

Log-Domain Circuit Models of Chemical Reactions

Soumyajit Mandal and Rahul Sarpeshkar

Department of Electrical Engineering and Computer Science
Massachusetts Institute of Technology, Cambridge, MA 02139

Email: rahuls@mit.edu

Abstract—We exploit the detailed similarities between electronics and chemistry to develop efficient, scalable bipolar or subthreshold log-domain circuits that are dynamically equivalent to networks of chemical reactions. Our circuits can be used for transient and steady-state simulations, parameter estimations and sensitivity analyses of large-scale biochemical networks. They allow the topology, rate constants, inputs, outputs and initial conditions of the reaction network to be programmed. When reactants are present in low concentrations, random fluctuations in reaction rates become significant; we can also model such stochastic effects. We present experimental results from a proof-of-concept chip implemented in 0.18 μm CMOS technology.

I. INTRODUCTION

Chemical kinetics and electronics are analogous at several levels. Chemical potentials map naturally to voltages, i.e., electronic potentials, while molecular fluxes map to electron flows, i.e., currents. Enzyme or catalyst concentration $[A]$ controls the energy barrier of a chemical reaction, exponentially changing its speed. In an analogous fashion, gate voltage V_G controls the electron energy barrier between source/drain terminals and the channel of a transistor, exponentially changing electron flow rate. The detailed analogy between chemical reactions and transistors with exponential I-V characteristics, in this case, subthreshold MOSFETs, is shown in Figure 1. We have exploited this analogy to build an integrated-circuit analog computer [1] for simulating systems of chemical reactions.

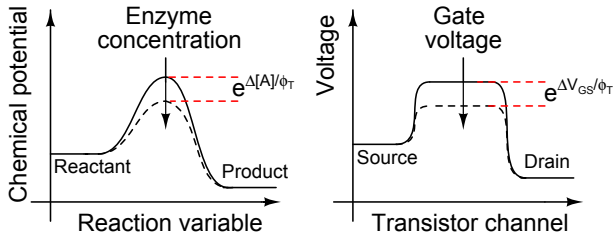


Fig. 1. Chemical potentials, molecular flux and enzyme concentration in a chemical reaction (left) are analogous to source/drain voltages, electron flow and gate voltage in a subthreshold MOS transistor (right).

II. THEORETICAL FORMULATION

A. Chemical Reaction Networks

Consider a reaction network composed of N distinct molecular species. The reaction medium is assumed to be a single-phase system, such as a dilute aqueous solution. It is also assumed to be well-stirred, i.e., spatial concentration gradients are negligible and the concentration of any species can be uniquely represented by a single number. The set of reactant concentrations at any time t forms a vector $\mathbf{x}(t) = [x_1(t), x_2(t), \dots, x_N(t)]$ of length N , where $x_i(t)$, $1 \leq i \leq N$ is the concentration of the i -th species.

A set of chemical reactions can be decomposed into elementary molecular steps in many ways. The correct set of such steps is known as the *mechanism* of the reaction. Since each elementary step follows

mass-action kinetics [2], the rate of change of x_i with time is given by

$$\frac{dx_i}{dt} \equiv \dot{x}_i = c_i + \sum_{j=1}^N d_{ij}x_j + \sum_{j=1}^N \sum_{k=1}^N e_{ijk}x_jx_k + \dots \quad (1)$$

where the first, second, third... terms on the right-hand side correspond to zeroth, first, second... order kinetics, respectively. Also, c_i , d_{ij} , e_{ijk} ,... are constants known as *kinetic rate constants*. Each rate constant can be positive (if species i is being produced in that reaction) or negative (if it is being consumed). In this formulation each reaction is unidirectional, i.e., the forward and backward parts of a reversible reaction are considered separately.

In principle, M -body molecular collisions result in M -th order mass-action kinetics. However, the probability of three or more molecules colliding simultaneously is usually negligible at practical temperatures, concentrations, and pressures. Therefore elementary steps are limited to zeroth, first or second order kinetics and the series in (1) can be safely terminated after the first three terms. In order to generalize our formulation we also note the following:

- Species concentrations can also depend on external inputs to the system. Let the vector of such external inputs be denoted by $\mathbf{u}(t) = [u_1(t), u_2(t), \dots, u_M(t)]$, where in general $N \neq M$. Inputs can affect species concentrations directly (resulting in first-order kinetics) or in combination with other species (resulting in second-order kinetics).
- The outputs of interest may consist of linear combinations of all the N species in the reaction network. Let the vector of such outputs be denoted by $\mathbf{y}(t) = [y_1(t), y_2(t), \dots, y_P(t)]$, where in general $P \neq N$ or M .

