Chapter 4

Area-Constrained Design of Differential Input Stage at Mid/Moderate Frequency Ranges

4.1 Introduction

In this chapter, the proposed methodology of designing the differential input stage amplifier using the concept of Figure of Merit at low operating frequencies in the previous chapter has been revalidated in the mid/moderate frequency ranges.

4.2 Figures of Merit

The concept of Figure of Merit was proposed in chapter 3 for the low frequency applications where flicker noise is the dominant noise. This figure of merit is not suitable for the first stage of a two-stage compensated operational amplifier operating in mid frequency ranges for two reasons. One, at these frequencies, thermal noise dominates the flicker noise and secondly, the pole of the first stage has to be shifted towards left in order to provide proper splitting of poles.

Here, two expressions for the figure of merit are proposed. Firstly, the same expression, as proposed before in chapter 3 with flicker noise replaced by thermal noise, as thermal noise will dominate the flicker noise at mid/moderate frequencies, given by

$$FoM2 = \frac{UGB * Ad}{IRN(th)} \tag{4.1}$$

where UGB is the unity-gain bandwidth, Ad is the differential dc gain and $IRN(th)$ is the input-referred thermal noise spectral density in the band of interest for the differential amplifier illustrated in Fig. 4.1.

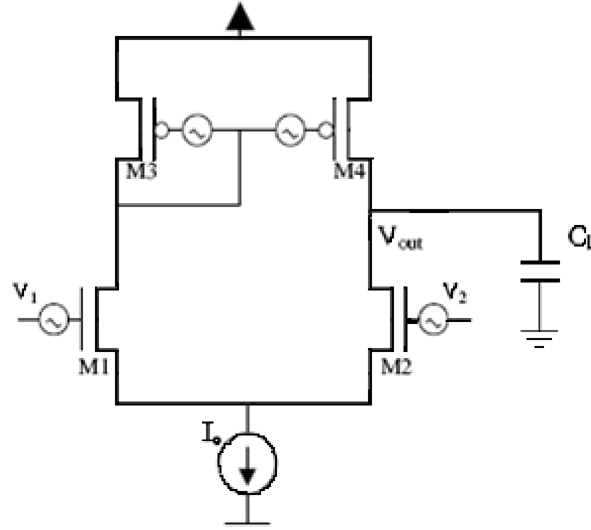


Figure 4.1 Differential amplifier.

Another figure of merit has been proposed for differential input stage of a two stage compensated operational amplifier operating at mid frequency that takes into account the maximization of differential gain with minimization of total mean square noise in the gain band and is proposed to be given by:

$$FoM3 = \frac{Ad}{UGB * (IRN(th))^2} \tag{4.2}$$

where UGB is the unity-gain bandwidth, Ad is the differential dc voltage gain and $(IRN(th))^2$ is the mean square input-referred thermal noise spectral density.

4.3 Thermal Noise Model for MOS Transistor

Each semiconductor device in the circuit introduces noise. Out of the various types of noises that could be possible in a device, thermal noise starts dominating at mid frequencies.

In a real resistor R, the electrons are in random thermal motion. As a result, a fluctuating voltage v_{nT} appears across the resistor even in the absence of a current from an external circuit as shown in Fig. 4.2(a). Thus, the Thévenin model of the real (noisy) resistor is that shown in Fig. 4.2(b). Clearly, the higher the absolute temperature (T) of the resistor, the larger is v_{nT}. In fact, it can be shown that the mean square of v_{nT} is given by

$$\overline{v_{nT}^2} = 4kTR\Delta f \ . \tag{4.3}$$

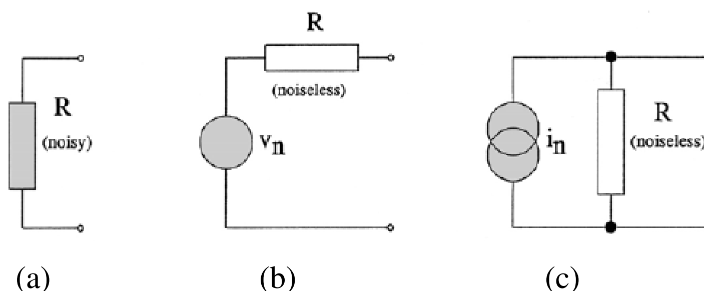


Figure 4.2 (a) Noisy resistor, (b) Thévenin equivalent circuit, (c) Norton equivalent circuit.

