CMOS transconductance amplifiers, architectures and active filters: a tutorial

E. Sánchez-Sinencio and J. Silva-Martínez

Abstract: An updated version of a 1985 tutorial paper on active filters using operational transconductance amplifiers (OTAs) is presented. The integrated circuit issues involved in active filters (using CMOS transconductance amplifiers) and the progress in this field in the last 15 years is addressed. CMOS transconductance amplifiers, nonlinearised and linearised, as well as frequency limitations and dynamic range considerations are reviewed. OTA-C filter architectures, current-mode filters, and other potential applications of transconductance amplifiers are discussed.

1 Introduction

An operational transconductance amplifier (OTA) is a voltage controlled current source (VCCS). The authors present an updated version of a tutorial paper published in 1985 [1]. One of the first papers on OTAs in the literature appeared nearly 30 years ago [2]. This paper described a bipolar OTA. At that time the emphasis was on amplifiers with feedback, such as op-amps. Thus the commercial OTAs were not meant to be used in open loop mode. The maximum input voltage for a typical bipolar OTA is of the order of only 30mV, but with a transconductance gain tunability range of several decades. Since then, a number of researchers have investigated ways to increase the input voltage range and to linearise the OTA. Some of the key attractive properties of OTAs are their fast speed in comparison with conventional low output impedance op-amps, and their bias dependence transconductance programmability (tunability). The wideband of the OTA is due in part to the fact that their internal nodes are low impedances. However, the internal low impedance and parasitic capacitance still cause a non-zero transconductance phase shift, known as 'excess phase' \( \phi_0 \). When the OTAs are connected in a system in closed loop, the excess phase makes the actual frequency response deviate from the ideal case, especially for high-\( Q \) systems. In the extreme case, the system may become unstable if the excess phase is not reduced. The main characteristics of a practical OTA are: (i) limited linear input voltage range, (ii) finite bandwidth, (iii) finite signal to noise ratio \( (S/N) \), and (iv) finite output impedance. The \( S/N \) is a function of the OTA architecture among other factors. The output impedance can be increased using cascode structures at the expense of reduced output signal swing. Their programmability is caused by the transconductance \( g_m \) bias dependence; this dependence allows several decades of tuning for transconductance with MOS transistors operating in weak inversion (or by using bipolar transistors) and about two octaves for MOS transistors operating in strong inversion.

The IC pioneer works on transconductors using BJT-JFET and CMOS were reported in 1980 [3, 4, 1981 [5], and 1984 [6], respectively. In 1985 an invited tutorial paper on OTAs [1] served to motivate a number of researchers to investigate new CMOS OTA architectures and their applications. For readers not familiar with OTAs, we suggest they read [1] to understand the background needed to take advantage of this tutorial. A number of significant contributions have been reported since 1985, including OTAs for open loop applications such as continuous-time filters, multipliers, nonlinear circuits and closed loop applications mainly for switched-capacitor circuits. The importance of the OTA is reflected by its inclusion in the textbooks [7-9].

2 Transconductance amplifiers: topologies

An ideal transconductance amplifier is an infinite bandwidth voltage-controlled current source, with an infinite input and output impedance. As shown in Fig. 1a, the simplest single input real transconductor is a MOS driver transistor \( M_1 \) operating in the saturation region. One of several drawbacks of this simple transconductor is its relative low output impedance. Several alternatives have been suggested to alleviate this problem. Figs. 1b-d show a group of cascode transconductors with high output impedance. Another useful simple transconductor is reported in [10]. It is often the case for Figs. 1b and c that \( M_1 \) operates in the ohmic region \([11-15]\). This provides better linearity, but the transconductance is reduced in comparison with \( M_1 \) operating in saturation. The amplifier A further increases the output resistance of the circuit shown in Fig. 1c; a simple MOS inverter or a bipolar inverter could replace the amplifier. Also, \( M_2 \) in Figs. 1b and c can be replaced by a BJT. The typical folded cascode structure is illustrated in Fig. 1d. A summary of the properties of these structures, when all transistors operate in the saturation region, is given in Fig. 2. If a positive simple \( g_m \) is required, the circuit of Fig. 1e is a possible implementation. Fig. 1f represents the symbol for the OTA with differential inputs, along with the ideal small signal equivalent circuit [16]. Note that \( g_m \) is a function of the amplifier bias current, \( I_{\text{DCE}} \). For the case of OTAs using MOS transistors in saturation...
Simpleila
Cascodellb
Enhancedk
Foldedld
Single input circuits