Using the usual Einstein summation-over-indices convention, our complete reaction model is given by

$$\begin{aligned} \frac{dx_i}{dt} &= c_i \cdot 1 + d_{ij}x_j + e_{ijk}x_jx_k + f_{ij}u_j + g_{ijk}x_ju_k \\ y_i &= h_{ij}x_j + k_{ij}u_j \end{aligned} \quad (2)$$

where c_i , d_{ij} , e_{ijk} , f_{ij} , g_{ijk} , h_{ij} and k_{ij} are constant coefficients. In matrix notation (2) becomes

$$\begin{aligned} \frac{d\mathbf{x}}{dt} &= \mathbf{C} + \mathbf{D}\mathbf{x} + \mathbf{E}(\mathbf{x} \otimes \mathbf{x}) + \mathbf{F}\mathbf{u} + \mathbf{G}(\mathbf{x} \otimes \mathbf{u}) \\ \mathbf{y} &= \mathbf{H}\mathbf{x} + \mathbf{K}\mathbf{u} \end{aligned} \quad (3)$$

where \otimes denotes the tensor or outer product. The similarity of (3) to the standard **ABCD** matrix model of a linear dynamical system is evident.

B. Electrical Circuit Equivalents

Our goal is to emulate the dynamics of the reaction system described in the previous section with an electrical circuit. We encode

the chemical potential of each species, i.e., the Gibbs free energy per molecule, as the voltage V on a capacitor of value C . In dilute solutions¹ the chemical potential of the i -th species is given by

$$\mu_i = \mu_0 + k_B T \ln \left(\frac{x_i}{X_0} \right) \quad (4)$$

where μ_0 and X_0 are constants referred to as the reference chemical potential and reference concentration, respectively. Also, x_i is the concentration of the species. To convert from μ to V we divide by κq , where κ is a constant and q is the electronic charge. Equation (4) can then be written as

$$\ln \left(\frac{x_i}{X_0} \right) = \frac{\kappa (v_i - V_0)}{\phi_T} \Rightarrow x_i = X_0 \exp \left(\frac{\kappa (v_i - V_0)}{\phi_T} \right) \quad (5)$$

where $\phi_T = k_B T / q$ is the thermal voltage and $V_0 = \mu_0 / (\kappa q)$ is a constant reference voltage. The concentrations of the input and output species are encoded similarly. Differentiating (5) on both sides, we get

$$\frac{d \ln(x_i)}{dt} = \frac{1}{x_i} \frac{dx_i}{dt} = \frac{\kappa}{\phi_T} \left(\frac{dv_i}{dt} \right) \quad (6)$$

For convenience we now convert concentrations to currents by defining $i_i / I_0 = x_i / X_0$, i.e., $i_i = I_0 \exp(\kappa (v_i - V_0) / \phi_T)$, where I_0 is a constant reference current. Similarly, we also define $i_{ui} / I_0 = u_i / X_0$ and $i_{yi} / I_0 = y_i / X_0$. Substituting (6) in (2), we get

$$C \frac{dv_i}{dt} = \frac{C \phi_T}{\kappa I_0} \left[\frac{c_i}{X_0} \frac{I_0^2}{i_i} + \sum_{j=1}^N d_{ij} \frac{I_0 i_j}{i_i} + X_0 \sum_{j=1}^N \sum_{k=1}^N e_{ijk} \frac{i_j i_k}{i_i} + \sum_{j=1}^M f_{ij} \frac{I_0 i_{uj}}{i_i} + X_0 \sum_{j=1}^N \sum_{k=1}^M g_{ijk} \frac{i_j i_{uk}}{i_i} \right] \quad (7)$$

$$i_{yi} = \sum_{j=1}^N h_{ij} i_j + \sum_{j=1}^M k_{ij} i_{uj} \quad (8)$$

Equations (7) and (8) are statements of KCL. The index i runs from 1 to N in the first equation (N state variables) and 1 to P in the second (P outputs). The reference concentration and current (X_0 and I_0) are normally chosen to be the geometric means of the minimum and maximum concentrations and currents of interest. In subthreshold CMOS implementations the minimum allowable current is set by leakage and parasitic capacitances, while the maximum is set by the onset of strong inversion.

