 DAC
- N進抵抗ラダーDACの構成
  - 抵抗ラダーを用いた電流分割
  - N進抵抗ラダーを用いた構成例
- 異なる分流比を持つ抵抗ラダーの接続
  - 接続の条件・手順
  - DAC構成案と非線形性シミュレーション
- 結論

構成例 $N = 4$, 4進ラダーDAC

- ラダー抵抗比
  $9R : 4R : 3R$
- 電圧ステップ数
  $N^K - 1$
  $= 4^3 - 1$
  $= 63$ 段階
- 出力電圧



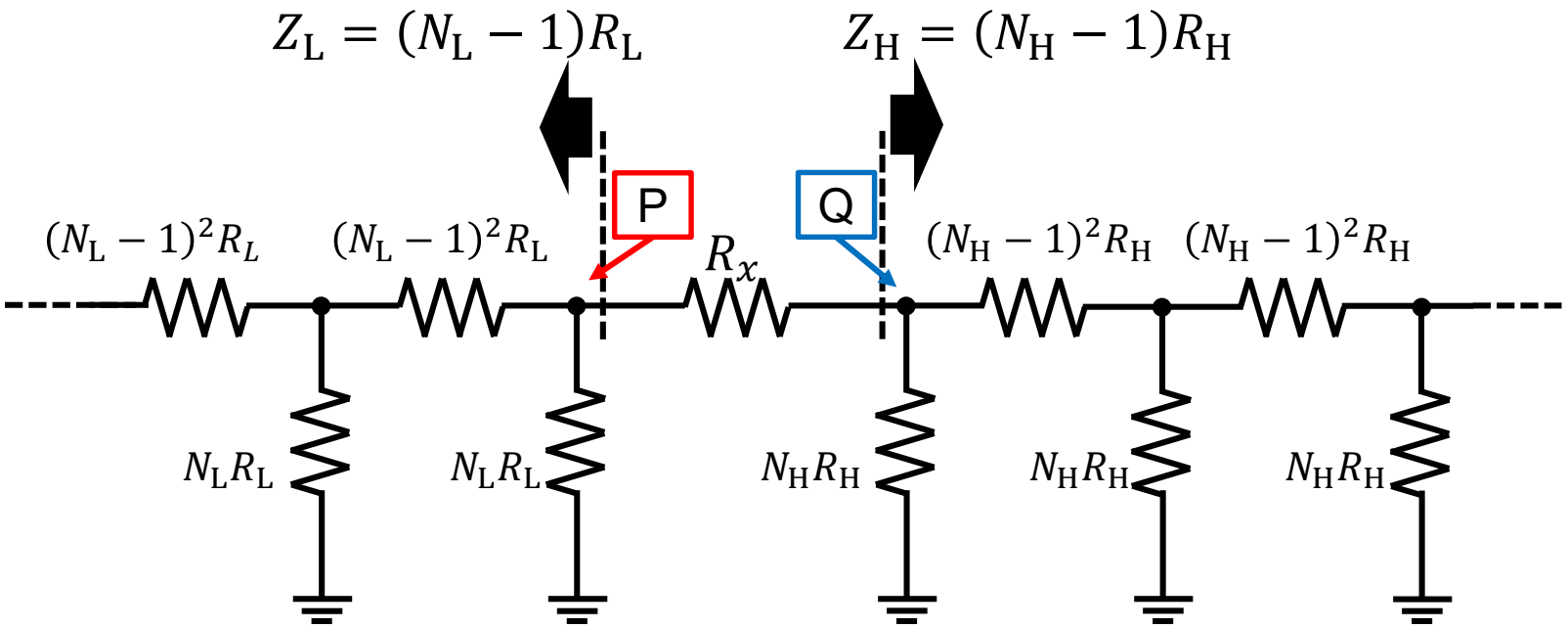
$$V_{OUT}(I_1, I_2, I_3, R_\mathrm{u}) = \underbrace{3R_\mathrm{u}\left(I_3 + \frac{1}{4^1} I_2 + \frac{1}{4^2} I_1\right)}$$

各段のI_j →出力に対して4倍ずつの重みをもつ

OUTLINE

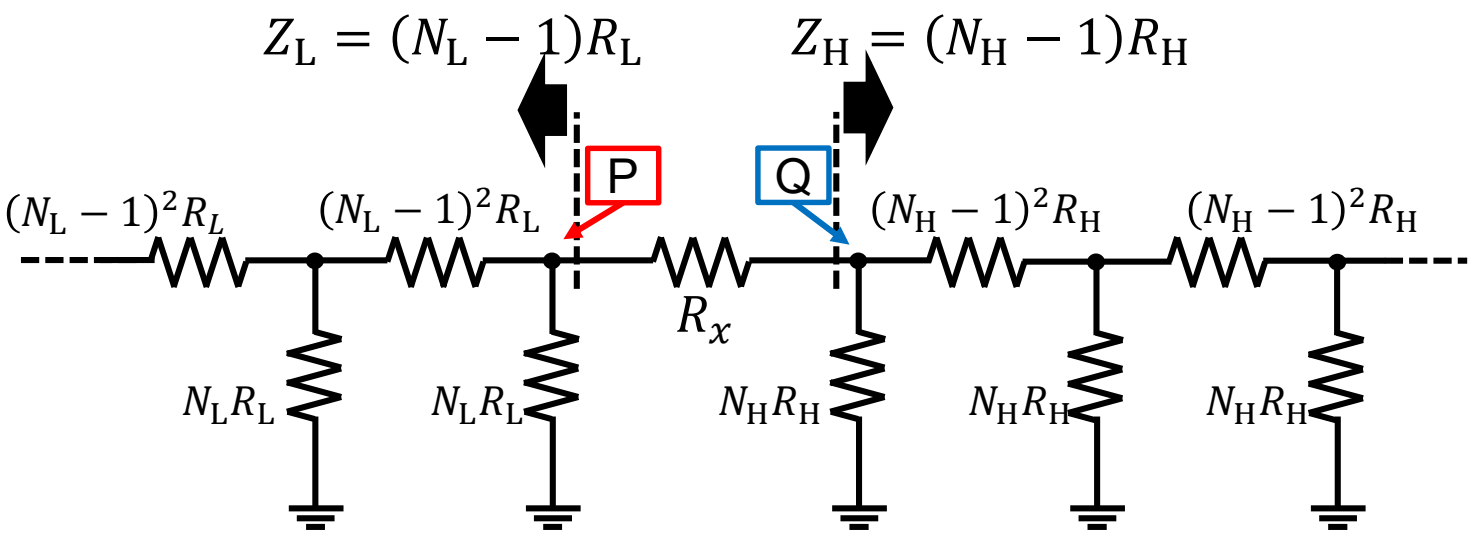
- 研究背景・目的
- DA変換器の構成
  - R-2R 電流源 DAC
- N進抵抗ラダーDACの構成
  - 抵抗ラダーを用いた電流分割
  - N進抵抗ラダーを用いた構成
- 異なる分流比を持つ抵抗ラダーの接続
  - 接続の条件・手順
  - DAC構成案と非線形性シミュレーション
- 結論

異なる分流比の抵抗ラダーの接続



- 接続する抵抗 R_x
- 下位側
  - 分流比 定数 N_{L}
  - 基準抵抗 R_{L}
  - P点から見た抵抗 Z_{L}
- 上位側
  - 分流比 定数 N_{H}
  - 基準抵抗 R_{H}
  - Q点から見た抵抗 Z_{H}

接続の条件



- 接続の条件
  1. P点から右を見込んだ場合、N_{H}進抵抗ラダーの特性。
  2. Q点から左を見込んだ場合、上位側N_{H}進特性が崩れない。

- $$\begin{cases} R_x + Z_{\mathrm{H}} = N_{\mathrm{H}} Z_{\mathrm{L}} \\ R_x + Z_{\mathrm{L}} = N_{\mathrm{H}} Z_{\mathrm{H}} \end{cases}$$

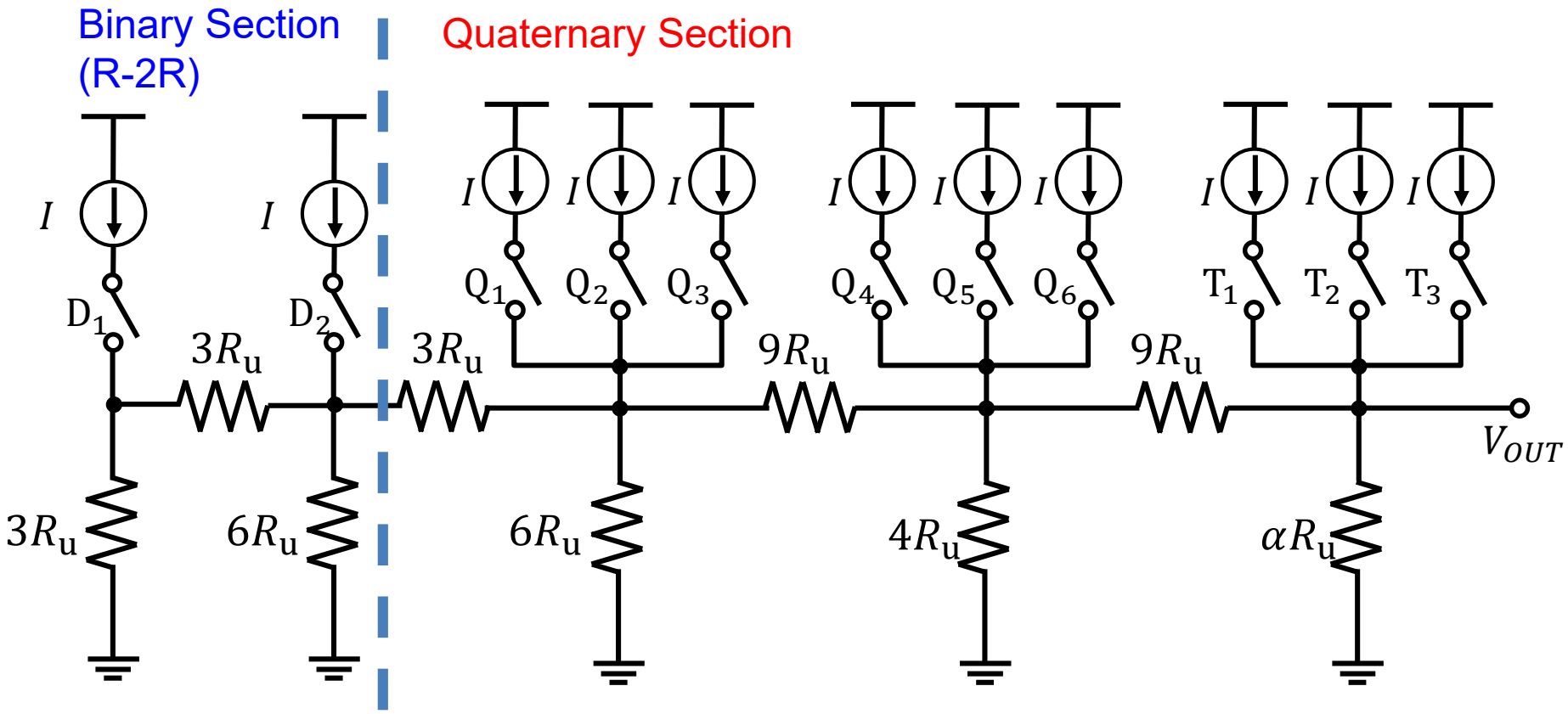
$$R_{\mathrm{H}} = \frac{N_{\mathrm{L}} - 1}{N_{\mathrm{H}} - 1} R_L, \qquad R_x = (N_{\mathrm{H}} - 1)(N_{\mathrm{L}} - 1)R_{\mathrm{L}}$$