- Negative simple transconductor
- Cascode transconductor
- Enhanced transconductor
- Folded-cascode transconductor
- Positive simple transconductor
f OTA symbol representation and equivalent model

<table>
<thead>
<tr>
<th>Structure/ Figure</th>
<th>Res</th>
<th>Min VDD *</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simple/la</td>
<td>1/2b</td>
<td>( \frac{2V_{DD}}{k(W/L)} + V_{Tn} + V_{sat,I_b} )</td>
</tr>
<tr>
<td>Cascode/lb</td>
<td>( \frac{g_m}{g_m^2} )</td>
<td>( \frac{2V_{DD}}{k(W/L)} + V_{Tn} + V_{sat,I_b} )</td>
</tr>
<tr>
<td>Enhanced/le</td>
<td>( \frac{A_g}{g_m^2} )</td>
<td>( \frac{2V_{DD}}{k(W/L)} + V_{Tn} + V_{sat,I_b} )</td>
</tr>
<tr>
<td>Folded/lc</td>
<td>( \frac{g_m}{g_m^2} )</td>
<td>( \frac{2V_{DD}}{k(W/L)} + V_{Tn} + V_{sat,I_b} )</td>
</tr>
</tbody>
</table>

Fig. 1  Properties of simple transconductors
* The bottom devices of the cascode pairs have an aspect ratio of \( (W/L)_1 = (W/L)_2 = n^2 \). k is a technological parameter determined by the mobility, and the gate oxide.

Fig. 2  OTA symbol representation and equivalent model

Fig. 3  Differential OTAs
a Simple differential OTA
b Balanced OTA
c Conventional fully differential OTA without CMF
d Fully differential OTA with inherent CMF
e Pseudo differential OTA, \( G_{out} = G_{out} \)

IEE Proc.-Circuits Devices Syst., Vol. 147, No. 1, February 2000
the \( g_{cs} \)s are proportional to \( V_{bs} \) for MOS transistors operating in weak inversion or bipolar transistors the \( g_{cs} \)s are directly proportional to \( I_{bs} \). Also observe that the ideal OTA has an infinite output impedance; in practice the output resistance \( R\) is as shown in Fig. 2. The circuits of Figs. 1a-c can be modelled by the symbol of Fig. 1f when one of the inputs is grounded. A non-ideal OTA macro-model [17] will have finite input and output impedances, and \( g_{cs} \) will have a single pole model to be discussed later.

Next, we discuss the basic transconductor (OTA) topologies with differential inputs. Fig. 3a shows a basic differential input OTA with one current-mirror; in Fig. 3b a balanced OTA with three current mirrors and a single output is shown. Fig. 3c illustrates a fully differential OTA; the common-mode feedback [18] is not shown. Fig. 3d is a really symmetric architecture, which has inherently common-mode feedforward (CMFF) [19, 20]. Also note that to obtain very high output impedance, the amplifier A in Fig. 1c might be substituted by the OTAs of Fig. 3a or Fig. 3b with the proper frequency compensation to guarantee stability. The complexity of these structures is also accompanied by an improvement in offset reduction and robustness, but not necessarily with an improvement for high-frequency applications. Thus trade-offs between speed and accuracy should be established for each particular application. Note that the circuits in Fig. 3 do not have very high output impedance. To accomplish that, the OTA output branches should be replaced by the architectures illustrated in Figs. 1b-d. To yield an improved performance, the simple current mirrors of Fig. 3 are often substituted by enhanced current mirrors (see [7] and [21], Chap. 6).