Equation (7) can be easily implemented in hardware using log-domain circuits [3]. The currents i_i are proportional to $\exp(\kappa v_i / \phi_T)$, where $i_i \geq 0, \forall i$. Thus each current can be created by a single BJT or subthreshold MOSFET operated in its forward active (BJT) or saturated (MOSFET) region. In addition, real biochemical networks are sparse: most species participate in fewer than four reactions. Because of this sparseness, most of the coefficients $c_i, d_{ij}, e_{ijk}, f_{ij}$ and g_{ijk} are zero (the reactions in question do not occur). Therefore only a small subset of the $1 + N + N^2 + M + MN$ terms on the right hand side of (7) are non-zero. Each of these contributes a current $\pm \beta i_1 i_2 / i_i$ to $C dv_i / dt$, where β is a dimensionless, non-negative constant and i_1 and i_2 are non-negative currents. As a result, (7) can

¹A solution is considered dilute when interactions between solute particles are negligible compared to solute-solvent interactions. In this situation solute molecules essentially behave like an ideal gas.

be easily implemented with single-quadrant log-domain integrators, which can be implemented with very few transistors.

Equation (8) is also easy to implement: the state variable currents i_j and input currents i_{uj} (we have N of the former and M of the latter) are summed together at a single node with appropriate weighting factors h_{ij} and k_{ij} . The result is the output current i_{yi} . We carry out P such summations to produce the P output currents.

All reaction networks must satisfy the thermodynamic constraint that the net change in chemical potential around any reaction loop is zero. This is an application of the first law of thermodynamics, i.e. that total energy is conserved. We may express this statement mathematically as

$$\sum_{i \in loop} \mu_i = 0 \Rightarrow \sum_{i \in loop} v_i = 0 \quad (9)$$

where the second equation follows from the first by using (4) and (5). However, this second equation is simply KVL, which is automatically satisfied by any electrical circuit. Therefore our circuit model incorporates thermodynamic constraints. However, it does not accurately model changes in reaction rates with temperature, since activation energies can depend on temperature in complicated ways. Intuitively, this is because molecules may have several internal degrees of freedom that affect how they react with each other. For example, diatomic molecules can rotate about the bond linking the two atoms, a process which has its own characteristic dependence on temperature. We may also note that an analogous phenomenon occurs in electronics: in general the threshold voltage of a transistor is also a complicated function of temperature. However, at a given temperature, as illustrated in Figure 1, flux (current flow) in both chemical and electronic systems is proportional to $\exp(-E/kT)$, where E is the height of the energy barrier (activation energy or threshold voltage) that controls the flow.

C. Polynomially Nonlinear Dynamical Systems

The circuit formulation described by the KCL equation in (7) can be extended to dynamically simulate any *polynomially nonlinear dynamical system*. Such a system can be used to model mass-action chemical kinetics of any order; it consists of a set of N differential equations of the form

$$\frac{dx_i}{dt} = \sum_{j=1}^R c_{ij} (x_1^{p_1} x_2^{p_2} \dots x_N^{p_N} u_1^{q_1} u_2^{q_2} \dots u_M^{q_M}) \quad (10)$$

where R is a positive integer, $[p_1 \dots p_N]$ and $[q_1 \dots q_M]$ are integers that in general are different for each value of i and j , the c_{ij} 's are real constants and, as before, we have N state variables x_i and M inputs u_i . Following the same procedure described in the previous section, equation (10) can be rewritten in terms of the rate of change of $\ln(x_i)$. The result, which is easier to implement in log-domain circuit form, is

$$\frac{d \ln(x_i)}{dt} = \sum_{j=1}^R c_{ij} \left(\frac{x_1^{p_1} x_2^{p_2} \dots x_N^{p_N} u_1^{q_1} u_2^{q_2} \dots u_M^{q_M}}{x_i} \right) \quad (11)$$

Equation (11) can be interpreted as KCL, i.e., the rate of change of $\ln(x_i)$, the voltage on a capacitor, is equal to the sum of R currents that add and subtract charge from it. Each term on the right hand side of (11) represents a current that is a multinomial function of the state variables and inputs. Log-domain circuits can easily implement such functions. Therefore any polynomially nonlinear dynamical system can be modeled using a dynamically equivalent log-domain circuit.

