A Resistor/Transconductor Network for Linear Fitting

Bertram E. Shi, Member, IEEE, Lina Gao, and Kwok Kit Lau, Student Member, IEEE

Abstract—We present a continuous time analog VLSI CMOS circuit consisting of resistors and transconductors for computing the best-fit line to a set of data points. The circuit can implement standard least-squares linear fitting, where the quality of the fit is determined by the sum of the squares of the deviations between the line and the data, as well as a form of linear fitting that is more robust to outliers. We analyze the static and transient response of the chip, and present design criteria given desired constraints on speed and accuracy. Finally, we describe the transistor level design and measurement results from a 50-input prototype fabricated using a 1.2 μm n-well process.

Index Terms—CMOS analog integrated circuits, least-squares methods, linear fitting, parallel architecture, robust statistics.

I. INTRODUCTION

This paper presents a continuous time analog VLSI CMOS circuit consisting of resistors and transconductors for computing the best-fit line to a set of data points. The circuit can implement standard least-squares linear fitting, where the quality of the fit is determined by the sum of the squares of the deviations between the line and the data, as well as a form of linear fitting that is more robust to outliers.

The circuit is targeted toward application in neuromorphic vision sensors, where analog pixel parallel processing circuits are implemented alongside photosensing circuits [1]. It is motivated along the same lines as circuits information from an array of pixels is “summarized” in a few voltages or currents which encode some global property of an image. This includes circuits for global velocity estimation [2], winner-take-all computation [3], [4], orientation computation [5] and centroid computation [6].

The architecture is perhaps most similar to the constraint-solving circuits of Tanner’s architecture for global velocity estimation [2]. A set of global wires distributes the current velocity estimates to the pixels. Each pixel checks to see if the velocity estimate satisfies its local constraint and generates a correcting current which charges or discharges the global wires. At steady state, the solution is a least-squares fit to the local constraints.

If the local image velocity estimate can be encoded as a voltage, this work opens the possibility to extend Tanner’s approach to the case where the velocity varies linearly over the array, i.e., affine motion. Affine models are used in parametric model-based algorithms for estimation of image motion, e.g. [7], [8]. For small fields of view and smooth changes in viewpoint, the image velocity field can be well approximated by an affine transformation

\[ v = T + Ar \]

where \( v \) denotes the image velocity, \( T \) is the translational velocity, and \( A \) is a linear transformation of the image coordinates \( r \) [9]. In the two-dimensional case, the matrix \( A \) can be decomposed into independent components, the divergence, curl and deformation. The divergence and deformation can be used to measure surface orientation and time-to-contact from a moving image [10]. In the 1D case as presented here, the divergence and deformation would be identical. However, the time-to-contact can be extracted if the camera translation is purely toward the surface patch (i.e., there is no translational velocity parallel to the image plane).

This work is also similar to work on image filtering using analog VLSI circuits [11]–[15]. The filtering operation can be considered as a more general form of curve fitting. The output at each pixel can be considered to be compromise between a data fidelity term and a regularization term. The data fidelity term ensures that the output is close to the input. The regularization term constrains the shape (e.g., smoothness) of the solution. It can also be chosen to account for intensity discontinuities [16]. The work here imposes a stricter linear model on the data.

The remainder of this section summarizes the basic concepts of linear fitting. Section II presents the proposed circuit architecture, as well as an analysis of its operation. Section III outlines the procedure by which the components of the architecture can be specified given desired computational characteristics of the array such as the speed and accuracy. Section IV details the transistor level design and test results from a 50-input prototype fabricated using the 1.2 μm AMI process available through MOSIS.

The problem of linear fitting is to find the line which best describes a set of data points \((x_n, y_n)\) for \(n \in \{1, \ldots, N\}\). Here, we restrict \(x_n = n\), so the data \(y_n\) correspond to samples equally spaced along the line. The “best fit” is determined by the choice of the cost function. Express the best-fit line by

\[ y = v_\alpha \left( \frac{n}{N+1} - \frac{1}{2} \right) + v_\beta \]

where \(v_\alpha\), the difference between the values of the line at \(n = 0\) and \(n = N + 1\), controls the slope of the line and \(v_\beta\), the value of the line at its midpoint \(n = (N + 1)/2\), controls the offset. Define the deviation between linear model and the data point \(y_n\) to be

\[ d_n = y_n - \left( v_\alpha \left( \frac{n}{N+1} - \frac{1}{2} \right) + v_\beta \right) \]

(1)
then for standard least-squares fitting, we wish to find the coefficients \(v_a\) and \(v_b\) that minimize

\[
E_k(v_a, v_b) = \frac{1}{2} \sum_{n=1}^{N} e_n^2.
\]  
(2)