A suitable architecture for low voltage power supplies is the pseudo-differential transconductance [19, 20, 22]. It consists of two single input transconductors (see Figs. 1a-d), and it looks like a differential pair, with the tail current of the differential pair substituted by a short-circuit. This configuration needs to have a common-mode circuit to drastically reduce the common-mode output voltage gain. One approach [19] consists of using an additional, two (equal) output current transconductor, with non-differential inputs as depicted in Fig. 3e. Note that the transconductor B does not have a differential output, but has two equal outputs which are added to the pseudo-differential transconductor such that common-mode signal can be rejected. This structure utilises a common-mode feedforward (CMFF) circuit implemented by the double input transconductance amplifier B. The performance of single-ended structures can be further improved by using fully differential topologies. In these topologies, the signal is referred to differential signal paths instead of to the commonly used analogue ground. The differential circuits are fully symmetrical, as shown in Figs. 3c and 3d, and their main advantages are due to this characteristic. The supply noise is injected to both OTA outputs with the same amplitude and same phase, hence they can be considered as common-mode noise. If the fully-differential transconductor presents nonlinear characteristics, the output currents, for \( v_2 = v' \) and \( v_1 = v_2 \), can be expressed by the following series expansions:

\[
i_{O1} = i_{Q1} = I_{B1} + a_1(v_2 - v_1) + a_2(v_1 - v_2)^2 + a_3(v_1 - v_2)^3 + \ldots
\]

and

\[
i_{O2} = i_{Q2} = I_{B1} + a_1(v_2 - v_1) + a_2(v_2 - v_1)^2 + a_3(v_2 - v_1)^3 + \ldots
\]

where \( I_{B1} \) is the amplifier bias current. It is evident from these expressions that an inversion of the differential input signal produces an inversion on the odd-order terms, while it has no effect on the phase of the squared components (even-order distortions). The even harmonic distortion components appear at the outputs with the same amplitude and same phase, and they ideally cancel each other when the differential output current is processed. In practice, process parameter tolerances and temperature gradients introduce transistor mismatches, avoiding the complete cancellation of common-mode signals. An additional advantage of fully differential systems is that the output signal swing is larger. According to eqns. 1 and 2, the fundamental output component at each output is given by \( -a_1V \), while the differential output \( (v_0 = v_1 - v_2) \) is \( -2a_2V \), where \( v_0 = v_1 - v_2 \). The main advantages of a fully differential topology are due to its symmetry, making the structure less sensitive to common-mode signals. However, mismatches in the N-type and P-type current sources might push both OTA outputs to the supply rails, and due to the differential nature of the system this effect is neither detected by the next stage nor corrected. To overcome this shortcoming, a common-mode feedback loop (CMFB) that controls the operating point is commonly used. The design of the CMFB [18] is not straightforward because the main signals are differential and the common-mode signals must be detected and suppressed with simple and fast circuitry. The circuit must present a very small impedance for the common-mode signals but be transparent (very high impedance) for the differential ones. The basic concept of the CMFB loop is shown in Fig. 4. The output voltage for this

---

**Fig. 4** Common-mode feedback basic circuit concept

- a Basic common-mode detector
- b CMOS CMFB implementation

IEE Proc.-Circuits Devices Syst., Vol. 147, No. 1, February 2000
circuit is taken across the drains of the transistors M₀. Fig. 4a shows conceptually a basic OTA with a common-mode detector implemented by two resistors. Practical solutions can be found elsewhere [18, 22–26]. The common-mode signals are sensed by averaging the OTA outputs, and compared with the AC ground (AC GND in Fig. 4) by the additional differential pair composed by the Mₘₖ transistors, and the resulting current is used to adjust the bias current of the main OTA. The common-mode open loop impedance ($R_{cm} = \frac{V_{cm}(s)}{i_{o}(s)}$) is determined by the small-signal transconductance of the common-mode loop. Transistors Mₘₖ compare the common-mode voltage with the AC ground, and transistors Mₚ mirror the resulting current to the OTA outputs. As a result, the common-mode transconductance is determined by transistors Mₘₖ. The parasitic pole of the common-mode loop is associated with the common mirror, transistors Mₚ in Fig. 4, and reduces the loop gain at higher frequencies. This yields:

$$g_{cm,jb}(s) \approx \frac{g_{cm}}{1 + \frac{3C_{Gsp}}{g_{m}} s^{2}}$$

(3)