The order S of the each term of the summation in equation (11) is defined as the sum of all the power-law coefficients in the numerator, i.e.,

$$S = \sum_{k=1}^N p_k + \sum_{k=1}^M q_k \quad (12)$$

The system of chemical reactions modeled by (7) is a special case of (11) when $S \in [0, 1, 2]$, i.e. only zeroth, first and second-order kinetics are allowed.

D. Noise

Individual chemical reaction events are usually uncorrelated. As a result, molecular fluxes exhibit shot noise. This behavior is exactly analogous to electronic shot noise, which is caused by diffusion currents within physical devices. In both cases individual fluxes exhibit Poisson statistics. However, in log-domain circuits noisy fluxes (currents) do not directly act on a state variable, i.e., species concentration. Instead, they add or subtract charge from a capacitor, the voltage on which is log-compressed, i.e., must be exponentiated to get a current that is the state variable. Because this operation is nonlinear, positive and negative fluxes that affect state variables do not display Poisson statistics. Thus, chemical and electronic state variables that behave identically in the high SNR or deterministic limit will have different noise properties.

If we ignore the noise produced by input (compression) and output (expansion) transistors, log-domain circuits produce total noise voltages of the form $\sqrt{\alpha kT/C}$, where the excess noise factor α is the effective number of noise sources (transistors) affecting a given log-compressed voltage. After exponentiation the SNR of each state variable becomes independent of its mean value and equal to $C\phi_T/(\alpha\kappa q)$. We have designed a feedback loop that modifies this behavior by dynamically adjusting the SNR of each state variable based on its mean value. In this way we can ensure that electronics and chemistry have similar noise properties. While we do not describe it further here owing to lack of space, we note that the loop allows our circuits to perform fast, accurate stochastic simulations. This ability is important because while noise has important effects in many biological systems, noisy systems are numerically stiff and simulate slowly on digital computers.

III. CIRCUIT IMPLEMENTATION

A. Chip Design

We reference all state variables to V_{DD} since PMOS transistors are our exponential elements. We use the log-domain integrator proposed in [4] as our primary building block, since it is guaranteed to be stable at all current levels and can be implemented on low power-supply voltages. Each integrator only needs to be unidirectional, since it models how flux from an unidirectional reaction changes the concentration of one species. In other words, the capacitor storing the chemical potential of the species is either charged or discharged by a current $i_C = \beta i_1 i_2 / i_i$, depending on whether the species is a product or a source, respectively. In some cases one of the inputs (i_1 or i_2) to the integrator is equal to the output i_i . In these cases i_C simplifies to either i_1 or i_2 and the integrator can be replaced by a current mirror. A complete reaction is modeled by using an integrator or current mirror for every participating species.

Both transient and steady-state behaviors of chemical networks can be simulated using our circuits, since the circuit equations shown in (7) and (8) are dynamically equivalent to the original chemical equations. Dynamical equivalence refers to the fact that the dynamics

of *normalized* chemical and electrical state variables, i.e., x_i/X_0 and i_i/I_0 , respectively, are identical. However, in order to simulate typical biochemical time constants of seconds to hours rapidly, our electronic circuits should be dynamically equivalent, not to the chemical dynamics themselves, but a time-scaled (sped-up) version of them. In order to get a speedup factor of α , normalized electronic state variables i_i/I_0 must have time derivatives that are α times larger than their chemical equivalents. As a result, the dimensionless number β that scales the capacitor current i_C is given by

$$\beta = \alpha\tau_0 X_0^{S-1} k \quad (13)$$

where the characteristic electronic time constant $\tau_0 = C\phi_T/(\kappa I_0)$, $S \in [0, 1, 2]$ is the order of the chemical reaction, and k is its kinetic rate constant. From (7), a single second-order reaction of the form $A + B \rightarrow C$ is described by the following equations:

$$\begin{aligned} Cdv_A/dt &= +\beta i_A i_B / i_A = +\beta i_B \\ Cdv_B/dt &= +\beta i_A i_B / i_B = +\beta i_A \\ Cdv_C/dt &= -\beta i_A i_B / i_C \end{aligned} \quad (14)$$