The optimal values of \(v_a\) and \(v_b\) are

\[
v_a^* = \frac{12}{N(N-1)} \sum_{n=1}^{N} \left(n - \frac{N+1}{2}\right) \cdot y_n
\]
\[
v_b^* = \frac{1}{N} \sum_{n=1}^{N} y_n.
\]  
(3)

One problem with using the sum of squares cost function is the sensitivity of the optimal point to outliers, data points which are far from the linear fit. The sensitivity arises because large deviations between the linear model and the data are heavily penalized by the squaring operation. A cost function which is less sensitive to outliers is given by

\[
E_r(a, b) = \sum_{n=1}^{N} f(d_n)
\]  
(4)

where

\[
f(x) = \begin{cases} 
\frac{1}{2} x^2, & x < -d_T \\
\frac{1}{2} x^2 - d_T^2/2, & -d_T \leq x \leq d_T \\
d_T x - d_T^2/2, & x > d_T 
\end{cases}
\]

and \(d_T\) is a positive constant. The contribution by data points which deviate from the linear model by less than the threshold \(d_T\) still increases as the square of the deviation. However, the contribution by data points whose deviation exceeds \(d_T\) increases only linearly. As \(d_T \to \infty\), the cost function approaches (2). Unlike the least-squares case, there is no closed form expression for the optimal values of \(v_a\) and \(v_b\).

II. CIRCUIT ARCHITECTURE

The proposed circuit for linear fitting is shown in Fig. 1. The output voltages of the two op amps represent the endpoint values of the best-fit line at the origin \(\eta = 0\) and at the point \(\eta = N+1\). The upper resistor line computes the linear estimates of the data points \(y_n\) by linear interpolation between these two values

\[
v_n = v_0 + \frac{n}{N+1} (y_{N+1} - v_0),
\]  
(5)

The row of transconductance amplifiers computes the deviations between the predicted and actual data points. To implement the least-squares cost function, the output currents depend linearly upon the deviation. To implement the robust cost function, the output currents are linear for deviations less than \(d_T\), but saturate for larger deviations. The lower resistive line distributes these error currents appropriately to each capacitor so that its voltage moves in the correct direction to decrease the total squared error. The remainder of this section analyzes the operation of this circuit under the assumption that the op amps are ideal. The effects of a more realistic op-amp model are studied in the next section.

We consider the evolution of the differential and common mode components of the endpoint voltages

\[
v_a = v_{N+1} - v_0, \quad v_b = \frac{v_0 + v_{N+1}}{2}
\]

since their dynamics are independent, while the dynamics of \(v_0\) and \(v_{N+1}\) are coupled. Under the ideal op amp assumption, the voltages at the ends of the lower resistive line are held at virtual ground. This implies

\[
C_a \frac{dv_a}{dt} = -i_a, \quad C_b \frac{dv_b}{dt} = -i_b
\]

where \(i_a = i_n - i_t\) and \(i_b = (i_t + i_n)/2\). The evolution of the differential component depends upon the difference of the currents at the two ends of the lower resistive grid, while the evolution of the common mode component depends upon their average.

By superposition

\[
i_a = 2 \sum_{n=1}^{N} \left(\frac{n}{N+1} - \frac{1}{2}\right) \cdot i_n
\]
\[
i_b = \frac{1}{2} \sum_{n=1}^{N} i_n.
\]  
(6)
To implement least-squares fitting, we choose $i_n = G(v_n - y_n)$.

The currents $i_a$ and $i_b$ can each be split into two components, one due to the input data

$$i_{a1} = -2G \sum_{n=1}^{N} \left( \frac{n}{N+1} - \frac{1}{2} \right) y_n = -G a_i$$

$$i_{b1} = \frac{G}{2} \sum_{n=1}^{N} y_n = -G b_i$$

and one due to the best-line estimate

$$i_{a2} = G_a v_a \quad i_{b2} = G_b v_b$$

where we have substituted (5) and

$$G_a = \frac{G N (N+1)}{2(N+1)} \quad G_b = \frac{GN}{2}$$  \hspace{1cm} (7)

$G_a$ is the effective transconductance from the differential voltage to the differential current between the two ends of the lower resistive line. $G_b$ is the transconductance from the common mode voltage to the average of the two currents. Combining the above

$$\tau_a \frac{dv_a}{dt} = -v_a + v_a^s \quad \tau_b \frac{dv_b}{dt} = -v_b + v_b^s$$

where

$$\tau_a = \frac{C}{G_a} \approx \frac{6C}{GN} \quad \tau_b = \frac{C}{G_b} = \frac{2C}{GN}$$  \hspace{1cm} (8)

are time constants determining the response speed of the array. The response speed is determined by the settling time of the differential component. For time varying inputs, $y_n(t)$, the outputs are low pass filtered versions of the optimal estimates $v_a^s(t)$ and $v_b^s(t)$ with transfer function

$$\frac{V_a(s)}{V_a^s(s)} = \frac{1}{\tau_a s + 1} \quad \frac{V_b(s)}{V_b^s(s)} = \frac{1}{\tau_b s + 1}.$$  \hspace{1cm} (9)