where the subscript $p$ refers to parameters of Mₚ. The factor 3 appears due to the connection of three transistors in node A. The operating point of the OTA outputs ($v_{o1} + v_{o2}$) is forced by the CMFB to be around the analogue ground (i.e. an appropriate DC bias voltage). Note in eqn. 3 that due to the parasitic pole the common-mode transconductance is reduced at higher frequencies, hence increasing the common-mode impedance and being less efficient for the rejection of common-mode signals. For the common-mode feedback loop two poles should be considered. The dominant pole is associated with node $V_{o1}$ (and $V_{o2}$), and the non-dominant pole is associated with node A. Similarly to the typical differential loops, the phase margin must be larger than 45°, otherwise common-mode oscillations could appear in the system. As a rule of thumb, the common-mode gain-bandwidth product ($g_{cm}C_{l}$) must be smaller than the non-dominant pole ($g_{cm}C_{Gsp}$) obtained in eqn. 3. Ideally the bandwidths for both differential and common-mode gains should be comparable.

3 Linearisation techniques

The structures discussed in the previous Section are nonlinear, which means that they have a very small input voltage range yielding say 1% total harmonic distortion (THD). A solution to this problem requires techniques to linearise the transconductor. There are three types of linearisation techniques reported in the literature, i.e. (a) attenuation, (b) nonlinear terms cancellation, and (c) source degeneration. The ideal output current of a differential input transconductor is:

$$i_{O}(v_{1}, v_{2}) = (v_{1} - v_{2})g_{m}$$

(4)

where $v_{1}$ and $v_{2}$ are the positive and negative input signals of the transconductor. In reality, since the transconductors use MOS transistors for their implementations, they are nonlinear devices. For simplicity of the discussion we will assume only nonlinearities of practical interest. In general, we can assume that $i_{O}(v_{1}, v_{2})$ is given by:

$$i_{O}(v_{1}, v_{2}) = \sum_{i=1}^{\infty} a_{i}v_{1}^{i} + \sum_{i=1}^{\infty} b_{i}v_{2}^{i} + \sum_{i=1}^{\infty} c_{i,j}v_{1}^{i}v_{2}^{j} + i_{OS}$$

(5)

From this expression we can infer that in order to have a linear transconductor one option is that the input voltage be made small, such that $i_{O}$ yields:

$$i_{O}(k_{1}v_{1}, k_{2}v_{2}) = k_{1}i_{O}(v_{1}, v_{2}) = k_{1}(v_{1} - v_{2})g_{m}$$

(6)

Thus a basic linearisation idea consists of attenuating the input signals by a factor $k$. This attenuation yields a linearised approximation that can be expressed as:

$$i_{O}(v_{1}, v_{2}) \approx k g_{m}(v_{1} - v_{2})$$

(7)

There exist several practical techniques to implement the attenuation factor. The concept is illustrated in Fig. 5a. The circuit in Fig. 5b is often used for commercial discrete OTAs. Figs. 5c–e refer to attenuation to the driving transistor M₁ in Fig. 1 and to M₁ and M₂ (differential pairs) of Fig. 3. The use of floating gate techniques ([21], chaps. 5 and 6) as depicted in Fig. 5e yields a capacitance divider, where $C_{1}$ and $C_{Gsp}$ are the capacitances associated to the input signal $v_{i}$ and the bias voltage, respectively. Fig. 5f shows a bulk driven transistor [27–30] attenuation technique, which can operate at low and medium frequency ranges. An active attenuator [31] with good linearity is illustrated in Fig. 5e. In these linearised schemes, the OTA deals with an attenuated version of the input signal. To compensate this attenuation, the transconductance gain must be increased by the same factor, increasing both power consumption and silicon area. If the noise contribution of the attenuator is negligible, the input referred thermal noise of the transconductor (attenuator and OTA) increases by the square root of the attenuation factor.

3.1 Active attenuation

Fig. 5 Attenuation implementation

- a Concept: $i_{O}(v_{i}) = k g_{m}v_{i} + g_{m}k_{1}v_{2}^{2} + g_{m}k_{2}v_{2}^{2} + \ldots$
- b For discrete OTAs, $k = R_{o}/(R_{i} + R_{o})$
- c Using floating gate techniques $k = C_{1}/C_{Gsp}$
- d Using a bulk-driven transistor $k = \gamma$, $0 < \gamma < 0.4$
- e Active attenuation $k = 1 - \sqrt{1 + E_{in}(W/L)_{F}}$ for $v_{o1} = v_{o2}$