OUTLINE

- 研究背景・目的
- DA変換器の構成
  - R-2R 電流源 DAC
- N進抵抗ラダーDACの構成
  - 抵抗ラダーを用いた電流分割
  - N進抵抗ラダーを用いた構成
- 異なる分流比を持つ抵抗ラダーの接続
  - 接続の条件・手順
  - DAC構成案と非線形性シミュレーション
- 結論

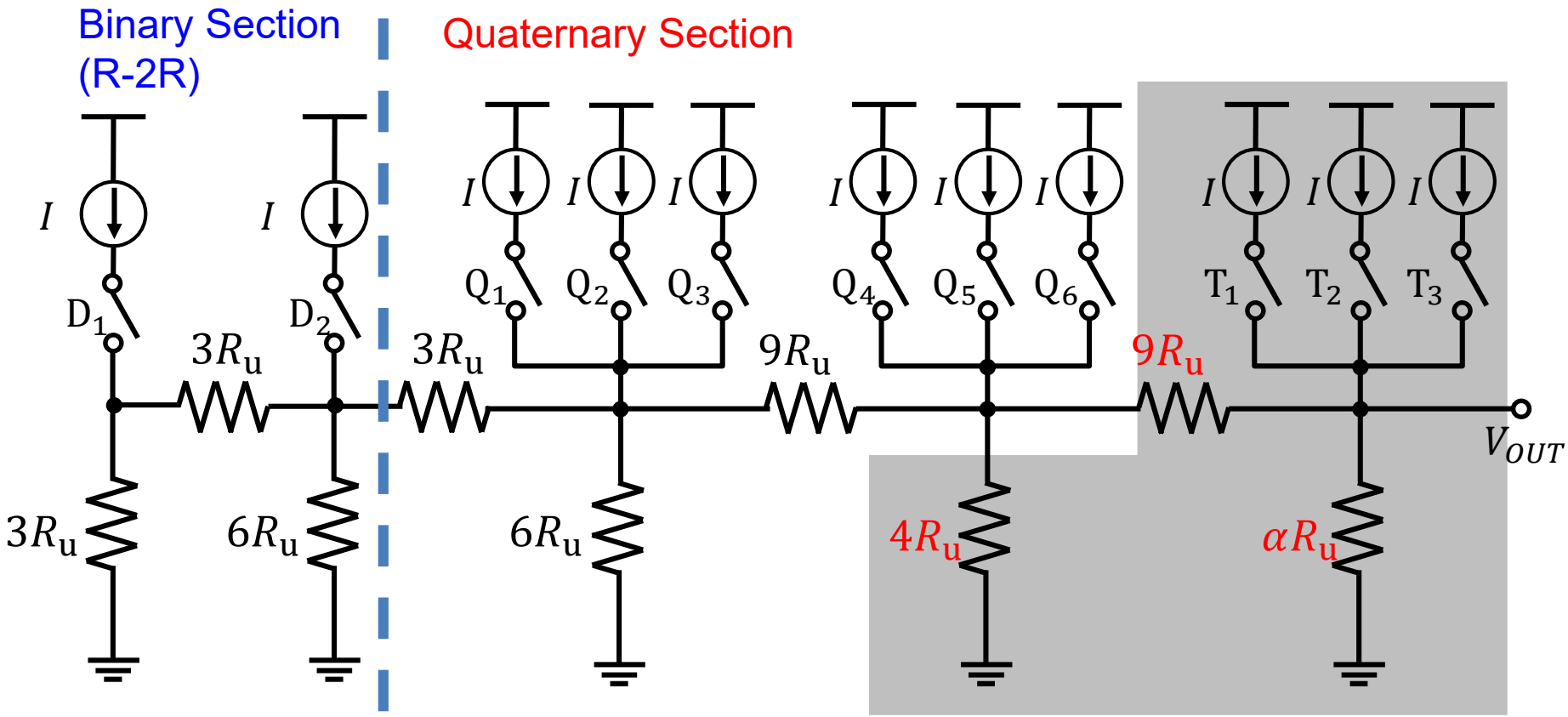
構成例

- 8-bit相当 2進-4進混成 抵抗ラダーDAC
  - $R_{\mathrm{L}} = 3R_{\mathrm{u}}, R_{\mathrm{H}} = R_{\mathrm{u}}, R_x = 9R_{\mathrm{u}}$

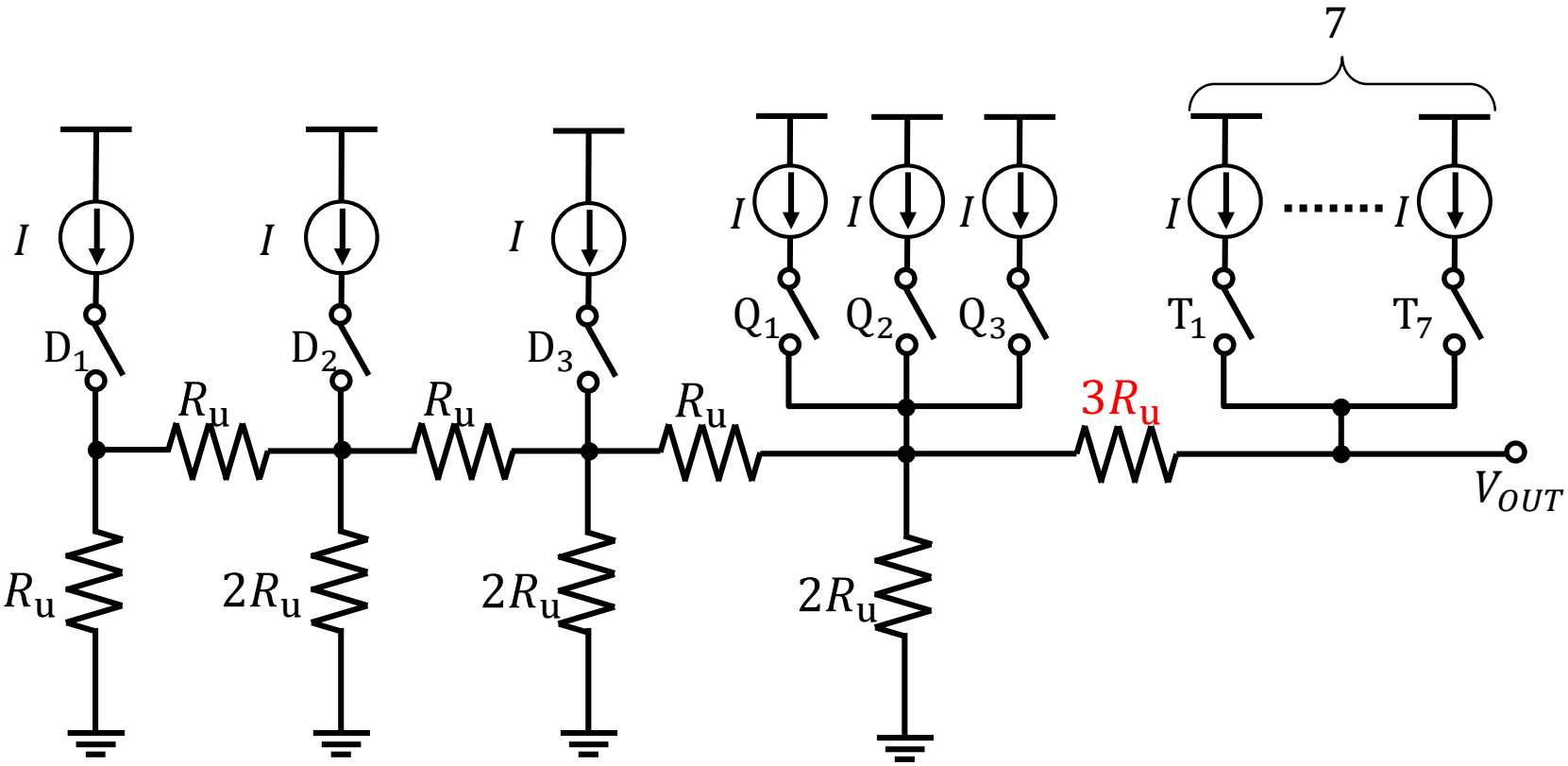


構成例

- 8-bit相当 2進-4進混成 抵抗ラダーDAC
  - $R_{\mathrm{L}} = 3R_{\mathrm{u}}, R_{\mathrm{H}} = R_{\mathrm{u}}, R_x = 9R_{\mathrm{u}}$