The same array can implement the robust cost function in (4) by exploiting the natural saturation characteristic of many implementations of transconductance amplifiers. This idea has been exploited in other analog computational circuits [17], [18]. Differentiating (4) with respect to $v_a$ and $v_b$, we obtain

$$\frac{dE_r}{dv_a} = \sum_{n=1}^{N} \left( \frac{n}{N+1} - \frac{1}{2} \right) \cdot f'(d_n)$$

where $d_n$ is given by (1) and

$$f'(x) = \begin{cases} \frac{dE_r}{dv_a} & x > d_T \\ x, & \text{otherwise} \end{cases} \quad f''(x) = \begin{cases} \frac{dE_r}{dv_a} & x < -d_T \\ -\frac{dE_r}{dv_a} & x > d_T \\ -\frac{dE_r}{dv_b} & x < -d_T \end{cases}$$

Comparing with (6), if we let the output of the transconductors be $i_n = G f'(d_n)$ then

$$\frac{dv_a}{dt} \propto -\frac{dE_r}{dv_a} \quad \frac{dv_b}{dt} \propto -\frac{dE_r}{dv_b}.$$  \hspace{1cm} (9)

In other words, the differential and common mode voltages evolve such that the cost function is minimized. Since the cost function is convex, stability is guaranteed.

**III. CIRCUIT DESIGN CRITERIA**

In this section, we derive constraints upon the circuit elements used to implement the array which are used in the transistor level designs described in the next section. In particular, we replace the ideal op amp model used in the previous analysis with a more realistic model of a voltage-controlled current source (VCCS) in parallel with an output resistance. We assume that the capacitance $C$ is large enough that the internal dynamics of the operational amplifiers are negligible. In addition, we consider constraints arising from the finite output range of the transconductance amplifiers. In the end, we find that we can consider the number of input data $N$, the desired speed of the array, the capacitance, the desired accuracy, and the ratio of the output and input ranges of the transconductance amplifiers as the free parameters which determine the component values.

The analysis assumes that the transconductance amplifiers are linear, i.e., the least-squares cost function is implemented. The designed parameters should be effective for the robust cost function since for a small number of outliers, most of the transconductors should be operating in their linear region. In a sense, the linear analysis represents a worst case for the robust cost function, since if some of the transconductors are saturated, the aggregate gains $G_a$ and $G_b$ are reduced, improving the stability of the array at the expense of a decrease in speed.
A. Analysis With Nonideal Op Amp

Replacing the operational amplifier with a VCCS with gain $g_m$ in parallel with an output resistance $r_o$, we obtain the circuit shown in Fig. 2. Appendix I shows that

\[ v_A = \frac{1}{1 + \left(\frac{A_U A_L}{r_o}\right)^{-1}} \left( G_A - \frac{1}{2} \right) \]

where

\[ A_U = g_m \left( \frac{1}{r_o} + \frac{2}{R_{LT}} \right) \]

and $R_{LT} = (N+1)R_L$ is the total resistance of the lower resistive line and $R_{RT} = (N+1)R_U$ is the total resistance of the upper resistive line. $A_U$ is the open-loop voltage gain from the voltage difference between the left and right op-amp inputs to the voltage difference between the op amp outputs. $A_L$ is the similar open-loop voltage gain, but from output to input

\[ A_L = \frac{G_A R_{LT}}{2} \]

The finite transconductance of the op amp increases the response speed, although the array may be unstable if $g_m$ is too low. There is also less attenuation of high frequency components in the input (e.g., noise) due to an additional zero to the transfer function

\[ V_A(s) = \frac{-\tau_A s + k_A}{\tau_A s + 1} \quad V_I(s) = \frac{-\tau_{BC} s + k_A}{\tau_{BC} s + 1} \]

To limit the effect of this zero, we choose $\tau_{BC} > 10 \tau_A$. If the effect of the op-amp output impedance is negligible, this is equivalent to

\[ g_m > 5.5 G N \]

If the inputs are constant, the steady state values of the differential and common-mode voltages are

\[ v_A^\infty = \frac{v_A^\infty}{1 + \left(\frac{A_U A_L}{r_o}\right)^{-1}} \approx v_A^\infty \left(1 - \left(\frac{A_U A_L}{r_o}\right)^{-1}\right) \]

\[ v_B^\infty = v_B^\infty \]

The common mode voltage still settles to its ideal value. However, the differential voltage is decreased by a factor which depends upon the product $A_U A_L$. In order to ensure a given accuracy, say $A \%$, we must choose

\[ A_U A_L = g_m \left( \frac{1}{r_o} + \frac{2}{R_{LT}(N+1)} \right)^{-1} \frac{GR_{LT}N(N-1)}{12} > A^{-1} \]

For accuracy, $R_L$, $R_U$, and $r_o$ should be large. Typically, $N$ is fixed by the application and $G$ is fixed by the desired speed of the array. There is also an upper limit on the $R_L$ determined by the maximum output range of the transconductor. Increasing $R_L$ increases the node voltages $v_n$, which may exceed the output range of the transconductance amplifiers. We examine this constraint in more detail below.