More elegant techniques exist to linearise transconductors by means of an optimal algebraic sum of nonlinear terms [7, 21, 32–39] yielding ideally only a linear term. Fig. 6 illustrates the conceptual ideas of this linearisation technique. This can be done in practice by interconnecting several transconductors, which ideally will cancel the nonlinearities yielding only a linear relation between the input voltage and the output current. In fact, the same techniques to obtain multipliers [32] of the type of $k_{1}v_{i}y$ are applicable to linear transconductors, where $k_{1}$ is a multiplication constant and one of the inputs $x$ or $y$ becomes a DC constant. Practical implementations of Fig. 6 are shown in Fig. 7. The transconductance [11, 33, 34] of Fig. 7a must operate its bottom transistors in the ohmic region, and the top driving transistors in saturation. For proper operation
of this transconductor the input signals should have a suitable DC bias voltage. Furthermore, a variation of this transconductor can be obtained by applying the input signals to the bottom transistors with appropriate DC gate bias for all transistors, to keep top and bottom transistors operating in saturation and ohmic regions, respectively. The transconductor \cite{39} of Fig. 7b is an example of the implementation of structure of Fig. 6b; the floating voltage source \(V_{f}\) can be implemented in several practical ways \((\cite{21}, Section 7.22)\) including a simple source follower.

Comparison of linearisation techniques employing source degeneration \((\cite{23, 41-45})\) are often used. Figs. 8a and b illustrate two possible implementations. Although both topologies realise the same transconductance, they present different properties. Fig. 8a, the noise contribution of the current sink is divided in both branches appearing at the outputs as common-mode noise. For the structure in Fig. 8b, the noise of each current sink is injected to a single output, appearing as a differential noise current. On the other hand, the voltage drop at the resistors of Fig. 8a reduces the common mode swing of the input signals; this is particularly critical for low-voltage applications. For the source degenerated structure, the output current is related to the input voltage by the following relationship \(\cite{44}\):

\[
\begin{align*}
\hat{i}_{o,sd} & = \sqrt{1 - \left(\frac{v_{id}}{2(1 + N)V_{DSsat}}\right)^2} \\
& \times \frac{2\mu_c C_{ox} I_B W_{n}}{1 + N} v_{id}
\end{align*}
\]

where \(N = G_r R\) is the source degeneration factor. This expression can also be used for the conventional differential pair with \(N = 0\). From this equation, the small-signal transconductance and the third harmonic distortion can be found; the resulting expressions for the elementary differential pair and source degeneration structure are given in Fig. 9. While the linearisation scheme reduces the small signal transconductance by \(1 + N\), the third harmonic distortion is reduced by the square of the same factor. Note that increasing the source degeneration factor the harmonic distortion is reduced even if the saturation voltage is limited. This additional degree of freedom is an important advantage of these structures. The derivation of the input referred noise is tedious, especially for source degenerated topologies. The results for Fig. 8b are given in Fig. 9. In those expressions, \(g_{mP}\) and \(g_{mN}\) are the small-signal transconductance of the transistors used as P-type and N-type current sources, respectively. For a lossless integrator and if the input referred noise density is integrated up to the unity gain frequency, the linearised integrator’s dynamic range can be approximated as follows:

\[DR_{sd} \approx \sqrt{(HD3)(CL)(1 + N)^{\frac{11}{2}}V_{DSsat}}\]

where the noise factor \(NF_{sd}\) is

\[NF_{sd} = 1 + \frac{g_{mP} + \left(\frac{N}{1 + N}\right)^2 g_{mN}}{G_m}\]

These expressions apply to the simple differential pair based OTA with \(N = 0\). The source degeneration OTA noise factor is larger than that of the elementary differential based OTA, mainly due to the noise contribution of the N-type current sources, and can be maintained at low levels if
the small-signal transconductances of the current sources are reduced. Although the noise factor is slightly larger, the linear range is increased by a factor \((1 + N)\). More detailed discussions on noise are given in [46-48]. Regarding optimal dynamic range see [49]. In general, the source degeneration reduces the small-signal transconductance by the degeneration factor \((N)\). Scaling up both transistor dimensions and drain currents compensates this reduction. The benefits of the source degenerated structures are seen in terms of higher power consumption and additional silicon area. Sometimes, the bandwidth (BW) of linearised OTAs is also severely limited, especially if many additional nodes are introduced. The BW issue will be discussed in Section 4.