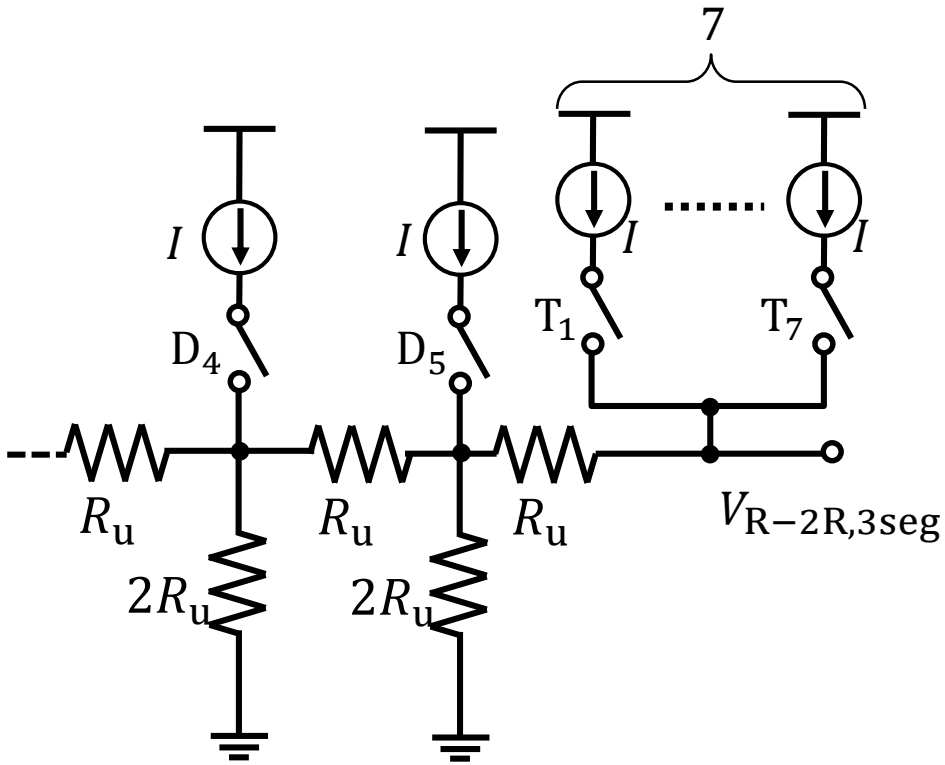


2進-4進-Unaryコード混成 DAC

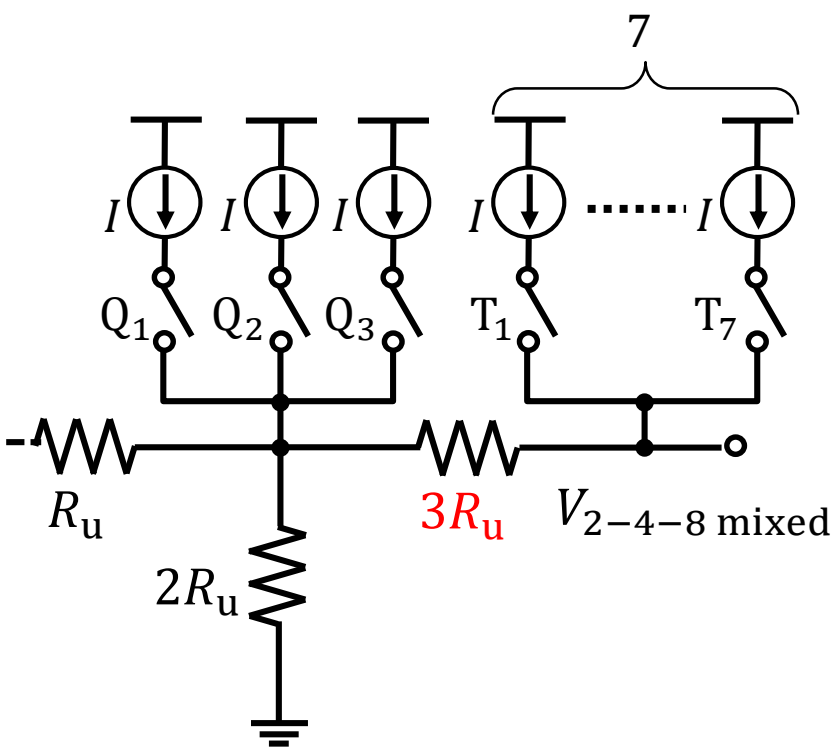


- 下位側はR-2R のまま、上位と4進部分の間を$3R_u$
  - 回路の面積はセグメント化 R-2R と同等
  - 要: D_4, D_5 の2 bit 分の追加の温度計デコーダ

3-bit セグメント化 R-2R DACとの比較



セグメント化R-2R DAC



2進-4進-Unaryコード混成

$$V_{\mathrm{R-2R,3seg.}} = \frac{1}{16} R_{\mathrm{u}} I \cdot \left\{ \sum_{\mathrm{p}=1}^{4} \left(\mathrm{D_p} \cdot 2^{\mathrm{p}-1} \right) + 32 \cdot \sum_{\mathrm{r}=1}^{7} \mathrm{T_r} \right\}$$

$$V_{\mathrm{2-4-8\ mixed}} = \frac{1}{8} R_{\mathrm{u}} I \cdot \left\{ \sum_{\mathrm{p}=1}^{3} \left(\mathrm{D_p} \cdot 2^{\mathrm{p}-1} \right) + \sum_{\mathrm{q}=1}^{3} \mathrm{Q_q} + 32 \cdot \sum_{\mathrm{r}=1}^{7} \mathrm{T_r} \right\}$$

同等の回路面積でゲインが2倍

DNLの計算

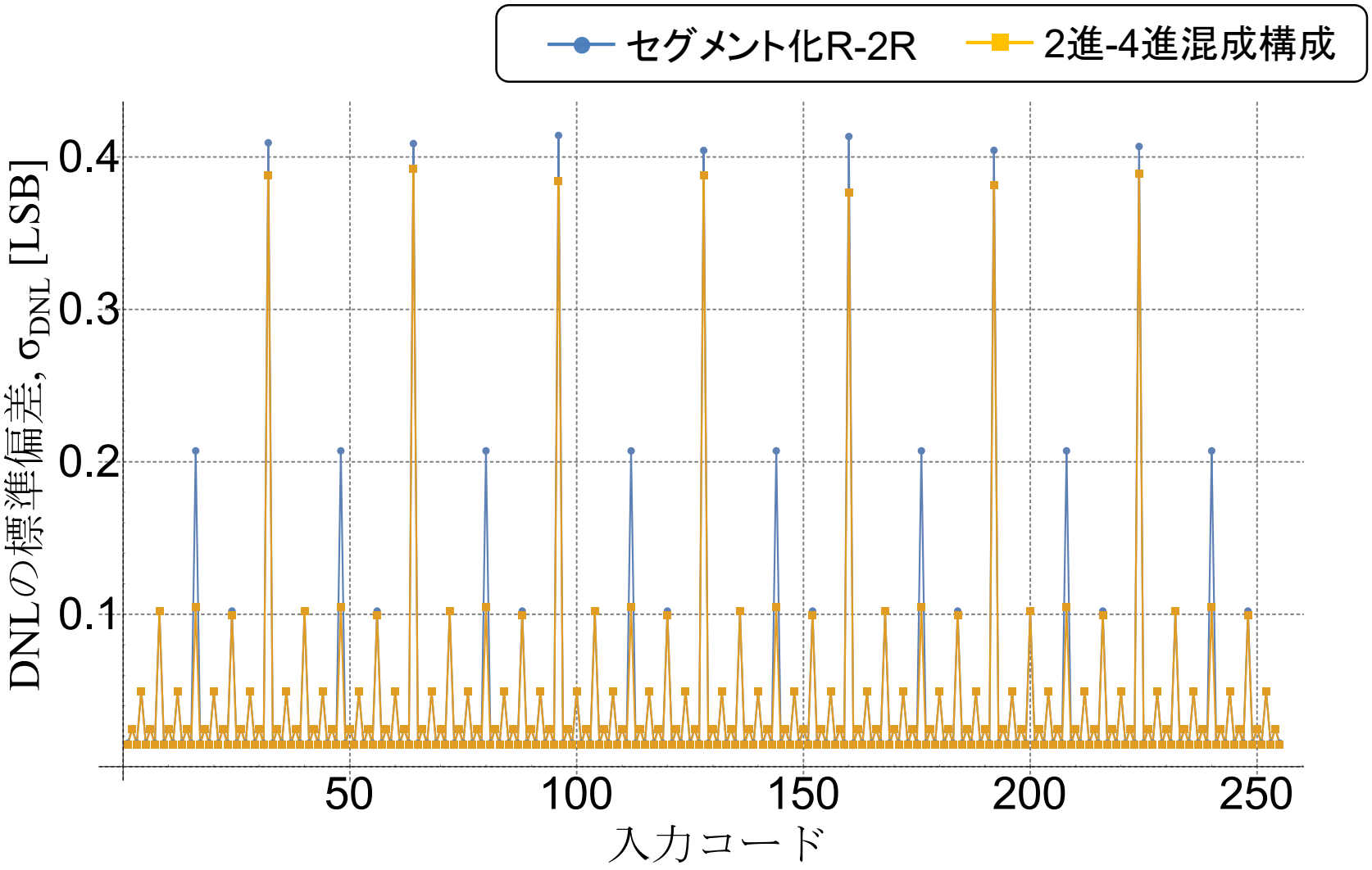
- DNLの定義

$$DNL(n) = \frac{V_{\mathrm{OUT}}(n) - V_{\mathrm{OUT}}(n-1)}{V_{\mathrm{LSB}}} - 1 \quad \text{[LSB]}$$