Given a set of transconductor output currents $\{i_m\}_{m=1}^N$, and assuming the op amps are ideal, the voltage at node $n$ of the lower resistive line is

\[ \hat{v}_n = \frac{R_L(N+1-n)}{N} \sum_{m=1}^n m i_m + \frac{n R_L}{N} \sum_{m=n+1}^N (N+1-m) i_m \]

This equation is derived by superposition. The first sum corresponds to the voltage at node $n$ due to currents entering nodes to the left and the second corresponds to currents entering nodes to the right.

Below, we derive an estimate of the variance of $\hat{v}_n$ at steady state assuming the variance of the noise in the data is known. This enables us to bound the probability that $\hat{v}_n$ exceeds the output range of the transconductors. Appendix II derives a stronger bound guaranteeing that $\hat{v}_n$ does not exceed the output range of the transconductor even during the transient. The bound is based upon the operating condition, where the input is a horizontal line with value $-V_{inax}$, but the current estimate is a horizontal line with value $V_{inax}$. This condition would be rarely observed in practice and never observed at steady state. That bound is also less convenient from a design point of view since it dictates a smaller value of $R_L$, which in turn requires larger values of $R_U$ and $r_o$ (i.e., a larger op-amp open-loop voltage gain) to satisfy a given accuracy constraint.

Thus, we feel the steady-state variance statistical bound derived below, which should be sufficient to ensure correct operation for slowly varying stimuli under most conditions, is preferable.

Assume that the input is a noisy line

\[ y_n = a \left( \frac{n}{N+1} - \frac{1}{2} \right) + b + Z_n \]

where $Z_n$ are random variables which satisfy

\[ \frac{1}{N} \sum_{n=1}^N Z_n = 0, \quad \frac{1}{N} \sum_{n=1}^N \left( \frac{n}{N+1} - \frac{1}{2} \right)^2 \cdot Z_n = 0. \]

Assume that $E[Z_n] = 0$ and $\sigma^2[Z_n] = \sigma^2$ for all $n$.

At steady state, the differential and common-mode voltages settle to $v_A = a$ and $v_B = b$. The output currents of the transconductors are given by $i_n = GZ_n$. The variance $\sigma^2[Z]$ measures the expected deviation between the data $y_n$ and a line. Assuming that the correlation between $Z_m$ and $Z_n$ for $m \neq n$ is negligible, the variance of $\hat{v}_n$ is

\[ \sigma^2_{\hat{y}_n} = \frac{\sigma^2 G^2 R_L^2 (N+1)}{6N^2} \cdot n(2n^2 - (2N+2)n - 1)(n - N - 1) \]
which is maximized for 
\[ n = \frac{(N + 1)}{2}; \]
\[ \sigma_{\text{min}}^2 \leq \frac{g^2_0 R^2}{48N^2} (N + 1)^3 (N^2 + 2N + 3). \]
Since it is a weighted sum of \( N \) random variables \( Z_n \), we assume \( \tilde{Z}_n \) to be Gaussian. If \( 3\sigma_{\text{min}} < V_{\text{max}} \) then the probability that \( \tilde{Z}_n \) exceeds the output range of the transconductance amplifiers is less than 0.3%. This requires that
\[ GR_L < \frac{4N}{\sqrt{3(N + 1)^3(N^2 + 2N + 3)}} \frac{V_{\text{max}}}{\sigma_z} \]
\[ \approx \frac{4}{\sqrt{3N^2/2}} \frac{V_{\text{max}}}{\sigma_z} \]  
(14)
If \( Z_n \) is uniformly distributed between \( \pm V_{\text{max}} \) then \( \sigma_z = V_{\text{max}}/\sqrt{3} \). This implies that
\[ GR_L < \frac{4}{N^3/2} \frac{V_{\text{max}}}{\sigma_z} \]  
(15)

B. Component Specification Procedure
Combining the results above, we find that we can consider the number of input data points \( N \), the desired speed of the array \( \tau^a \), the capacitance \( C \), the desired accuracy of the differential voltage \( A \), and the ratio of the output and input ranges of the transconductance amplifiers \( V_{\text{max}}/V_{\text{max}} \), as free parameters. The component values \( G, R_L, R_U, g_m \) and \( r_o \) can be determined using the procedure outlined below.