Note that the signs of the currents have been reversed since the state variable is now referenced to V_{DD} , i.e., given by $V_{DD} - v_i$. As a result increasing v_i decreases the state variable, and vice-versa. The first two equations in (14) require current mirrors, while the third requires a log-domain integrator. A simplified circuit implementation is shown in Figure 2. The W/L ratio of some transistors, indicated in the figure, are made β_1 and β_2 times larger than the other transistors using binary-weighted N -bit transistor arrays ($N = 5$ in this implementation). Therefore β_1 and β_2 can vary between 1 and 2^N , and $\beta = \beta_1\beta_2$ between 1 and 2^{2N} . Once the chemical rate constants k are given, α , τ_0 and X_0 must be chosen such that β for all reactions falls within this range.

We also add a fixed current I_{min} to $\beta_1 i_B$ in the actual implementation to ensure that i_B does not become small enough for parasitic capacitances inside the integrator to noticeably affect the dynamics of the state variables, i.e., A , B and C . An additional integrator and current mirror (not shown) is used to remove the effect of I_{min} , as follows:

$$\begin{aligned} Cdv_A/dt &= +\beta_2 (\beta_1 i_B + I_{min}) - \beta_1 \beta_2 I_{min} \\ &= +\beta_1 \beta_2 i_B \\ Cdv_C/dt &= -\beta_2 (\beta_1 i_B + I_{min}) i_A / i_C + \beta_2 I_{min} i_A / i_C \\ &= -\beta_1 \beta_2 i_A i_B / i_C \end{aligned} \quad (15)$$

A first-order reaction $A \rightarrow B$ is defined by the following equations:

$$\begin{aligned} Cdv_A/dt &= +\beta I_0 i_A / i_A = +\beta I_0 \\ Cdv_B/dt &= -\beta I_0 i_A / i_B \end{aligned} \quad (16)$$

These equations can be implemented with a current mirror and an integrator. However, since the value of the constant current I_0 is known *a priori*, I_{min} is not needed. This fact simplifies the circuit implementation. Finally, a zeroth-order reaction $[\] \rightarrow A$, where the species A is produced by an external flux (current source) is defined by the equation

$$Cdv_A/dt = -\beta I_0^2 / i_A$$

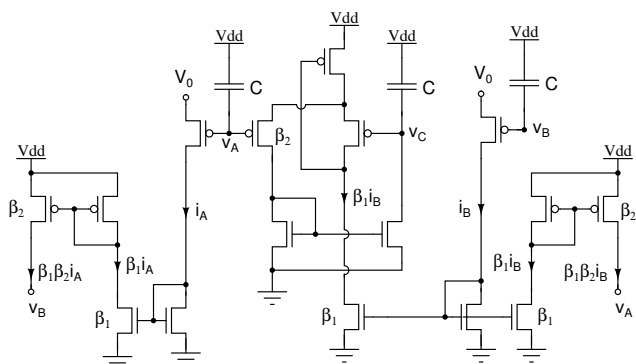


Fig. 2. Simplified schematic of a circuit that models a second-order chemical reaction.

This equation can be implemented with a single integrator, and again, I_{min} is not needed. The value of β for both first and zeroth-order reaction circuits is set in a similar way to the second-order case, i.e., by factorizing β into β_1 and β_2 , which are set by binary-weighted transistor arrays. We can now combine reaction circuits of various types to implement arbitrarily complicated systems of chemical reactions. We have designed a chip for this purpose that contains 81 second-order equation blocks, 40 first-order equation blocks, 40 zeroth order equation blocks, 32 state variables, 16 inputs and 8 outputs. State variables are stored on capacitors in log-compressed form, i.e., as chemical potentials.

The chip occupies 1.5mm x 1.5mm in a 0.18 μ m CMOS process. The topology of the reaction network is completely programmable, i.e., any terminal in any of the equation blocks can be connected to any of the state variables. The sizes of these capacitors can be individually set, allowing us to simulate systems where reactants and products are present in compartments with different volumes (such systems are common in biology). The parameters of a given network topology, i.e., reaction rates and initial conditions, are also individually programmable. Component mismatches will cause static offsets in the values of these parameters. However, in principle such offsets can be measured in an initial calibration step and subtracted out. Finally, this chip, being a prototype, does not contain any SNR-adjustment loops. As a result the SNR of any state variable is independent of the mean value of that variable.