V_{LSB}：最小桁の変化に相当する電圧

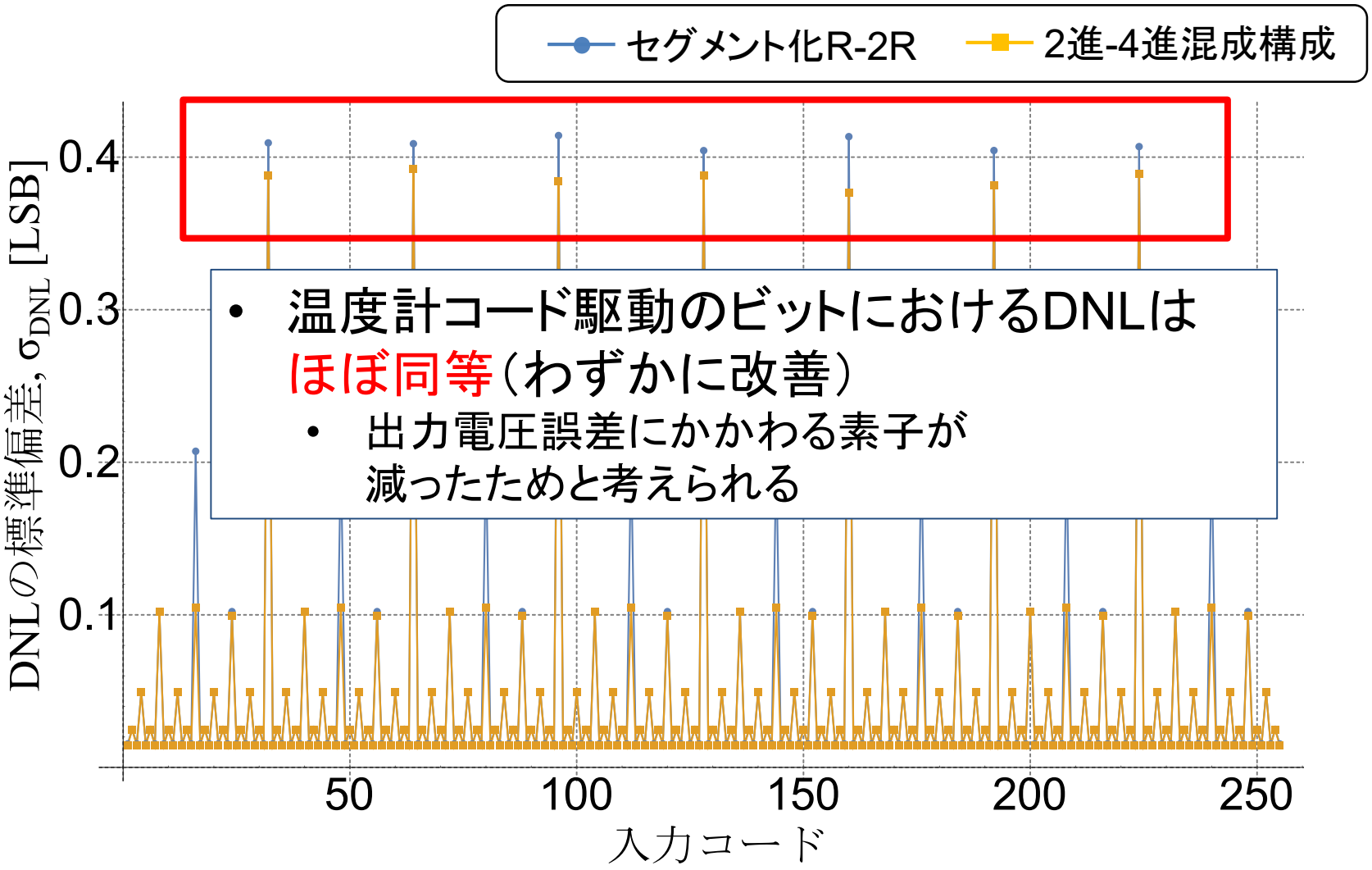
- シミュレーションの条件
  - 抵抗と電流源のみからなる線形回路モデルを仮定した
    モンテカルロシミュレーション
  - シミュレーションセット数　3000回
  - 単位抵抗R_{u}と単位電流Iに
    正規分布のばらつきを仮定
  - 標準偏差σは平均値の1%