Given \( N, \tau^a, C, A \) and \( V_{\text{max}}/V_{\text{max}} \)
1) choose \( G \) using (8)
\[ G > \frac{GC^2}{\tau^a} \frac{1}{N} \]
2) choose \( g_m \) according to (10)
\[ g_m > 5.5G N \]
3) choose \( R_L \) according to (15)
\[ R_L < \frac{4G^2}{C^2} \frac{V_{\text{max}}}{V_{\text{min}}^2} \]
4) choose \( r_o \) and \( R_U \) according to (11)
\[ \frac{1}{r_o} + \frac{2}{R_U(N + 1)} < \frac{g_m G R_L N(N - 1)}{A}. \]

For example, the following component parameters satisfy the requirements \( N = 50, \tau^a < 7 \mu s, C = 1.2 \text{nf}, A = 0.5\%, \) and \( V_{\text{max}}/V_{\text{min}} = 2 \)
\[ G = 24 \frac{\mu A}{V} \]
\[ g_m = 15 \frac{mA}{V} \]
\[ R_L = 250 \Omega \]
\[ R_U = 1 \text{k}\Omega \]
\[ r_o = 40 \text{k}\Omega. \]

IV. EXPERIMENTAL RESULTS
We have implemented a 50-cell array based on the design specified above on a 2 mm \( \times \) 2 mm die using the 1.2 \( \mu \)m AMI process provided through MOSIS. The prototype requires \( \pm 2.5 \) V supplies. This section describes the CMOS transistor level design of the circuits as well as the measured results from the prototype.

A. Transconductance Amplifier
We use the standard 5 transistor NMOS input differential pair transconductance amplifier shown in Fig. 3 to implement the transconductor. The transconductor is designed for an input range of \( \pm 300 \text{ mV} \). The measured transconductance is \( 24 \mu A/V \). The circuit is biased at \( I_b = 15 \mu A \).

B. Resistors
The resistors in the lower line \( R_L \) are implemented using polysilicon lines. Their measured resistance is \( 227 \Omega \). The resistors in the upper line \( R_U \) are implemented using complementary transistors in parallel biased in their linear region [Fig. 4(a)]. A global bias circuit generates the bias voltages \( V_{Gn} \) and \( V_{Gp} \) which are distributed to all circuits in the array.

Although complementary transistors enable large resistances in smaller area than required by polysilicon, the current varies nonlinearly with the terminal voltages. Using the EKV model [19], the current through the resistors can be approximated to second order by
\[ I = (G_U + KV_{cm})\Delta V \]  
(16)
where \( V_{cm} \) is the common mode voltage across the resistor and \( \Delta V \) is the differential voltage across the resistor and
\[ G_U = (R_U)^{-1} = n_m^1 R_{n} - n_p^2 V_{P} \]
\[ K = n_m^2 - n_p^1. \]
The pinchoff voltages \( V_{n} \) and \( V_{p} \) are referenced with respect to ground potential and are functions of the gate voltages \( V_{Gn} \) and \( V_{Gp} \). Note that they have opposite signs. The \( \beta \) parameters are given by \( \mu C_{XX}(W/L) \). The slope parameters \( n_m \) and \( n_p \) are approximately equal to one, but depend weakly upon the gate voltages. The \( W/L \) ratios of the transistors can be sized to provide the desired resistance while decreasing the effect of the common mode voltage. We chose the \( W/L \) ratios to minimize \( K \) based on BSIM3 simulation models.

Due to parameter mismatch, the measured variation in current due to common mode offsets is larger than the simulated variation. Fig. 4(b) plots the current through the resistor circuit versus the common mode voltage across the terminals for differential voltage varying between \( \pm 300 \text{ mV} \). Ideally, the lines should be horizontal. We obtain the parameters \( G_U = 87.3 \mu A/V \) \( (R_U = 11.5 \text{ k}\Omega) \) and \( K = 19.5 \mu A/V^2 \) by least-squares fitting of (16) to the data for \( -300 \text{ mV} \leq V_{cm} \leq 300 \text{ mV} \), which is the expected operating range of the resistors. Over this range, the variation in the conductance is 6.7\%.
Fig. 4. (a) Circuit used to implement the resistors $R_{ij}$ in the upper line. Transistor sizing given in units of $\lambda = 0.6 \mu m$. (b) Current through the resistor circuit plotted versus common mode voltage across the terminals $V_{cm}$ at 10-mV intervals. The seven curves represent different differential voltages ($\Delta V$) ranging between $-300 \text{ mV}$ and $300 \text{ mV}$.

C. Operational Amplifier

The operational amplifier is implemented using the conventional two stage op amp shown in Fig. 5. It is designed to have output resistance $r_o = 47 \text{ k}\Omega$ and transconductance $g_m = 15 \text{ mA/V}$.