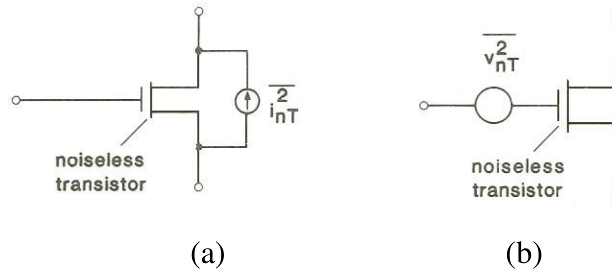
Here, k is the ubiquitous Boltmann's constant, and Δf is the bandwidth in which the noise is measured, in Hz.

If the above equation was true for any bandwidth, then the energy of the noise would be infinite. In fact, however, for very high frequencies ($\approx 10^{13}$) other physical phenomena enter, which cause $\overline{v_n^2}$ to decrease with increasing frequency so that the overall noise energy is finite.

The average value (dc component) of the thermal noise is zero. Since its spectral density $\overline{v_{nT}^2}/\Delta f$ is independent of frequency (at least for lower frequencies), it is a "white noise". Clearly, Fig. 4.2(b) may be redrawn in the form of a Norton equivalent, that is as a noiseless resistor R in parallel with a noise current source i_{nT} as shown in Fig. 4.2(c). The value of the latter is given by

$$\overline{i_{nT}^2} = \frac{4kT\Delta f}{R}. \tag{4.4}$$

Since the channel of a MOSFET in conduction contains free carriers, it is subject to thermal noise. Therefore, the two equations (4.3) and (4.4) will hold, with R given by the incremental channel resistance. The noise can then be modeled by a current source, as shown in Fig. 4.3(a). If the device is in saturation, its channel tapers off, and the approximation $R \cong \frac{3}{2g_m}$ can be used in the above equation.



(a) (b)

Figure 4.3 Equivalent models for the thermal noise in a MOS transistor.

In most circuits, it is convenient to pretend that i_{nT} is caused by a voltage source connected to the gate of an (otherwise noiseless) MOS transistor as shown in Fig. 4.3(b). This "gate referred" noise voltage source is then given by

$$\overline{v_{nT}^2} \cong \overline{(i_{nT}/g_m)^2} = \frac{8}{3}\frac{kT}{g_m}\Delta f. \tag{4.5}$$

Both i_{nT} and v_{nT} thus depend on the dimensions, bias conditions, and temperature of the device. If the device is switched off, then R becomes very high, and the equivalent noise current source $\overline{i_{nT}^2}$ is very small; hence, for usual (low or moderate) external impedance levels, the MOS transistor can be regarded as a noiseless open circuit if it is turned off.

In the SPICE2 circuit simulator the drain-current-referred thermal noise power spectral density (PSD) for MOS transistor utilizes the model of equation (4.5) and is given by

$$S_i = \frac{\overline{i_{nT}^2}}{\Delta f} = \left(\frac{2}{3}\right) 4kTg_m \tag{4.6}$$

This model is clearly invalid in the linear region [Tsividis84], because for the drain voltage $V_d = 0$, it predicts $S_i = 0$, whereas the PSD should be $4kTg_{d0}$, where g_{d0} is the drain conductance g_d at $V_d = 0$. There exists many analog applications *e.g.* switch-capacitor and MOS-C continuous-time filters, in which MOS transistor operates in linear region and hence the noise must be modeled in the linear region.