シミュレーションによるDNL標準偏差



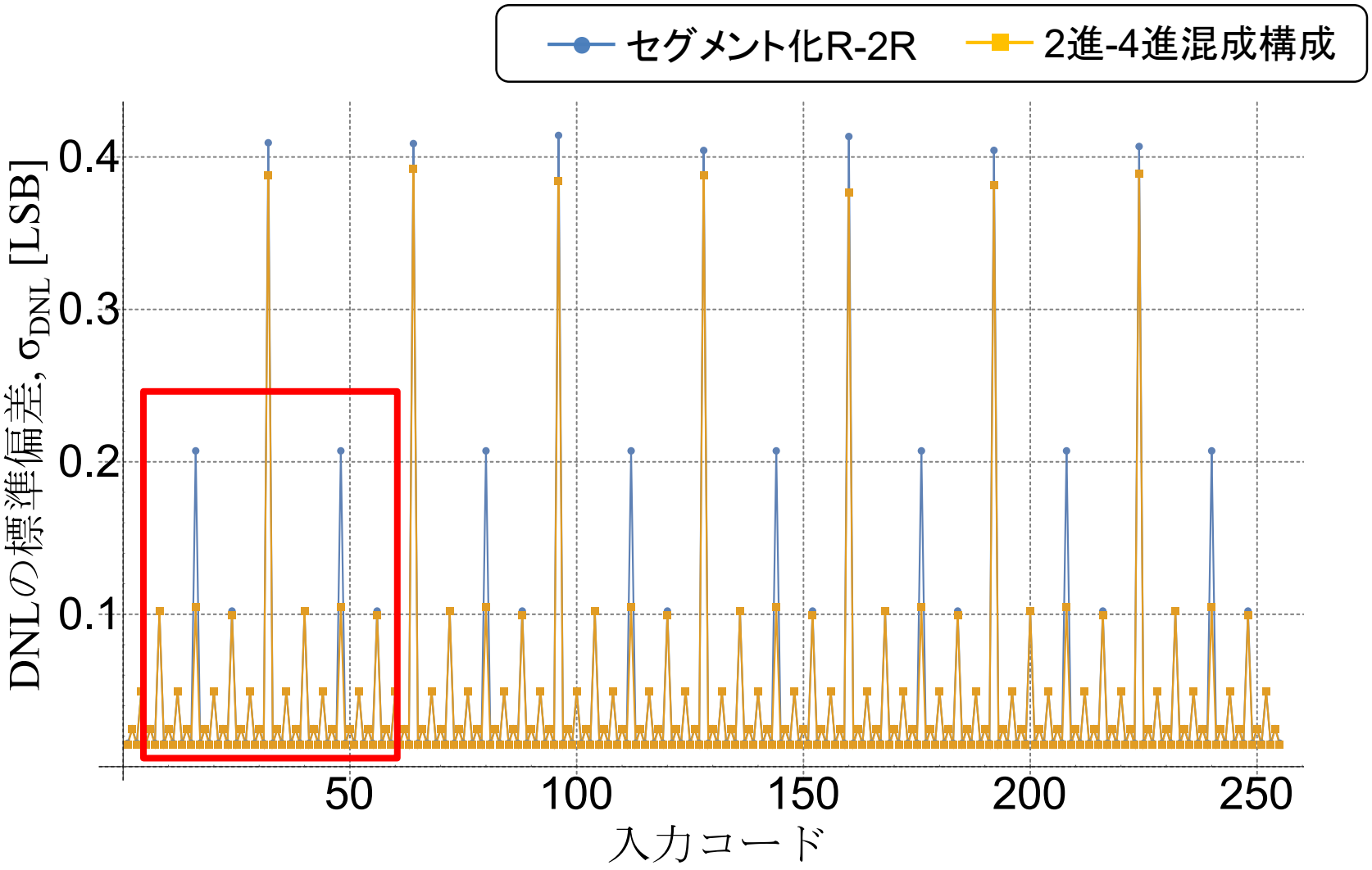
シミュレーションによるDNL標準偏差

- 最大DNLについて



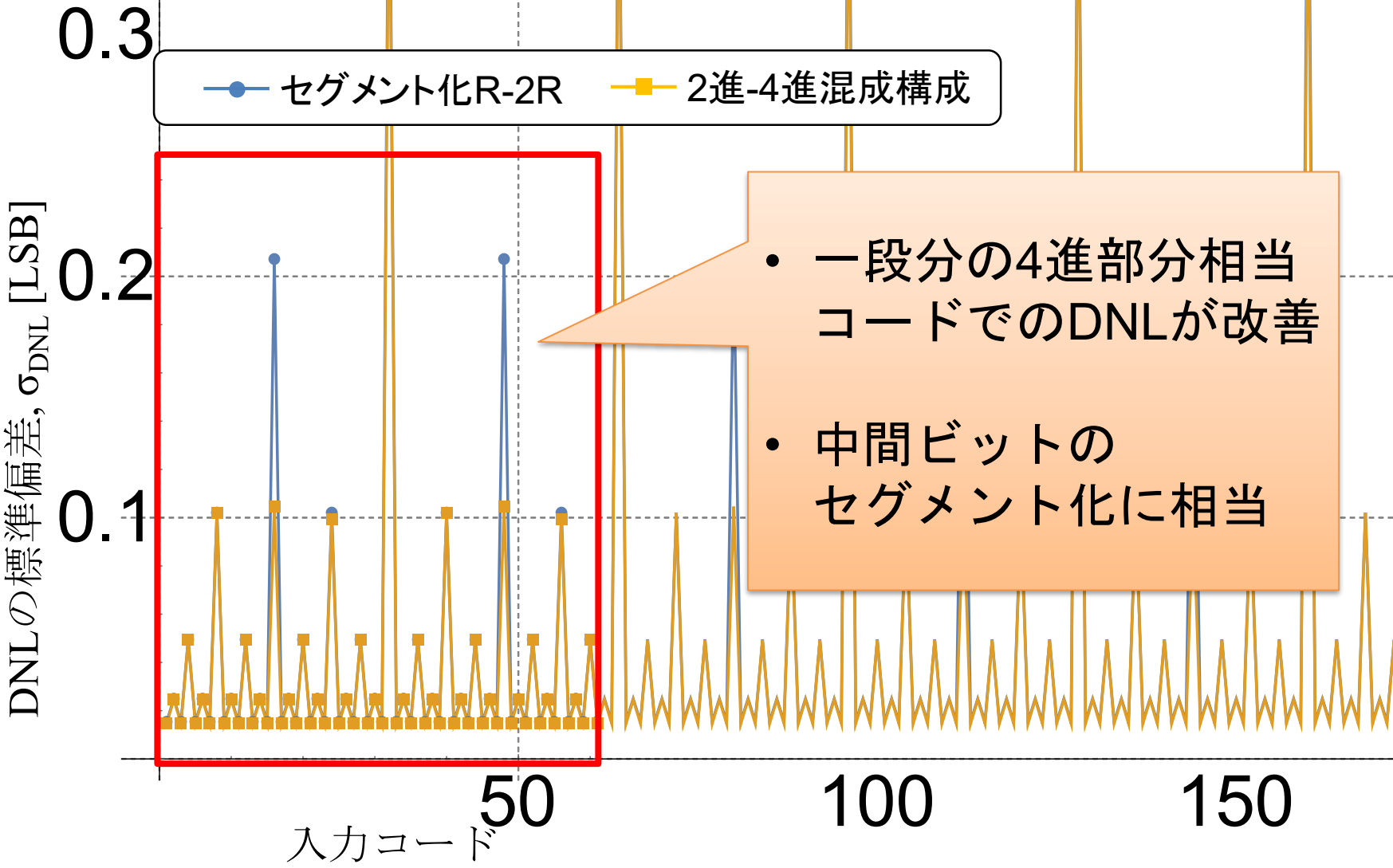
シミュレーションによるDNL標準偏差

- 4進部分を駆動するコードでのDNL



シミュレーションによるDNL標準偏差

- 4進部分を駆動するコードでのDNL



OUTLINE

- 研究背景・目的
- DA変換器の構成
  - R-2R 電流源 DAC
- N進抵抗ラダーDACの構成
  - 抵抗ラダーを用いた電流分割
  - N進抵抗ラダーを用いた構成
- 異なる分流比を持つ抵抗ラダーの接続
  - 接続の条件・手順
  - DAC構成案と非線形性シミュレーション
- **結論**

結論

- まとめ
  - 抵抗ラダーの分流特性をR-2Rラダーから変化させ、
    分流の比が一定でない抵抗ラダーを用いた場合でも
    DACが構成できる
  - 抵抗ラダーを用いたDACにおいても
    上位ビットと下位ビットの中間にセグメント化が可能
  - 2進-4進-温度計コードの混成構成は、
    従来のセグメント化 R-2R DACとほぼ同等面積・DNLで
    ゲインを大きくできる
    - モンテカルロシミュレーションによって、DNL標準偏差を確認し、
      比較を行った
- 今後の検討課題
  - 周辺の回路を含めた設計を行い、その動的な特性を評価

質疑応答

- 回路規模というのは、抵抗・電流源の数という意味か。
  - A. そうです。
    抵抗ラダー部については、上位側・下位側のすべて単位抵抗で構成すると仮定して、チップ内の面積について述べました。

- 実際の電流源はどのように構成するのか。
  - MOSのカレントミラーで構成します。

以降　資料

two-step segmentation

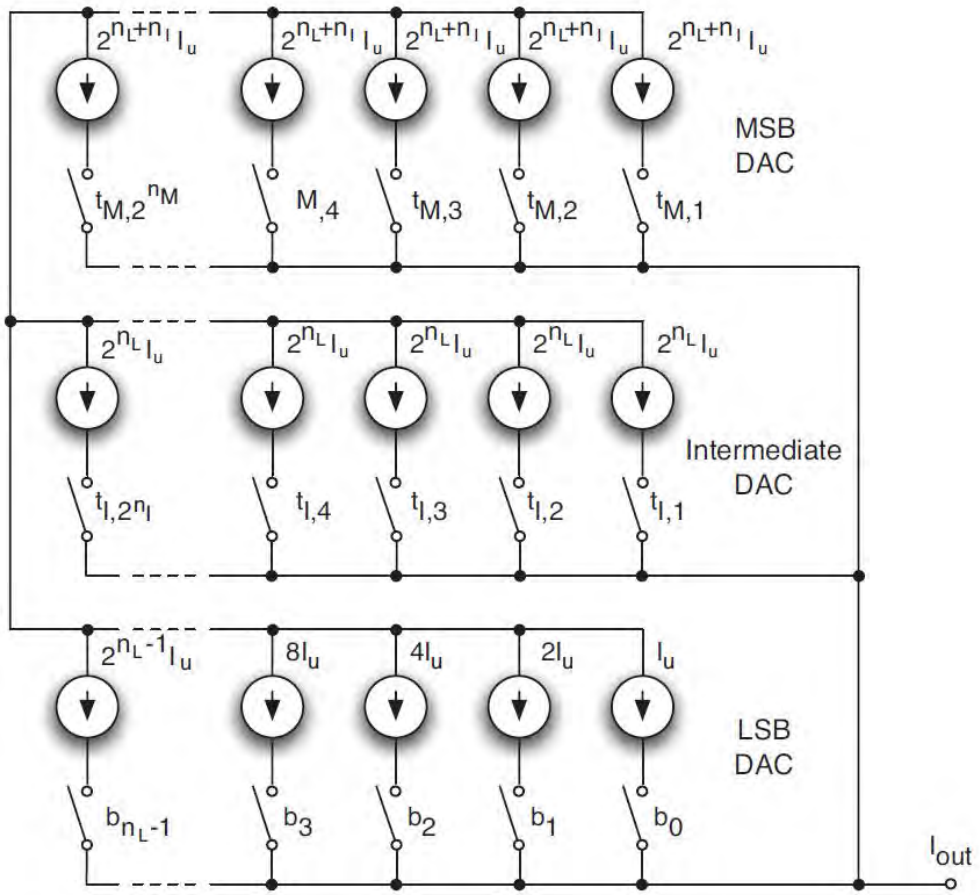


Figure 3.41. Conceptual schematic of a current steering segmented *DAC*.

F. Maloberti, Data Converters, 3.5.5 P127, Springer (2007).

DEM with Nested-Segment Structure

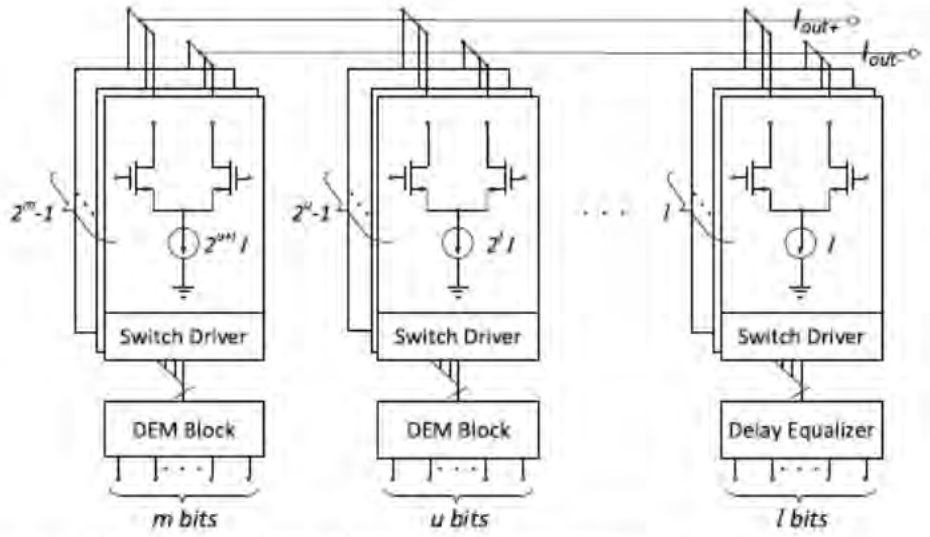


Fig. 1. Conventional DEM DAC with separate-segment structure.

TABLE I

MUX COUNT COMPARISONS FOR DEM DACS WITH DIFFERENT STRUCTURES

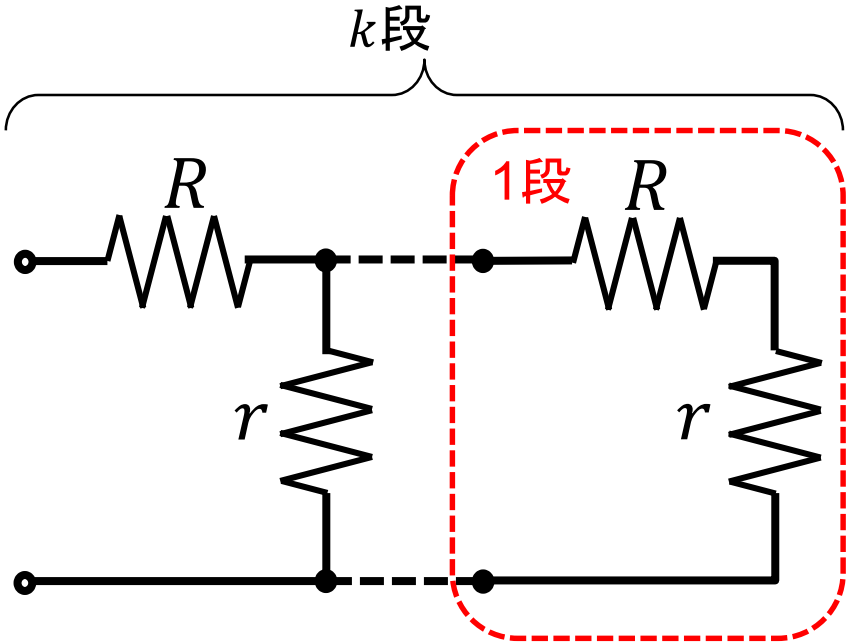
MSB Bits	Seg. Ratio	SFDR (dB)	MUX Count RRBS	MUX Count GRTC	Seg. Ratio	SFDR (dB)	MUX Count RRBS	MUX Count GRTC	Seg. Ratio	SFDR (dB)	MUX Count Proposed
2	2T+10B	66.99	6	6	2T+...+2T	69.21	36	36	2T+2T+8B	73.52	14
3	3T+9B	72.09	21	14	3T+...+3T	74.29	84	56	3T+3T+6B	79.04	38
4	4T+8B	75.90	60	45	4T+4T+4T	77.81	180	135	4T+4T+4B	80.32	94
5	5T+7B	78.71	155	127	5T+5T+2T	79.69	316	260	5T+5T+2B	80.46	222
6	6T+6B	80.27	378	255	6T+6T	80.51	756	510	6T+6T	80.58	510

Wei Mao, “High Dynamic Performance Current-Steering DAC Design WithNested-Segment Structure”, January 2018IEEE Transactions on Very Large Scale Integration (VLSI) Systems PP(99):1-5.