What follows is a discussion on the implementation of the resistors \(R\) of Fig. 8. These realisations involve MOS transistors either operating in the ohmic region or in saturation. Three popular implementations of the \(R\) for the linearised OTA are illustrated in Fig. 10. The advantages and disadvantages are summarised in Fig. 11. Observe that a combination of these linearisation techniques can be implemented in a circuit. Of course each addition of a linearisation scheme will improve the overall performance at the expense of a reduction of the transconductance gain. The authors in [44] use two source degeneration techniques and current addition to partially increase the linearised transconductance gain, yielding a well linearised OTA. Other authors have combined resistors with parallel combinations of transistors operating in triode and saturation regions [50, 51].

In real OTAs the transconductance gain has a finite bandwidth. This can be modelled as a first-order low-pass, i.e. \(G_m = G_{m,degb} \times (1 + s/BW)\). For open loop applications the resulting excess phase \(\phi_e = \omega_0B W\) can be compensated by either connecting two OTAs in parallel [17] with one of the OTAs with reverse input polarity; another approach for integrators consists of adding a resistor [42] in series with

---

### Table: Parameter Comparison

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Differential pair</th>
<th>Source degeneration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small-signal transconductance</td>
<td>(G_m = \frac{W}{L} \sqrt{\frac{1}{N}})</td>
<td>(G_{m,dl} = \frac{G_m}{1+N})</td>
</tr>
<tr>
<td>Third Harmonic Distortion (HD3)</td>
<td>(HD3 = \frac{1}{3 \times G_m} \left( \frac{v_{in}}{V_{DSAT}} \right)^2)</td>
<td>(\frac{1}{3 \times G_{m,dl}} \left( \frac{v_{in}}{V_{DSAT}} \right)^2)</td>
</tr>
<tr>
<td>Input referred thermal noise density</td>
<td>(\frac{16 kT}{3 \times G_m} \left( \frac{G_{m,dl}}{G_m} \right))</td>
<td>(\frac{16 kT}{3 \times G_{m,dl}} \left( \frac{G_{m,dl}}{G_m} \right))</td>
</tr>
<tr>
<td>Dynamic Range</td>
<td>(DR = \left(\frac{HD3}{C_m} \right) \times 10^6 V_{DSAT})</td>
<td>(\frac{1}{N} \left( N + 1 \right) \times 10^6 V_{DSAT})</td>
</tr>
<tr>
<td>Current consumption*</td>
<td>(2L)</td>
<td>(2(1+N)L)</td>
</tr>
<tr>
<td>Transistor dimensions*</td>
<td>(\frac{W}{L})</td>
<td>(\frac{(1+N)W}{L})</td>
</tr>
</tbody>
</table>

---

**Fig. 9** Main characteristics for differential pair based OTAs

* For comparison, same transconductance and same \(V_{DSAT}\) are fixed for both structures. \(NF = N_{F_{dl}}\) for \(N = 0\)

---

**Fig. 11** Properties of OTAs using source degeneration

---
the integrating capacitance. These techniques are illustrated in Fig. 12. The RC compensation shown in Fig. 12b consists of replacing \( R \) by a transistor operating in the triode (also called ohmic) region. Note that changing \( g_m \) requires adjustment of \( R \) in Fig. 12b. Another particular case of phase compensation consists of connecting an optimal value capacitor \([42]\) in parallel with the resistor associated with the source degeneration linearisation technique (Fig. 10a) discussed earlier.

Fig. 12 Phase compensation techniques for integrators
\( a \) Active  
\( b \) Passive

![Diagram showing phase compensation techniques for integrators](image)