Another model [Nicollini87] defines the drain-current referred thermal noise power spectral density as

$$S_i = \alpha kT\left(g_m + g_{mb} + g_d\right) \tag{4.7}$$

where g_{mb} is the back-gate trans-conductance, and $\alpha = 1 - V_b/(3V_{dsat})$ for $V_d < V_{dsat}$ or $\alpha = 2/3$ for $V_d \geq V_{dsat}$. V_{dsat} is the V_d value at the border of the linear and saturation regions. The added term g_d ensures that the noise is nonzero in the linear region, and the factor α varies linearly from 1 (at $V_d = 0$) to $2/3$ (at $V_d = V_{dsat}$) to give the proper noise values in saturation and at $V_d = 0$. The serious deficiency of this model is that it is inconsistent with SPICE model levels due to

the g_{mb} term in equation (4.7) which can be a substantial fraction of $g_m + g_d$ [Fox93].

In the present work, the most popular model given by equation (4.6) has been used. According to it, the thermal noise due to a MOS transistor can also be lumped as a voltage source at its gate and can be written as

$$\overline{V_{eq}^2} = 4kT \cdot \left(\frac{2}{3g_m} \right) \cdot \Delta f \tag{4.8}$$

or

$$S_v = \frac{\overline{V_{eq}^2}}{\Delta f} = 4kT \cdot \left(\frac{2}{3g_m} \right) \tag{4.9}$$

in the noise bandwidth of Δf and g_m is the device trans-conductance at the operating point.

However, the overall circuit noise depends on the circuit configuration.

4.3.1 Input-Referred Noise of a Differential Input Stage

The differential input stage amplifier is shown in Fig. 4.1.

If the noise generated in the channel of each MOSFET is represented by a current source in parallel with the channel which can be represented by its equivalent input noise voltage generator as shown in Fig. 4.1, the equivalent input thermal noise voltage of the circuit, $\overline{V_{eqT}^2}$ at input (gate of M1) can be given by the following equation [Gray93].

$$\overline{V_{eqT}^2} = \overline{V_{eq1}^2} + \overline{V_{eq2}^2} + \left(\frac{g_{ml}}{g_{mi}} \right)^2 \cdot \left(\overline{V_{eq3}^2} + \overline{V_{eq4}^2} \right) \tag{4.10}$$

where $\overline{V_{eq1}^2}$, $\overline{V_{eq2}^2}$, $\overline{V_{eq3}^2}$, $\overline{V_{eq4}^2}$ are the noise sources at the gates of transistors M1, M2, M3 and M4. g_{mi} and g_{ml} are the transconductances of the input (M1 and M2) and load (M3 and M4) transistors, respectively, and are given by

$$g_{mi} = \sqrt{2 \cdot k_n \cdot \left(\frac{W_i}{L_i}\right) \cdot \left(\frac{I_o}{2}\right)}, \qquad \text{and}$$

$$g_{ml} = \sqrt{2 \cdot k_p \cdot \left(\frac{W_l}{L_l}\right) \cdot \left(\frac{I_o}{2}\right)}. \tag{4.11}$$

where W_i, W_l, L_i, and L_l are the widths and lengths of input and load transistors respectively, k_n and k_p are the process trans-conductance parameters for n-channel and p-channel MOS transistors, and I_o is the tail current of the differential amplifier.

Using equations (4.9), (4.10) and (4.11), the Power Spectral Density (PSD) of noise voltage at the gate of M1 is written as

$$S_{VG}^2 = \frac{\overline{V_{eqT}^2}}{\Delta f} = 4kT \frac{4}{3\sqrt{2 \cdot k_n \cdot (W_i/L_i) \cdot (I_o/2)}} \cdot \left[1 + \sqrt{\frac{k_p \cdot (W_l/L_l)}{k_n \cdot (W_i/L_i)}}\right] \tag{4.12}$$

Therefore, root mean square value of spectral power density better known as input-referred noise is written as

$$IRN = IRN(th) = \sqrt{S_{VG}^2} = \sqrt{4kT \frac{4}{3\sqrt{2 \cdot k_n \cdot (W_i/L_i) \cdot (I_o/2)}} \cdot \left[1 + \sqrt{\frac{k_p \cdot (W_l/L_l)}{k_n \cdot (W_i/L_i)}}\right]} \tag{4.13}$$

4.4 Formulation of Figure of Merit

Unity-gain bandwidth of the circuit, UGB is given by

$$UGB = \frac{g_{mi}}{2 \cdot \pi \cdot C_L} = \frac{1}{2 \cdot \pi \cdot C_L} \sqrt{\frac{k_n \cdot W_i \cdot I_o}{L_i}} \tag{4.14}$$

where C_L is the total load capacitance at the output node.