無限抵抗ラダーの合成抵抗の収束

$$Z_k = \frac{\alpha\gamma^k - \beta}{\gamma^k - 1}$$

$$\alpha = \frac{1}{2}\left(R + \sqrt{R^2 + 4rR}\right),$$
$$\beta = \frac{1}{2}\left(R - \sqrt{R^2 + 4rR}\right),$$
$$\gamma = \frac{R + r - \beta}{R + r - \alpha}, 1 < \gamma$$



k段 R-r抵抗ラダー

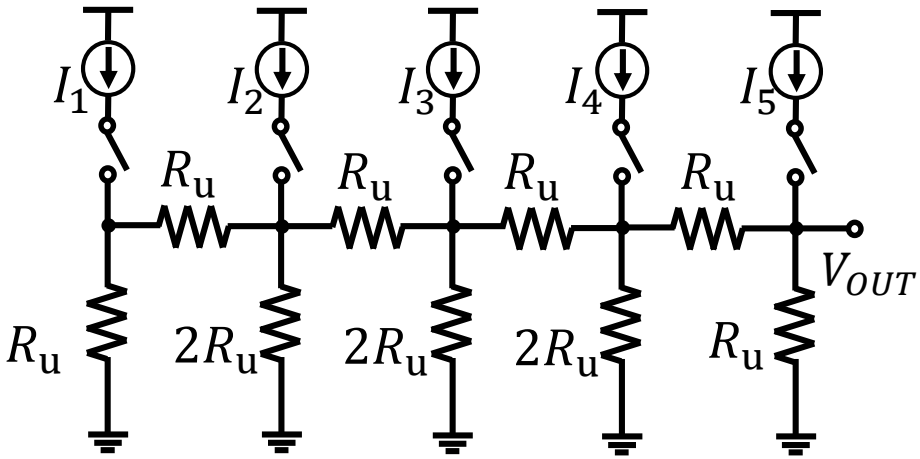
段数kの増加で収束

合成抵抗の収束値:

$$Z = \frac{R}{2} + \frac{\sqrt{R(R + 4r)}}{2}$$

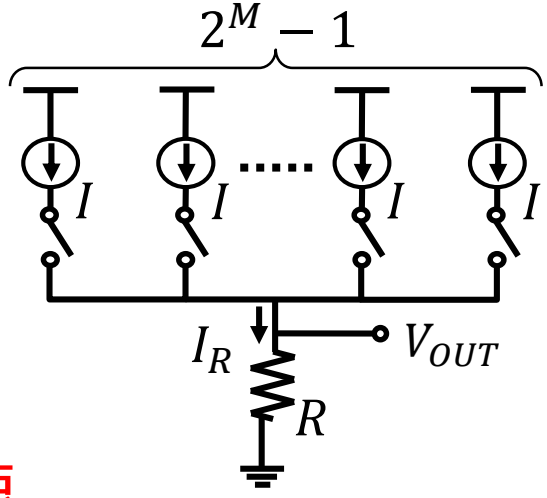
R-2R DAC と Unary DACの比較

R-2R DAC



- 利点
  - 回路構成が単純
    1段の増加＝1bit の増加
  - デコーダが不要
- 欠点
  - 素子誤差でDNLが劣化
  - 上位ビット変化タイミングにより
    グリッチが発生

Unary DAC



- 利点
  - 1LSB変化＝電流源1つの変化
    →グリッチが出にくい・単調性良
- 欠点
  - 要　温度計コードへの変換
  - M-bit DACに2^M-1個電流源
    →デコーダ・回路規模の増大

電流非一定分割抵抗ラダーを用いたDA変換器構成と微分非線形性の解析

平井愛統*（群馬大学），谷本　洋（北見工業大学），源代裕治（群馬大学），
山本修平，桑名杏奈，小林春夫（群馬大学）

Digital-to-Analog Converter Configuration based on Non-uniform Current Division Resistive-Ladder and its Nonlinearity Analysis

Manato Hirai*(Gunma University), Hiroshi Tanimoto (Kitami Institute of Technology), Yuji Gendai, Shuhei Yamamoto, Anna Kuwana, Haruo Kobayashi (Gunma University)

キーワード： デジタルアナログ変換器，微分非線形性、抵抗ラダー、モンテカルロ法、非２進重みづけ
Keywords: Digital-to-Analog Converter, Differential Nonlinearity, Resistive Ladder, Monte- Carlo Simulation, Non-binary weighting

1. はじめに

近年の社会において、コンピュータの普及によって信号をデジタルで処理するデジタル信号処理が広く普及している。信号処理を行った出力を人間が知覚したり物理的な実体を伴わせて伝送したりするためには、デジタル信号をアナログ信号に変換する必要があり、デジタル信号とアナログ信号のインターフェースとしてのデジタルアナログ変換器（Digital-to-Analog Converters, DACs）の高性能化は重要である。

ところで、抵抗ラダー回路は、デジタルアナログ変換器、アナログデジタル変換器、アナログ空間フィルタなどの内部回路として用いられる[1]-[4]。その中で、R-2R 抵抗ラダーは主に DAC の内部回路として用いられ、そのシンプルな構成から R-2R DAC は広く普及している。

本稿では、抵抗ラダーの電流分割特性に着目し、R-2R 抵抗ラダーから電流分割特性を変化させた抵抗ラダーで DAC を構成することを提案する。

2. R-2R DAC の構成

〈2･1〉R-2R DAC の概要

R-2R ラダーは、Rと$2R$ の 2 値の抵抗を梯子状に接続した回路で、ラダーの各ノードから見た合成抵抗が一定である性質を持つ。この性質によって、各スイッチの状態を入力コードに応じて変化させ、2 進に重みづけられた出力を得ることができる。R-2R 抵抗ラダーを用いて電流源から流れる電流を重みづけし、出力を得る構成を図 1 に示す[2]。

図 1 において、回路中の電流は出力に対して 2 進に重みづけられる。この回路の原理は、図 2 のようにノートンの定理を用いて、最下位ビット（Least Significant Bit, LSB）側から順に電流源と抵抗を等価回路に置き換えることで説明できる。この説明から、出力端子と接地電位との間に抵抗を接続した場合にも、単にフルスケールを変化させるだけであり、回路内の電流を 2 進に重みづけする性質が変わることはない。出力のノードに接続される抵抗の抵抗は DAC としての動作を変えることはないため、出力終端の抵抗を除いて回路の動作を考察することもできるといえる。

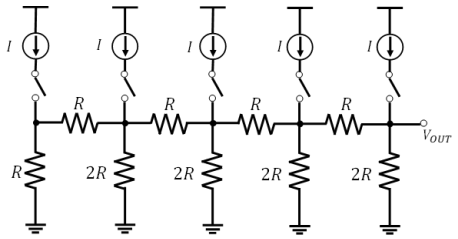


図 1　5-bit R-2R 電流源 DAC
Fig. 1　5-bit R-2R current-steering DAC

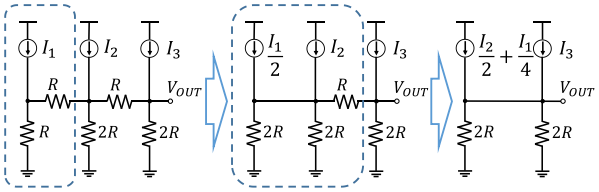


図 2　3-bit R-2R 電流源 DAC の等価回路
Fig. 2　Equivalent circuits of 3-bit current-steering DAC with an R-2R ladder

〈2･2〉R-2R DAC の利点と欠点

R-2R ラダーと電流源を用いた DAC 構成の利点は、他の方式に対して比較的高速である点、R-2R ラダーを用いて電流の 2 進重みづけを行いバイナリコードで動作させるため、デコーダ回路が不要である点が挙げられる[2]。

R-2R DAC の欠点としては、素子のマッチングが悪い場合に、入力コードの中央において微分非線形性（Differential Nonlinearity, DNL）が大きくなることや大きなグリッチが発生する点がある。

〈2･3〉セグメント化 R-2R DAC

単純な 2 進重みづけ（R-2R）構成の場合、DNL を小さくして単調性を保証し、発生するグリッチを小さくするためには、素子マッチングへの要求が厳しくなる。そこで、分解能が高い場合には、上位ビット側を温度計コード（ユナリコード）による駆動、下位ビット側をバイナリコードによる駆動に分けるセグメント化構成（Segmented Architecture）が用いられる。例として、図 3 に上位 3 bit を温度計コードによる駆動のユナリ型、下位 5 bit を R-2R ラダーを用いた電流重みづけで構成した 8 bit セグメント化 DAC を示す。この構成では、温度計コードによって操作されるビット数が増えるほど、バイナリコードを温度計コードに変換するデコーダ回路の規模が増大する。