B. Experimental Results

As an example, we implemented the simple reaction system $A + A \rightarrow B$, $B \rightarrow C$, which consists of one second-order and one first-order reaction, both in software (using MATLAB) and on our chip. The system was initialized at time $t = 0$ with a high initial concentration of A and low initial concentrations of B and C . The MATLAB simulation used an optimized version of the Gillespie stochastic simulation algorithm (SSA) [5], with the initial number of molecules of A set to a value that results in the same SNR as obtained experimentally from our chip (approximately 32dB).

Figure 3 compares the results of simulation and experiment. The two sets of trajectories are very similar, being always within 10% of each other. Since biological systems are both noisy and heterogenous, this level of accuracy may be sufficient for simulating many interesting phenomena. We see that the chip runs approximately 30 times faster than the simulation, which was performed on a 2.4GHz quad-core desktop computer. The speed advantage increases with the complexity of the reaction network: The simulation time

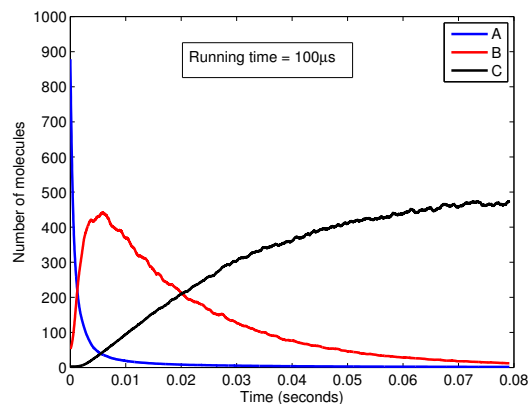
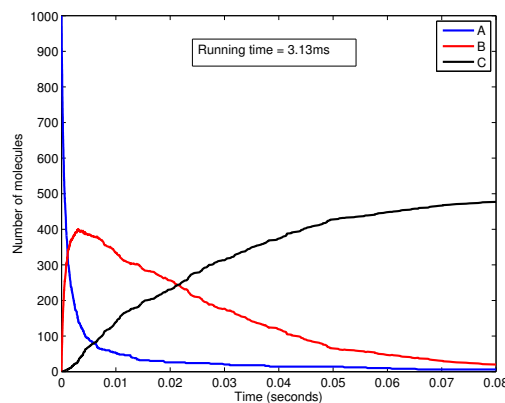


Fig. 3. Software simulation (top) and measurements from our chip (bottom) of the dynamics of the system of chemical reactions described in the text.

of this optimized SSA scales as $\log(r)$, where r is the number of reaction channels, whereas it is independent of r on the chip. When the SNR of every species is high enough, we can in principle use deterministic differential equations instead of the SSA. Since they ignore noise, the former run much faster than the latter, particularly when SNR levels are high. However, the SNR of each species varies with time, making it difficult to determine *a priori* if the resultant loss in accuracy will be acceptable. We can avoid this issue entirely by using our chips, since they run stochastic simulations with no performance penalties.

ACKNOWLEDGMENT

Soumyajit Mandal acknowledges support from a Poitras predoctoral fellowship, and Prof. Bruce Tidor for helpful comments.

REFERENCES

- [1] G. E. R. Cowan, R. C. Melville, and Y. P. Tsividis, "A VLSI analog computer/digital computer accelerator," *IEEE Journal of Solid-State Circuits*, vol. 41, no. 1, pp. 42–53, Jan. 2006.
- [2] L. Pauling, *General Chemistry*, 3rd ed. Mineola, NY: Dover Publications, 1988.
- [3] D. R. Frey, "Log-domain filtering: An approach to current-mode filtering," *IEE Proceedings-G: Circuits, Devices and Systems*, vol. 140, no. 6, pp. 406–416, Dec. 1993.
- [4] D. Python, M. Punzenberger, and C. Enz, "A 1-V CMOS log-domain integrator," *Proceedings of the IEEE Symposium on Circuits and Systems (ISCAS)*, vol. 2, pp. 685–688, June 1999.
- [5] D. T. Gillespie, "A general method for numerically simulating the stochastic time evolution of coupled chemical reactions," *Journal of Computational Physics*, vol. 22, no. 4, pp. 403–434, Dec. 1976.