And the differential dc gain, Ad of the differential input amplifier given by

$$Ad = \frac{g_{mi}}{g_{di} + g_{dl}} = 2 \cdot \sqrt{\frac{k_n}{I_o} \cdot \left(\frac{W_i}{L_i}\right)} \cdot \left(\frac{1}{L_i} \left(\frac{dx_d}{dV_{DS}} \right)_n + \frac{1}{L_l} \left(\frac{dx_d}{dV_{DS}} \right)_n \right)^{-1} \tag{4.15}$$

where g_{di} and g_{dl} are the drain conductances of input and load transistors, respectively. The drain conductance g_d is approximated as

$$g_d = \frac{I_o}{2.L} \cdot \left(\frac{dx_d}{dV_{DS}} \right) \tag{4.16}$$

where $\left(\frac{dx_d}{dV_{DS}} \right)$ (known as channel-length modulation parameter) is a process parameter [Gray92] and its value has been taken as 0.1 µm/V for nMOS and 0.05 µm/V for pMOS transistors for the 1.25 µm CMOS technology.

Substituting the values of UGB, Ad and IRN from equations (4.13), (4.14) and (4.15) in equations (4.1) and (4.2), we get

$$FoM\ 2 = \frac{\sqrt{3} I_o^{1/4} k_n^{7/4} (W_i L_i)^{7/4}}{4\pi (kT)^{1/2} C_L L_i^{3/2}} \cdot \left[\left(\frac{dx_d}{dV_{DS}} \right)_n + \frac{L_i}{L_l} \left(\frac{dx_d}{dV_{DS}} \right)_p \right]^{-1} \cdot \left(1 + \frac{L_i}{L_l} \cdot \sqrt{\frac{k_p W_l L_l}{k_n W_i L_i}} \right)^{-1/2} \tag{4.17}$$

$$FoM\ 3 = \frac{3\pi C_L k_n^{1/2} (W_i \cdot L_i)^{1/2}}{4kT\sqrt{I_o}} \cdot \left[\left(\frac{dx_d}{dV_{DS}} \right)_n + \frac{L_i}{L_l} \left(\frac{dx_d}{dV_{DS}} \right)_p \right]^{-1}$$

$$\cdot \left(1 + \frac{L_i}{L_l} . \sqrt{\frac{k_p W_l L_l}{k_n W_i L_i}} \right)^{-1} \tag{4.18}$$

4.5 Maximization of Figure of Merit Under Area Constraints

If A is the total area available for the devices in the differential amplifier, let us assign fraction x of A *i.e.* $x \cdot A$ to the input transistors and $(1-x) \cdot A$ to load transistors. Then, writing the expressions for UGB, Ad and IRN in terms of x, area (A), bias current (I_o) and technology parameters, we get

$$A_d = 2.\sqrt{\frac{2 \cdot k_n \cdot (x \cdot A)}{I_o}} \cdot \left(\left(\frac{dx_d}{dV_{DS}} \right)_n + \frac{L_i}{L_l} \cdot \left(\frac{dx_d}{dV_{DS}} \right)_p \right)^{-1} \tag{4.19}$$

$$UGB = \frac{\sqrt{k_n . I_o \cdot (x \cdot A)}}{2\sqrt{2}.\pi \cdot C_L} \cdot \frac{1}{L_i} \tag{4.20}$$

$$IRN^2 = \frac{16.k.T}{3.\sqrt{k_n.(x.A).I_o}} . \frac{1}{L_i} . \left(1 + \frac{L_i}{L_l} . \sqrt{\frac{k_p.(1-x)A}{k_n.(xA)}} \right) \tag{4.21}$$

$$IRN = \left[\frac{16.k.T}{3.\sqrt{k_n.(x.A).I_o}} . \frac{1}{L_i} . \left(1 + \frac{L_i}{L_l} . \sqrt{\frac{k_p.(1-x)A}{k_n.(xA)}} \right) \right]^{1/2} \tag{4.22}$$