D. Input Stage

The inputs to the array are scanned in and stored by a sample and hold circuit implemented at every cell and shown in Fig. 6 [20]. The voltages are connected to a common line and a shift register provides the voltages $V_{\text{in}}$ and $V_{\text{in}}'$ which select the capacitor which will store the current voltage on the input line. The capacitor voltage is connected directly to the inverting input of the transconductance amplifier to supply the voltage $v_n$.

E. Array Performance

To test the chip’s ability to perform linear fitting with clean data, inputs of the form

$$y_n = a \left( \frac{n}{N+1} - \frac{1}{2} \right) + b$$

with $b = 0 \text{ mV}$ and $a$ varying between $\pm600 \text{ mV}$ were applied to the chip. The steady-state values of $v_{\text{in}}$ and $v_{\text{in}}'$ are plotted versus $a$ in Fig. 7(a).

Ideally, the graph of $v_{\text{in}}$ should be a line with slope one passing through the origin and the graph of $v_{\text{in}}'$ should be a line with zero slope passing through the origin. However, there are slope and offset errors in both curves. The offset errors are caused by offsets in the op amps, transconductance amplifiers and measurement circuits. The slope error for $v_{\text{in}}$ is partially due to the finite transconductance of the op-amp, as described above. However, the observed error is much larger than predicted by this factor alone. Our measurements reveal that the actual resistance of the upper resistors near the right-hand side of the array is lower than that of those to the left. For $v_{\text{in}} > 0$, the voltages $v_{\text{in}}$ are larger than those which would be obtained by linearly interpolating between the two op amp outputs, since $v_{\text{in}}$ increases with $n$ faster for small $n$ than for large $n$. Thus, the best-fit line to $v_{\text{in}}$ has a larger slope than predicted by $v_{\text{in}}$ and a larger (more positive) offset than predicted by $v_{\text{in}}'$. For $v_{\text{in}} < 0$, the slope and offset are more negative. The magnitudes of the slope and offset errors increase with $|v_{\text{in}}|$.

The errors described above can be largely compensated by a linear scaling plus constant offset for $v_{\text{in}}$ and an additive offset for $v_{\text{in}}'$, which is an affine function of $v_{\text{in}}$

$$v_{\text{in},\text{comp}} = c_1 v_{\text{in}} + c_2$$

$$v_{\text{in}},\text{comp} = v_{\text{in}}' + c_3 v_{\text{in}} + c_4.$$  \hspace{1cm} (17)

For this chip, $c_1 = 1.13$, $c_2 = 6.7 \text{ mV}$, $c_3 = 0.046$, and $c_4 = 6.9 \text{ mV}$. The result of applying this compensation to the measured data is shown in Fig. 7(b). The remainder of the data reported uses this compensation.

To test the effect of varying the offset, clean data with $a = 100 \text{ mV}$ and varying values of $b$ between $\pm250 \text{ mV}$ were applied to the chip. The steady state values of $v_{\text{in}}$ and $v_{\text{in}}'$ after compensation are plotted versus $b$ in Fig. 8. As expected, $v_{\text{in}}$ is approximately a line with unit slope and $v_{\text{in}}'$ is approximately a line with zero slope.

To test the linear fitting capabilities of the circuit for noisy data, similar measurements were taken using noisy data

$$y_n = a \left( \frac{n}{N+1} - \frac{1}{2} \right) + b + Z_n$$
where $Z_n$ was generated using independent random numbers uniformly distributed between $\pm 2z_{\max}$ which were corrected by subtracting a linear and offset term so that (13) is satisfied. The mean absolute error between the chip output and the true line parameters computed over 500 trials is shown in Fig. 9 for two lines with different slope. Values of $z_{\max}$ ranged from 0 to 200 mV. The error increases with the noise level, with the error in the differential voltage being larger than that in the common mode voltage. For the largest noise level $z_{\max} = 200$ mV corresponding to a input mean absolute error of 100 mV, the mean absolute error in the common mode voltage is about 6 mV. The error in the differential voltage is about 27 mV.

Provision for measuring the response speed of the array was made by implementing CMOS transmission gates across the two capacitors. During normal operation, these transmission gates are open. When closed, the output of the op amp is shorted to virtual ground. By first closing then opening the gate and observing the op amp outputs, the speed of the response can be measured. This measurement was performed on clean data with $a = 100$ mV and $b = 0$ mV. The measured rise time (10%–90%) of the differential component was 15 $\mu$s, corresponding to a time constant of 6.8 $\mu$s. The rise time of the common mode component was 5.0 $\mu$s, corresponding to a time constant of 2.3 $\mu$s. As predicted by (8), the common-mode component evolves three times faster than the differential component. The power dissipation of the array was 4.7 mW.