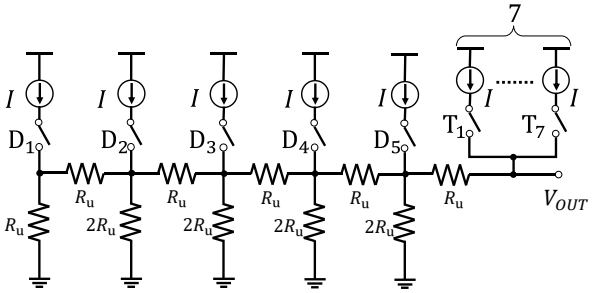


図 3　8-bit セグメント化 R-2R DAC

Fig. 3　8-bit partially segmented R-2R current-steering DAC

3.　分流比が一定でない抵抗ラダーを用いた DAC 構成

〈3･1〉非 2 進電流分割抵抗ラダーを用いた DAC 構成

抵抗ラダーを用いた DA 変換器の構成のために、DA 変換器等に用いられる有限の素子からなる抵抗ラダーを無限抵抗ラダーからの切り出しとみなして、抵抗ラダーの分流比に着目して検討を行った。

図 4 に抵抗Rとrからなる無限抵抗ラダーを示す。図 4 におけるZのように、あるノードから右もしくは左を見込んだ時の合成抵抗は次式であらわされる。

$$Z = \frac{R + \sqrt{R^2 + 4rR}}{2} \tag{1}$$

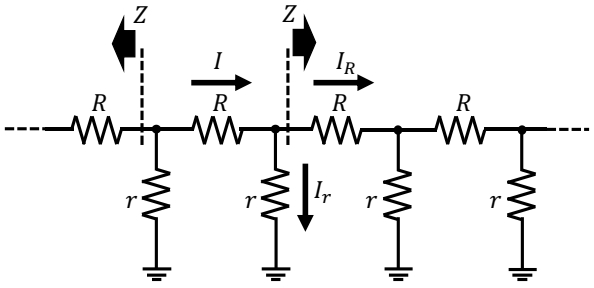


図 4　無限抵抗ラダーによる電流等比分割

Fig. 4　Current division with an infinite resistive ladder

図 4 の無限抵抗ラダーで、あるノードからその隣のノードに流れる電流Iは、抵抗Rからさらに隣のノードに流れる電流I_Rと、抵抗rを通り接地電位に流れる電流I_rに分割される。この時の電流の比はrとZの比によって決まる。Nを 2 以上の整数とし、電流を比$I_r : I_R = (N-1) : 1$ に分割する場合は、Rとrの比は以下になる。

$$\frac{R}{r} = \frac{(N-1)^2}{N} \tag{2}$$

現実の回路でこの分流比を変えずに有限の長さの抵抗ラダーを実現するためには、終端に抵抗を接続する必要がある。図 4 の無限抵抗ラダーと分流比が変わらない有限抵抗ラダーを図 5 に示す。また、この時の終端抵抗R_Tは次式であらわされる。

$$R_\mathrm{T} = Z - R = \frac{R}{N-1} \tag{3}$$

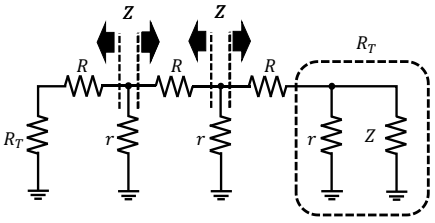


図 5　抵抗ラダーの終端

Fig. 5　The termination of the resistor ladder

(2)式と(3)式から、比$I_r : I_R = (N-1) : 1$に電流を分割する有限抵抗ラダーの抵抗比は次式であらわされる。

$$R : r : R_\mathrm{T} = (N-1)^2 : N : \ N-1 \tag{4}$$

この有限抵抗ラダーは整数比の抵抗によって実現できる。

電流等比分割の性質を用いると、ラダーの各段に$N-1$個ずつの単位電流源を接続することで、出力電圧範囲を等間隔に分割した出力を得ることができる。これを N進抵抗ラダーDAC と呼ぶことにし、図 6 にその回路図を示す。

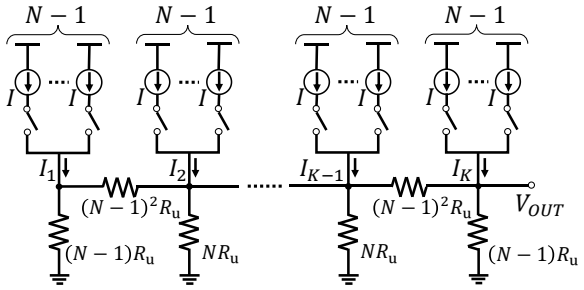


図 6　K段N進抵抗ラダー型 DAC

Fig. 6　*K*-stage, *N*-ary resistive ladder DAC

抵抗の比が(4)式を満たしている場合、図 6 におけるV_OUTは次式であらわされる。ここで、I_1からI_Kはラダーの各段に流し込む電流値、R_uは基準抵抗値、Nは電流分割比を決める定数、Kはラダーの段数である。

$$V_{OUT}(I_1, \cdots, I_K, R_\mathrm{u}, N, K) = R_\mathrm{u}\frac{N(N-1)}{N+1}\sum_{k=1}^{K}\left(\frac{I_k}{N^{K-k}}\right) \tag{5}$$

(5)式では、ラダーの各段に入力される電流が、出力側に近づくにしたがってN倍ずつの重みをもつようになっている。ラダーの各段に流し込む電流I_kを変化させることで、$N^K -$

1段の等間隔な出力電圧を得ることができるため、DAC として動作させることができる。

例として、Nを 4 とし電流が$I_r : I_R = 3 : 1$の比に分割されるようにした場合、ラダーの各段に 3 個の単位電流源とスイッチを接続することで、DAC としての動作が可能である。図 7 にこの方法による 4 段の 4 進構成の抵抗ラダーDAC を示す。各段に流し込まれる電流I_kは、出力に対して 4 倍ずつ重みづけられる。スイッチの制御により、255 $(= 4^4 - 1)$ の等間隔な出力電圧を得ることができる。

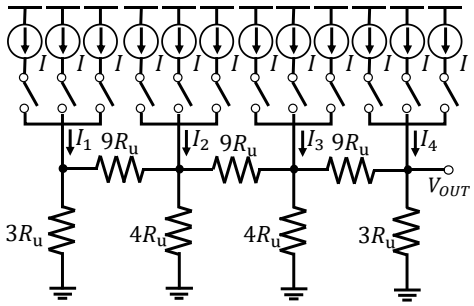


図 7　4 段 4 進抵抗ラダーDAC

Fig. 7　4-stage quaternary resistor ladder DAC

〈3･2〉非 2 進電流分割抵抗ラダーの接続

3.1 で示した N 進抵抗ラダーDAC において、分流比を定めている定数Nが 2 の冪の場合、出力電圧のステップ数が R-2R ラダーを用いた場合と一致する。そこで、Nが 2 の冪である抵抗ラダーどうしを、各部における分流比が変化しないよう接続した抵抗ラダーを用いた DA 変換器構成の検討を行った。

図 8 のように、上位側の抵抗ラダーの分流比定数NをN_H、基準抵抗をR_Hとし、下位側の抵抗ラダーについては同様にN_L、R_Lとし、抵抗R_xで接続するものとする。また、ノード P から左を見込んだ抵抗をZ_L、ノード Q から右を見込んだ抵抗をZ_Hとおく。

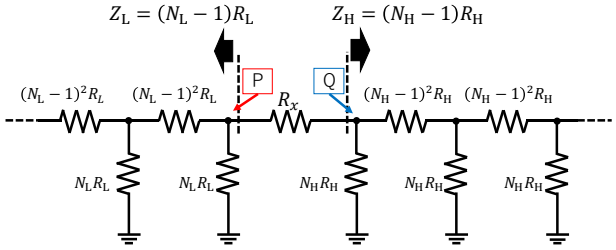


図 8　異なる分流比を持つ抵抗ラダーの接続

Fig. 8　Connection of resistive ladder

この時、接続後のN_L進部分とN_H進部分それぞれでの分流特性が変わらないために、以下の二つの条件を満たすように抵抗R_x、R_L、R_Hの関係を定める。

1. P 点から右を見込んだ場合、ラダーの接続部分はN_H進抵抗ラダーの特性を持つ。
2. Q 点から左を見込んだ場合、上位側の抵抗ラダーのN_H進特性が崩れない。つまり、ノード Q の左右にのびる抵抗ラダーの抵抗が同じである。

これらから、R_x、R_L、R_Hの関係は、次式で表すことができる。

$$\begin{cases} R_x + Z_\mathrm{H} = N_\mathrm{H} Z_\mathrm{L} \\ R_x + Z_\mathrm{L} = N_\mathrm{H} Z_\mathrm{H} \end{cases} \tag{6}$$

$$\leftrightarrow R_\mathrm{H} = \frac{N_\mathrm{L} - 1}{N_\mathrm{H} - 1} R_L, \qquad R_x = (N_\mathrm{H} - 1)(N_\mathrm{L} - 1) R_\mathrm{L} \tag{7}$$

この結果を 2 進ラダー（R-2R ラダー）と 4 進ラダーの合成に適用した場合、R_x、R_L、R_Hは以下の比である。

$$R_\mathrm{L} = 3R_\mathrm{u}, \qquad R_\mathrm{H} = \frac{1}{3} R_\mathrm{L} = R_\mathrm{u}, \qquad R_x = 3R_\mathrm{L} = 9R_\mathrm{u} \tag{8}$$

このようにして求めた抵抗ラダーを用いて構成した 2 進-4 進混成 抵抗ラダーDAC を図 9 に示す。ここで、出力側終端の抵抗αR_uは抵抗ラダー部分の電流分割にはかかわらないため、任意の値としてよい。