Hence, from equations (4.1), (4.2), (4.19), (4.20), (4.21) and (4.22) figure of merit $FoM\,2$ and $FoM\,3$, in terms of x, area (A), bias current (I_o) and technology parameters can be written as

$$FoM\ 2 = \frac{\sqrt{3} I_o^{1/4} k_n^{7/4} (xA)^{7/4}}{4\pi (kT)^{1/2} C_L L_i^{3/2}} \cdot \left(1 + \frac{L_i}{L_l} . \sqrt{\frac{k_p (1-x) A}{k_n x A}}\right)^{-1/2}$$
$$\cdot \left[\left(\frac{dx_d}{dV_{DS}} \right)_n + \frac{L_i}{L_l} \left(\frac{dx_d}{dV_{DS}} \right)_p \right]^{-1} \qquad (4.23)$$

$$FoM\,3 = \frac{3\pi . C_L \cdot k_n^{1/2} \cdot (xA)^{1/2}}{4kT \cdot \sqrt{I_o}} . \left(1 + \frac{L_i}{L_l} . \sqrt{\frac{k_p .(1-x) A}{k_n .(xA)}}\right)^{-1}$$
$$\cdot \left[\left(\frac{dx_d}{dV_{DS}} \right)_n + \frac{L_i}{L_l} \left(\frac{dx_d}{dV_{DS}} \right)_p \right]^{-1} \qquad (4.24)$$

From equations (4.23) and (4.24), following conclusions can be drawn

- Figures of merits, $FoM\,2$ and $FoM\,3$ are dependent on technology parameters *i.e.* process transconductance parameters, k_n for n-channel transistor, k_p for p-channel transistor; and device channel length modulation parameters, $\left(\frac{dx_d}{dV_{DS}} \right)_n$ for n-channel transistor and $\left(\frac{dx_d}{dV_{DS}} \right)_p$ for p-channel transistor.
- To maximize $FoM\,2$ and $FoM\,3$ for a given load C_L, the length of input transistor L_i should be kept minimum for a given area (A), as it maximizes all the three product terms where it appears.
- Length of load transistor L_l should be as large as possible under the constraints of area since it maximizes the last two product terms and hence maximizes Figures of merits $FoM\,2$ and $FoM\,3$.
- Figures of merits are independent of the width of the load transistor, W_l and hence it should be kept at minimum.

- Figure of merit, $FoM2$ increases with the increase in the bias current whereas Figure of merit $FoM3$ decreases with the increase in the bias current.
- Figure of merit, $FoM2$ is a strong function of the area allocated to input transistors (xA), whereas Figure of merit, $FoM3$ is relatively a weaker function of the area allocated to input transistors (xA).

4.6 Analytical Results

The above analytical formulations have been used to draw the following curves.