Robust linear fitting was tested by decreasing the bias current of the transconductance amplifier so that the output current saturated at a differential voltage across the inputs of 100
Fig. 10. Mean absolute error in the differential voltage ($v_d$) and the common mode voltage ($v_b$) calculated over 500 trials with the chip configured for least-squares (l.s) and robust (r) linear fitting. The input data consisted of a line corrupted by impulsive noise with fixed amplitude. The percentage of corrupted data points varied from 0% to 100% in 4.2% steps. The upper two traces show the error in the differential component is larger than that in the common mode component (the lower two traces). Results from the chip configured for robust fitting (solid lines) are better than those for the chip configured for least-squares fitting (dashed lines) if the percentage of corrupted data is less than 55%.

mV. Clean data with $a = 100$ mV and $b = 0$ mV was corrupted by additive impulsive noise with a fixed magnitude of 200 mV. The percentage of corrupted data points was varied from 0 to 100 with positive and negative impulses being equally likely. The mean absolute error between the chip output and the true line parameters are compared for the chip configured for least-squares fitting and for robust fitting in Fig. 10. For fewer than 55% corrupted data points, the mean absolute error for robust fitting is smaller than that for least-squares fitting, since the data points which are corrupted contribute less to the cost function. On the other hand, when most of the data points are corrupted, the least-squares circuit performs better since it better utilizes the information contained in the corrupted data to estimate the underlying slope and offset.

Reducing the bias current for robust fitting decreased the power consumption to 1.9 mW. The speed of the array was also reduced due to the decrease in the transconductance $G$. For clean data with $a = 10$ mV and $b = 100$ mV, the rise time of the differential component increased to 575 $\mu$s and the rise time of the common mode component increased to 266 $\mu$s.

V. CONCLUSION

We have described a continuous-time analog CMOS circuit architecture for performing least-squares and robust linear fitting targeted at applications in analog parallel processing arrays. Design criteria in terms of desired computational characteristics of the array, such as speed, accuracy and input range were derived. Test measurements from a 50-input prototype verify the functionality of this architecture.

APPENDIX I

Since we are no longer assuming an ideal op amp, the input voltage at the inverting inputs of the op amps $\hat{v}_a$ and $\hat{v}_N+1$ are no longer held at virtual ground. Define

$$\hat{v}_a = \hat{v}_{N+1} - \hat{v}_0, \quad \hat{v}_b = \frac{\hat{v}_0 + \hat{v}_{N+1}}{2}.$$ 

By superposition

$$i_a = -G_a \hat{v}_a + G_a v_a = \frac{2\hat{v}_a}{R_{LT}},$$

$$i_b = -G_b \hat{v}_b + G_b v_b$$

where $G_a$ and $G_b$ are given by (7) and $R_{LT} = (N+1)R_L$ is the total resistance of the lower resistive line. The factor of 2 arises in the first equation since the current due to the voltage difference $v_a$ which leaves the left side of the lower resistive grid is the negative of the corresponding current leaving the right side, and it is counted twice in the difference.

By KCL, the output currents of the op amps must satisfy

$$-g_m \hat{v}_{N+1} = \frac{v_{N+1}}{r_o} - \frac{v_a}{R_{UT}} - i_r,$$

$$-g_m \hat{v}_0 = \frac{v_0}{r_o} - \frac{v_a}{R_{UT}} - i_r$$

thus

$$\hat{v}_a = \left( g_m + \frac{2}{R_{LT}} \right)^{-1} \left[ \left( G_a \frac{1}{r_o} - \frac{2}{R_{UT}} \right) v_a - G_a v_a^* \right]$$

$$\hat{v}_b = \frac{1}{g_m} \left[ \left( G_b - \frac{1}{r_o} \right) v_b - G_b v_b^* \right].$$

The voltages across the left and right capacitors $v_{C_L}$ and $v_{C_R}$ can be split into differential and common mode components as well:

$$v_{C_a} = v_{C_R} - v_{C_L} = v_a - \hat{v}_a$$

$$v_{C_b} = \frac{v_{C_L} + v_{C_R}}{2} = v_b - \hat{v}_b$$

Substituting (19)

$$v_{C_a} = \left( g_m + \frac{2}{R_{LT}} \right)^{-1} \left[ \left( g_m + \frac{2}{R_{LT}} - G_a \right) + \frac{1}{r_o} \right] v_a + G_a v_a^*$$

$$v_{C_b} = \frac{1}{g_m} \left[ \left( g_m - G_b + \frac{1}{r_o} \right) v_b + G_b v_b^* \right].$$

Differntiating

$$\frac{dv_{C_a}}{dt} = \left( g_m + \frac{2}{R_{LT}} \right)^{-1} \left[ \left( g_m + \frac{2}{R_{LT}} - G_a + \frac{1}{r_o} \right) \frac{dv_a}{dt} + G_a \frac{dv_a^*}{dt} \right]$$

$$\frac{dv_{C_b}}{dt} = \frac{1}{g_m} \left[ \left( g_m - G_b + \frac{1}{r_o} \right) \frac{dv_b}{dt} + G_b \frac{dv_b^*}{dt} \right].$$