Fig. 13 Two-integrator biquads
\( a \) Single-ended  
\( b \) Fully differential

4 Transconductance and current-mode filters

There are two common techniques \([8]\) to implement OTA-C filters: (a) cascade of biquads \([1, 8, 52, 53]\), (b) RLC emulation either by implementing the equations describing the passive prototype \([54]\) or by direct simulation of components \([6, 8]\) and \([21]\), chap. 10). The two-integrator loop biquad is one of the most popular structures. The single-ended and the fully differential versions are shown in Fig. 13. The structure provides a lowpass (LP) output at \( v_{o2} \); if a bandpass (BP) is required the input OTA \((g_{01})\) is injected instead to node \( v_{o1} \) into the node \( v_{o2} \). Many other combinations yielding different types of filters are possible and are well documented in the literature \([1, 7, 8, 21]\). The biquad in Fig. 13 has suitable properties for high-frequency applications. For the single-ended biquad observe that one could save an OTA by injecting the input signal to the positive terminal of the OTA \((g_{01})\), but this will cause a feedforward path through the input capacitance of the OTA \((g_{in})\) and the capacitance \(C_2\); this creates an undesirable capacitance voltage divider. For high-frequency applications, a number of parasitics \([55]\) and finite OTA bandwidth can affect the filter performance. For instance, for a second-order bandpass filter, the actual quality factor \(Q\) is limited by the finite output impedance (or equivalently the finite voltage gain) and the excess phase \((\phi_E)\) of the transconductance amplifier. Assuming equal OTAs in the filter, \(Q\) can be expressed as a function of the desired \(Q\), the DC voltage gain \((A_{dc})\) and the excess phase \((\phi_E)\), that is:

\[
Q = \frac{Q}{1 + 2 \left(\frac{1}{A_{dc}} - \frac{\phi_E}{Q}\right)}
\]

For the fully differential version the performance benefits are significant at the expense of increased power consumption and silicon area. Also the use of common-mode feedback circuits \([18]\) is often needed, although under some conditions, typically with all lossy integrators filters, this CMFB can be avoided \([19, 56]\).

Current-mode filters might be generated based on OTA-C filters. Assume in the OTA-C version that every OTA with a load \( Z \) at the output will be substituted by an input load \( Z \) followed by the OTA. In this last case the signals at the output and input are current. In the practical implementation of current-mode (CM) filters \([20, 57, 58]\) the transconductance is of the type shown in Fig. 1, thus they are pseudo-differential types, which usually involve cross-coupled connections to enhance their common-mode performance. The current-mode filters frequently operate at very high frequencies, but often suffer high-sensitivity and good layout transistor matching becomes a vital task.

Tuning: Critical IC filters are frequently based on resonant loops. For the two-integrator loop shown in Fig. 13, the resonant frequency and the filter bandwidth are given in Fig. 14, where the load of each integrator consists of a capacitor and an OTA with a finite output resistance. \(1/g_{01}\), \(1/g_{02}\) and \(1/g_{03}\) are the finite output resistances for OTA1, OTA2 and OTA3, respectively. In the case of resonant loops \((g_{03} = 0)\), the pole frequencies are not very sensitive to the OTA finite DC gain. Notice that even if the OTA DC gain \((g_{in}/g_{01})\) is only around 50, the frequency error is typically below \(1\%\). The non-dominant pole \((g_{03}, C)\) introduces excess phase in the integrators; fortunately, the resonant frequency has low sensitivity to these effects too. On the other hand, both OTA finite DC gain and non-dominant poles affect the filter bandwidth (see eqn. 11). For narrow-band applications \(g_{03}\) must be reduced, therefore the factor \((g_{03} + g_{02})/g_{01}\) increases, leading to large bandwidth errors (see Fig. 14). Usually cascade output stages reduce these errors. The effects of the non-dominant poles are quite important for high-Q filters even if the second pole is placed at very high frequencies. As an example, for \(g_{03}/g_{01} = 1000\) and \(Q = 10\) the bandwidth errors are in the range of \(20\%\) \([59]\).