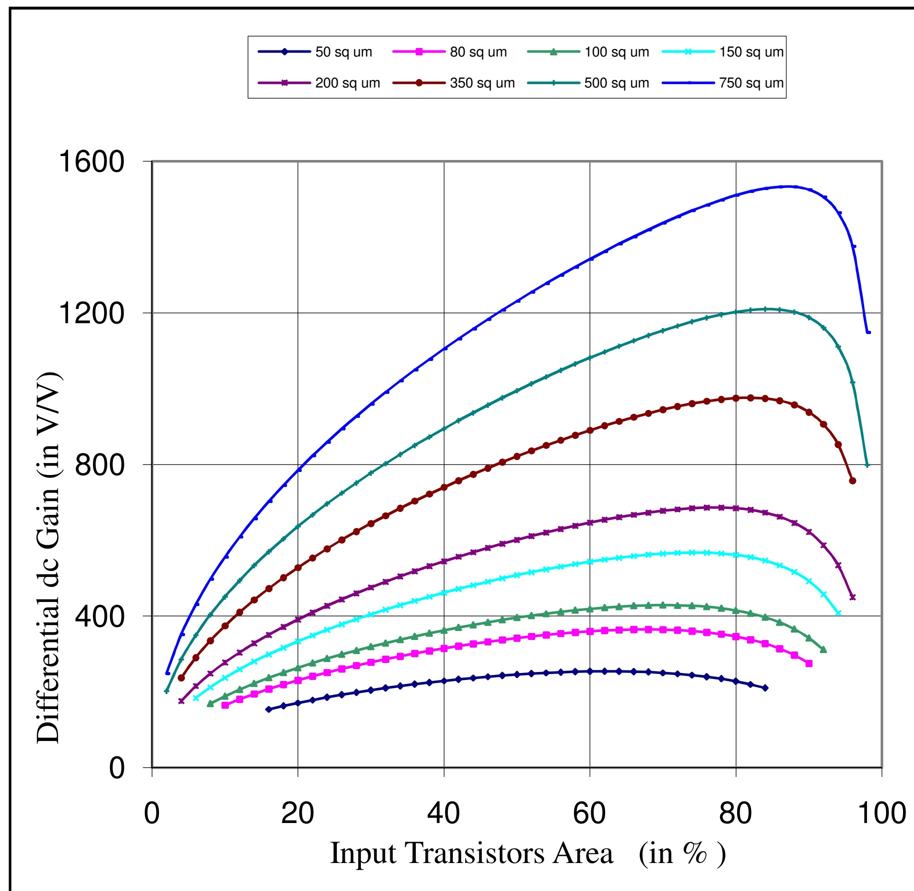


Figure 4.4 The differential dc voltage gain as a function of input transistor area.

The variation of differential dc voltage gain as a function of relative area allocated to input transistors (xA) at different values of total area, A is shown in Fig. 4.4. It is clear that it increases with increasing area allocation to input transistor and peaks at a certain value of x. Also, as the total area increases, the peak value of differential voltage gain increases. The peak value of voltage gain is obtained for x in the range of 0.64 to 0.86.

Fig. 4.5 shows that the unity-gain bandwidth is a monotonically increasing function of the relative area allocated to the input transistors x in percent.