(20)
Substituting (19) into (18)
\[ i_a = \frac{g_m G_a + \frac{2}{R_{LT}} \left( \frac{1}{\tau_o} + \frac{2}{R_{LT}} \right) v_a}{g_m + \frac{2}{R_{LT}}} v_a - \frac{g_m G_a}{g_m + \frac{2}{R_{LT}}} v_a^* \]
and substituting (20) and (21) into
we obtain
\[ C \frac{dv_a}{dt} = -i_a \quad C \frac{dv_b}{dt} = -i_b \]
where
\[
\tau_\alpha = \frac{1}{g_m} \left( G_a - \frac{1}{r_o} - \frac{2}{R_{LT}} \right) \]
\[
k_\alpha = \frac{1}{1 + (A_U/A_L)^{-1}}
\]
\[
\tau_{\alpha z} = \frac{1}{g_m} \left( 1 + (A_U/A_L)^{-1} \right)
\]
\[
\tau_\beta = \tau_\alpha \left( 1 - \frac{1}{g_m} \left( G_b - \frac{1}{r_\alpha} \right) \right)
\]
\[
\tau_{\beta z} = \frac{C}{g_m}
\]
The time constants \( \tau_\alpha \) and \( \tau_\beta \) are from the ideal op amp case given in (8). The term
\[ A_U = g_m \left( \frac{1}{r_\alpha} + \frac{2}{R_{LT}} \right)^{-1} \]
is the open-loop voltage gain from the voltage across the lower resistive line \( \hat{v}_b \) to the voltage across the upper resistive line \( v_a \).
The term
\[ A_L = \frac{G_a R_{LT}}{2} \]
is the open-loop voltage gain from \( v_a \) to \( \hat{v}_a \).

**APPENDIX II**

Here we derive an upper bound on the value of \( \hat{v}_n \) given the maximum voltage input to the transconductance amplifiers \( V_{\text{imax}} \). Unlike the estimate of the steady-state variance, this is a deterministic estimate which is valid for the entire transient. Since all of the coefficients in the sum (12) are positive, \( \hat{v}_n \) is maximized if all of the \( i_n \) achieve their maximum value \( i_n = 2G \hat{V}_{\text{imax}} \). This corresponds to the operating condition when the input data \( y_n \) is a horizontal line with value \( -V_{\text{imax}} \) but the current parameter estimates correspond to a horizontal line with value \( +V_{\text{imax}} \). Substituting the value of \( i_n \) into (12) and evaluating the sum, we obtain
\[
|\hat{v}_n| \leq \frac{G R_L V_{\text{imax}} (N + 1)^3}{4N}
\]
The right-hand side of the inequality reaches its maximum at \( n = (N + 1)/2 \). Let \( V_{\text{imax}} \) be the maximum output of the transconductance amplifiers with respect to ground. To guarantee proper operation
\[
G R_L < \frac{4N}{(N + 1)^3} V_{\text{max}}
\]
This bound is stronger than the variance based bound given in (14), assuming that \( \sigma_z = V_{\text{imax}} / \sqrt{3} \).

**REFERENCES**

Bertam E. Shi (S’93–M’95) received the B.S. and M.S. degrees in electrical engineering from Stanford University, Stanford, CA, in 1987 and 1988, and the Ph.D. degree in electrical engineering from the University of California at Berkeley in 1994.

He joined the Hong Kong University of Science and Technology, Kowloon, Hong Kong, as an Assistant Professor in the Department of Electrical and Electronic Engineering in 1994. His research interests are in analog VLSI neural networks, cellular neural networks, neuromorphic engineering, image processing, computer vision and speech recognition.

Dr. Shi served as an Associate Editor for the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS I between 1997 and 1999, and as Secretary for the IEEE CAS Society Technical Committee on Cellular Neural Networks and Array Computing in 1999.

Lina Gao received the B.S. degree from the Xi'an Jiaotong University, Xian, China, in 1985, the M.S. degree in 1988 from the China University of Science and Technology, and the Ph.D. degree from the Xi'an Jiaotong University in 1994.

She joined the Xi'an Jiaotong University in 1994 as an Assistant Professor in the School of Electrical Engineering. Between 1997 and 1998, she was a Research Associate at the Hong Kong University of Science and Technology, Kowloon, Hong Kong.

Kwok Kit Lau (S’99) was born in Hong Kong. He received the B.S. degree in electronic engineering from the Hong Kong University of Science and Technology, Kowloon, Hong Kong, in 1998, where he is currently working toward the M.Phil. degree.

He has been working on the estimation of image velocity. His research interest is on mixed-signal VLSI design, power electronics, and control applications.