Among the effects previously discussed, both temperature variations and process parameter tolerances affect the precision of OTA-C filters. The main characteristics of OTA-C filters are determined by the integrator’s time constant \(C/g_m\). Typical tolerances for both \(C\) and \(g_m\) are in the range of \(\pm30\%\), and these variations are uncorrelated, leading to very large variations in the filter characteristics. The accuracy of the OTA-C filters can be further improved by employing on-chip master–slave automatic tuning schemes \([3-6, 21-25, 43, 59-64]\). The basic idea behind these techniques is to extract the most important filter characteristics from a piece of additional hardware (the master system) and to lock them to stable and very well controlled external
references, assuming a good matching between the master and slave systems. Very often accurate clock frequencies already available in the system are employed. Most of the automatic tuning loops are based on phase-locked loops. A voltage controlled oscillator is employed; for a two integrator loop-based oscillator the oscillating frequency is given by $g_m/C$. This frequency is tracked to a clock frequency generated by an external crystal, as shown in Fig. 15a. From the error voltage the OTA small-signal transconductance is controlled; for most of the differential pair based OTAs the bias current is adjusted. For efficient tuning it is very important to minimise the mismatches between the master system and the main filter. Because OTA-C filters are sensitive to parasitic capacitors, the parasitics must be considered when the master system is designed. Another tuning scheme employs a second-order bandpass filter, as shown in Fig. 15b. In this tuning scheme the centre frequency of the BPF is tracked to the external frequency. For narrow-band filter applications the filter bandwidth must also be tuned. For this purpose, several approaches for Q-tuning and simultaneous frequency and bandwidth tuning have been addressed [22, 24, 59-64]. A Q-tuning technique yielding precision better than 1% for band-pass biquads is reported in [64]. In contrast to other Q-tuning techniques, in [64] no envelope detector circuits are involved. The Q-tuning technique involves a pseudo least mean square (LMS) implementation.

![Typical frequency tuning schemes](image)

**Fig. 15** Typical frequency tuning schemes

- a Based on a VCO
- b Based on a VCF

The matching between the main filter and the tuning system is better if both systems are loaded very close to each other and are as identical as possible. For high frequency applications, signals generated by the tuning system are fed through the substrate and parasitic capacitors and appear at the output of the main filter, reducing the filter signal to noise ratio. Shielding both the main filter and automatic tuning system reduces these signals. Other techniques use frequencies in the filter stop band to reduce these effects [25, 59].

## 5 Transconductance applications

Analogue multipliers play a very important role in several applications as mixers in communications, analogue multiplication for signal processors, adaptive schemes, programmable neural networks, and automatic control systems. Most of the high-frequency analogue multipliers are based on the popular Gilbert cell. It is based on two differential pairs biased by a third differential pair working as a voltage controlled current source. In fact, the Gilbert cell can be considered as an array of OTAs [32]. In the same paper, a number of different CMOS multiplier implementations are also discussed. A shortcoming of several analogue multipliers is the temperature dependence of the multiplication coefficient. Using an additional OTA can efficiently compensate these effects [65].

Other nonlinear operations [66] that generate arbitrary piecewise linear functions can also be implemented employing OTAs. As we discussed in previous Sections, for the tuning of OTA-C filters a control structure is employed. Based on these systems the realisation of automatic gain control systems is straightforward [67]. The OTA-based amplifier is composed of two transconductors. The voltage gain is very well controlled because it depends on the ratio of transistor dimensions and the ratio of bias currents. Both parameters can be controlled precisely in current CMOS technologies. By using a control loop driving the bias current (transconductance) of one of the OTAs, efficient and low-distortion AGC systems can be realised. OTA-C oscillators have also been proposed [21, 67].

OTA-C filters have been used in many practical applications. Usually, high-performance filters for intermediate frequencies, video [4, 10, 15, 23-25, 45, 61, 68, 69], and disk drive read channels [26, 70] employ this technique. In most of these papers several interesting circuits, including automatic tuning systems, are reported. The demand for higher frequency applications is moving toward faster continuous-time filters in the range of 100MHz and beyond, as noted in several recent published works [22, 71-75]. Nevertheless, some challenges still remain before continuous-time filters can be highly competitive at such high frequencies. Although many efficient tuning strategies have been reported, most of them are not efficient above 100MHz. Also, most of the linearisation schemes introduce parasitic poles, reducing their frequency response.

## 6 Conclusions

A brief summary of the operational transconductance amplifier has been given. Trade-offs of structures, technology implementation (CMOS, bipolar or BiCMOS), and speed are very much application dependent. Several of the design issues for high-performance continuous-time filters have been addressed. There are still many open problems in frequencies higher than 100MHz, and it is very challenging for frequencies of around a few gigahertz [76] where other non-conventional process technologies are employed.
Acknowledgment

WHEATLEY, C.F., and WITTLINGE, H.A.: ‘OTA obsoletes OP.


REFERENCES


61 KRAMMENACHÉR, F.: 'Design considerations in high-frequency CMOS transconductance amplifier (TAC) filters'. Proceedings of ISCAS'89, 1989, pp. 100-103