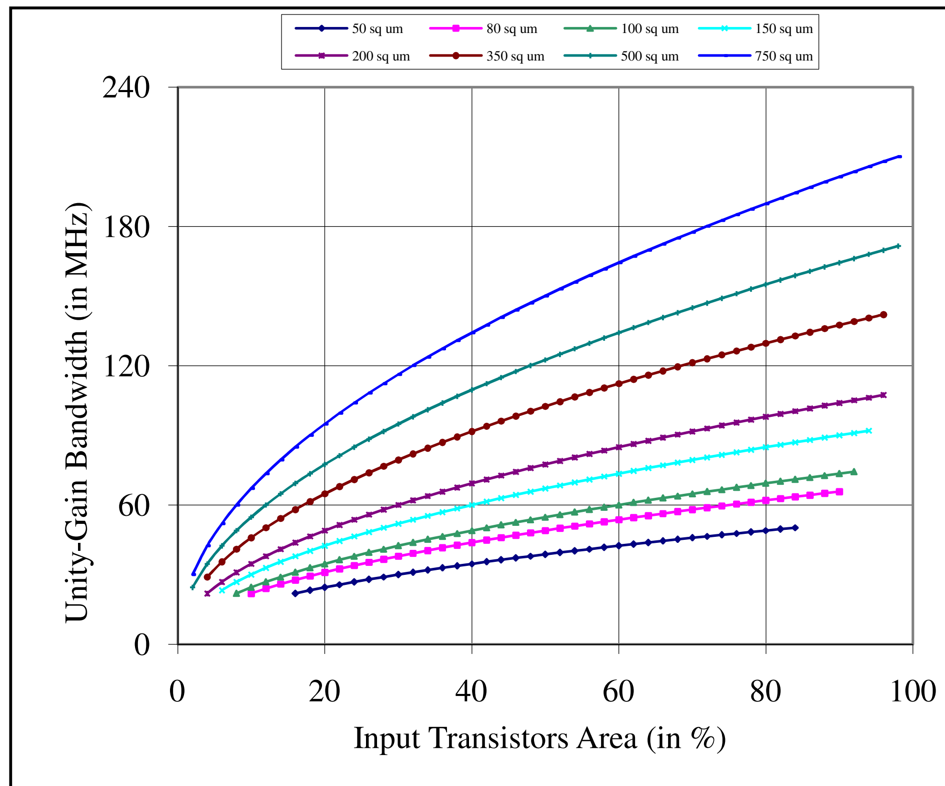


Figure 4.5 Unity-Gain Bandwidth (in MHz) as a function of input transistor area.

The variation of input-referred noise as a function of x is shown in Fig. 4.6. For larger values of x, the noise is reducing initially, because with increasing x the gate area of input transistors is increasing and their noise contribution is decreasing. But as the x increases beyond 0.4 the contribution of noise from load

transistors increases and starts dominating the noise contribution at the gate of transistor M1.

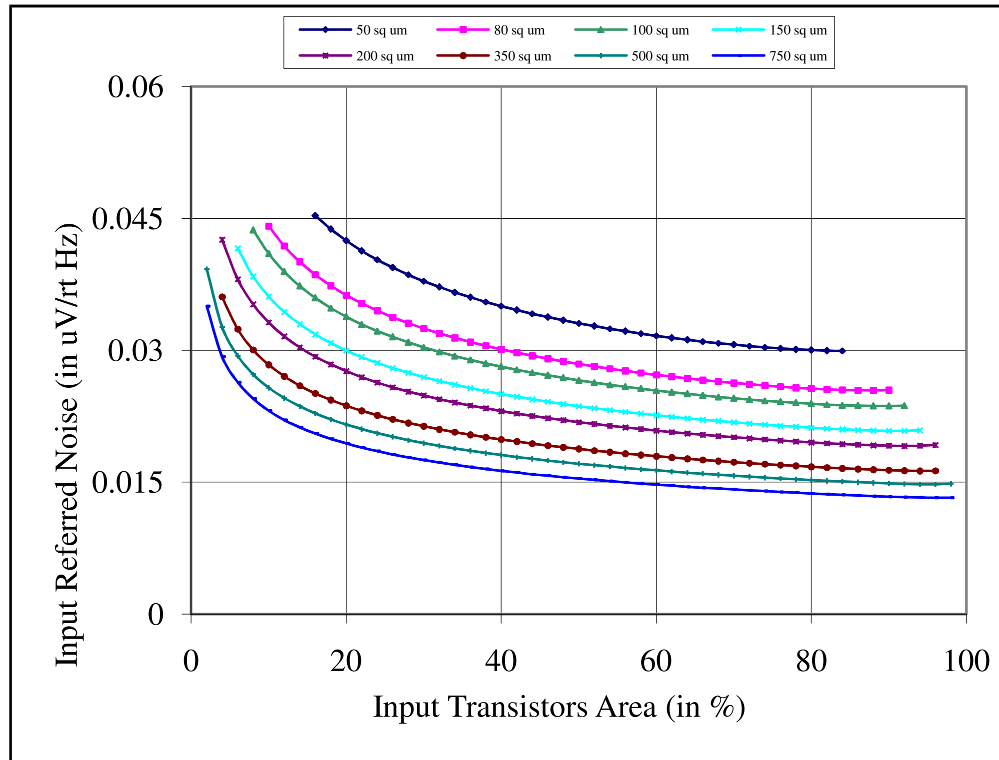


Figure 4.6 Input-Referred Noise (in nV/rtHz) as a function of input transistor area.