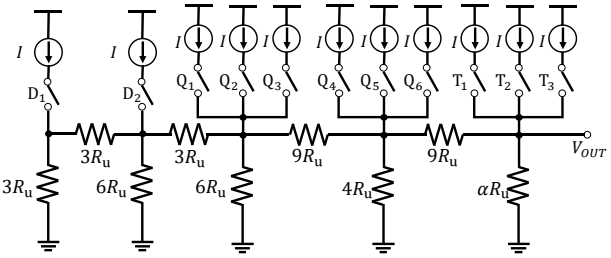


図 9　8-bit 相当 2 進-4 進混成 抵抗ラダーDAC

Fig. 9　8-bit R-2R and 9R-4R connected resistive ladder DAC

〈3･3〉2 進-4 進-温度計コード混成 抵抗ラダーDAC

図 9 に示した 2 進- 4 進混成 抵抗ラダーDAC は抵抗の比が複雑化している。2.3 にて述べた R-2R セグメント化電流源 DAC の上位側と下位側の間に 4 進部分を一段だけ挟む場合、図 9 における 4 進部分の $4R_\mathrm{u}$を除くことができるため、抵抗の比を単純化できる。この場合の例として、2 進 4 進混成セグメント化 DAC を図 10 に示す。上位 3bit を温度計コードによる駆動にした 8-bit R-2R 電流源 DAC について、R-2R 部分と温度計コード駆動部分との間に、4 進抵抗ラダーの特性を持つ部分を挿入している。また、2.1 から、抵抗ラダー出力側終端の抵抗を除いている。挿入された 4 進部は下位から 4 bit 目、5 bit 目の入力を温度計コードに変換して駆動する必要がある。そのため、4 進部分を駆動するための追加の温度計コードデコーダが必要である。

図 3 に示した上位 3 bit を温度計コードによる駆動にした 8-bit R-2R DAC から出力終端の抵抗を除いた回路（図 11）の出力電圧$V_{\mathrm{R-2R,3seg.}}$と、図 10 に示した 4 進部混成構成 8-bit DAC の出力電圧$V_{\mathrm{2-4-8\ mixed}}$は、次式で表せる。

$$V_{\mathrm{R-2R,3seg.}}(\mathrm{D}_1, \dots, \mathrm{D}_4, \mathrm{T}_1, \dots, \mathrm{T}_7, R_\mathrm{u}, I)$$

$$= \frac{1}{16} R_\mathrm{u} I \cdot \left\{ \sum_{\mathrm{p}=1}^{4} \left(\mathrm{D}_\mathrm{p} \cdot 2^{\mathrm{p}-1} \right) + 32 \cdot \sum_{\mathrm{r}=1}^{7} \mathrm{T}_\mathrm{r} \right\} \tag{9}$$

$$V_{\mathrm{2-4-8\ mixed}}(\mathrm{D}_1, \mathrm{D}_2, \mathrm{D}_3, \mathrm{Q}_1, \mathrm{Q}_2, \mathrm{Q}_3, \mathrm{T}_1, \dots, \mathrm{T}_7, R_\mathrm{u}, I)$$

$$= \frac{1}{8} R_\mathrm{u} I \cdot \left\{ \sum_{\mathrm{p}=1}^{3} \left(\mathrm{D}_\mathrm{p} \cdot 2^{\mathrm{p}-1} \right) + \sum_{\mathrm{q}=1}^{3} \mathrm{Q}_\mathrm{q} + 32 \cdot \sum_{\mathrm{r}=1}^{7} \mathrm{T}_\mathrm{r} \right\} \tag{10}$$

D_p、T_r、Q_qはそれぞれ、下位ビットによる駆動、温度計コー

ドによる駆動、挿入した 4 進部の温度計コードによる駆動で切り替わるスイッチを表し、0 または 1 である。4 進部混成構成 8-bit DAC は、セグメント化 R-2R DAC とほぼ同等のアナログ部面積であり、ゲインは 2 倍になっている。

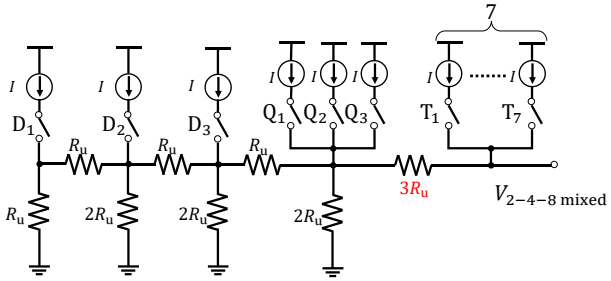


図 10　8-bit 2 進-4 進-温度計コード駆動混成 DAC

Fig. 10　8-bit binary-quaternary-unary connected resistor ladder DAC

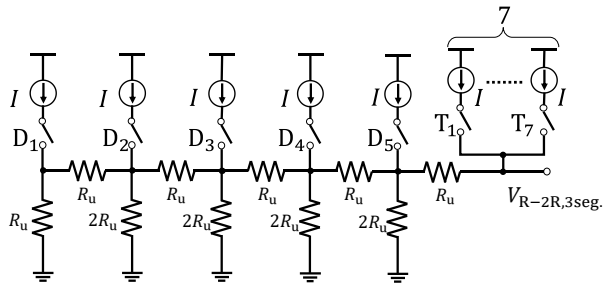


図 11　終端抵抗を除いた 8-bit セグメント化 R-2R DAC

Fig. 11　8-bit segmented R-2R current-steering DAC without output termination

4.　シミュレーションによる微分非線形性評価

〈4･1〉微分非線形性

DAC の DNL は隣接コード入力時の出力電圧差から計算され、(10)式で定義される[1]。ここで、$V_{\mathrm{OUT}}(n)$はコードnでの出力電圧、V_{LSB}は最小の出力電圧の理想値としている。

$$DNL(n)=\frac{V_{\mathrm{OUT}}(n)-V_{\mathrm{OUT}}(n-1)}{V_{\mathrm{LSB}}}-1 \qquad (11)$$

〈4･2〉シミュレーション条件

図 10 に示した 8-bit 2 進-4 進-温度計コード駆動混成 抵抗ラダーDAC について、モンテカルロシミュレーションによって電流と抵抗に誤差がある場合の出力電圧を求め、DNL の標準偏差を求めた。比較として、図 3 に示した 8-bit セグメント化 R-2R 電流源 DAC から出力終端の抵抗をのぞいた回路（図 11）についても同様の条件でシミュレーションを行った。

シミュレーションの条件は以下である。

- 単位抵抗と単位電流は、標準偏差が平均値の 1%である正規分布をする。
- 平均値が定数a倍の抵抗は、標準偏差が$\sqrt{a}$倍された正規分布をする。
- シミュレーションセット数は 3000
- 各セットについて(10)式から DNL を求める。

〈4･3〉シミュレーション結果による DNL 標準偏差

4.2 の条件で行ったシミュレーションから、DNL の標準偏差を計算した結果を、図 12 に示す。

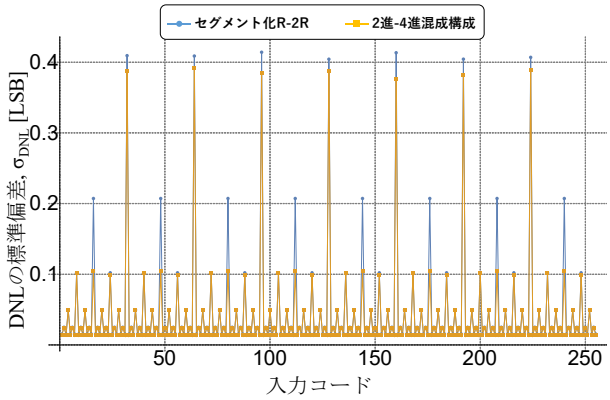


図 12　シミュレーションから計算した DNL 標準偏差

Fig. 12　standard deviation of DNL from simulation results

上位 3 bit にあたる温度計コード駆動部分が変化するコードでの DNL は、DNL 標準偏差がほぼ同等である。2 進-4 進-温度計コード駆動混成 8-bit 電流源 DAC（以下、混成構成）での DNL 標準偏差がわずかに小さくなっているが、これは抵抗ラダーの段数が一段減り、出力の誤差にかかわる要素が減少したためと考えられる。

また、$\sigma_{\mathrm{DNL}}(16)$など、混成構成における 4 進化部分のコードが変化したときの DNL 標準偏差については、4 進抵抗ラダーDAC の DNL 劣化特性が表れており、セグメント化 R-2R DAC の$\sigma_{\mathrm{DNL}}(16)$より小さくなっている。

5.　まとめ

本稿では、R-2R 抵抗ラダーとは異なり分流特性が一定でない抵抗ラダーを用いた DAC 構成を示した。R-2R 抵抗ラダーによるバイナリコード駆動部分と温度計コードで駆動されるユナリ部分との間にさらにセグメント化を行うことで、ほぼ同一の面積で 2 倍のゲインを得られる構成を提案した。また、線形回路モデルを仮定したモンテカルロシミュレーションでは、提案構成を用いた場合の素子のランダムばらつきによる微分非線形性劣化の度合いは、従来構成のセグメント化 R-2R DAC からわずかに改善した。今後の追加検討事項として、周辺回路を含めた DAC 全体を設計し、動的な特性を含めた従来構成との比較評価を行っていく。

文　献

[1] F. Maloberti, Data Converters, Springer (2007).
[2] B. Razavi, Principles of data conversion system design. New York: IEEE Press (1995).
[3] J. B. Ricketts, “Switching circuit for a ladder type digital to analog converter utilizing an alternating reference voltage,” U.S. Patent 3,092,735 (June 4, 1963).
[4] H. Kobayashi, J. L. White and A. A. Abidi, “An Active Resistor Network for Gaussian Filtering of Images", IEEE Journal of Solid-State Circuits, vol.26, no.5, pp.738-748 (May, 1991)