Fig. 4.7 shows the total band-noise *i.e.* the product of unity-gain bandwidth and input-referred thermal noise (mean square value) as a function of x for different values of A. It is clearly observed that it becomes almost constant for x greater than about 0.4 in all cases. It is interesting to note that the total band-noise does not change significantly with the total area A assigned to the circuit. It implies that by allocating more area to the input transistors, the unity-gain gain bandwidth of the circuit can be increased with reduced input referred noise such that the total band noise is almost the same.

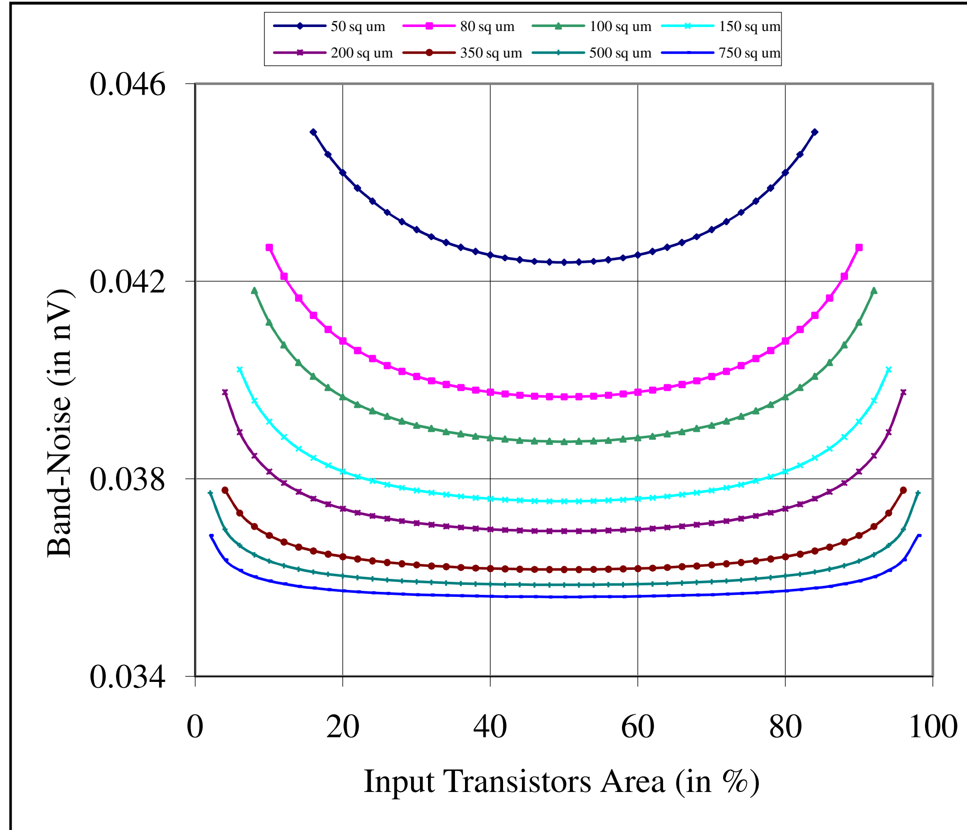


Figure 4.7 Total band-noise (in nV) as a function of input transistor area.

The figure of merit, $FoM2$ is plotted as a function of percentage area allocated to input transistors as shown in Fig. 4.8. It is a peaking function of x in the range 80% to 90%.

Next the figure of merit, $FoM3$ is plotted as a function of percent area allocated to input transistors. Figure of Merit, $FoM3$ is also a peaking function of x in the range 64% to 84% as shown in Fig. 4.9. The $FoM2$ is a much stronger function of x than $FoM3$. It is clear that both figures of merit increase with total area A. It also indicates that for a fixed value of total area, how much percentage of total area should be assigned to input transistors to obtain a maximum value of a figure of merit. Peak value of figure of merit, $FoM3$ as a function of total area is plotted in Fig. 4.10.

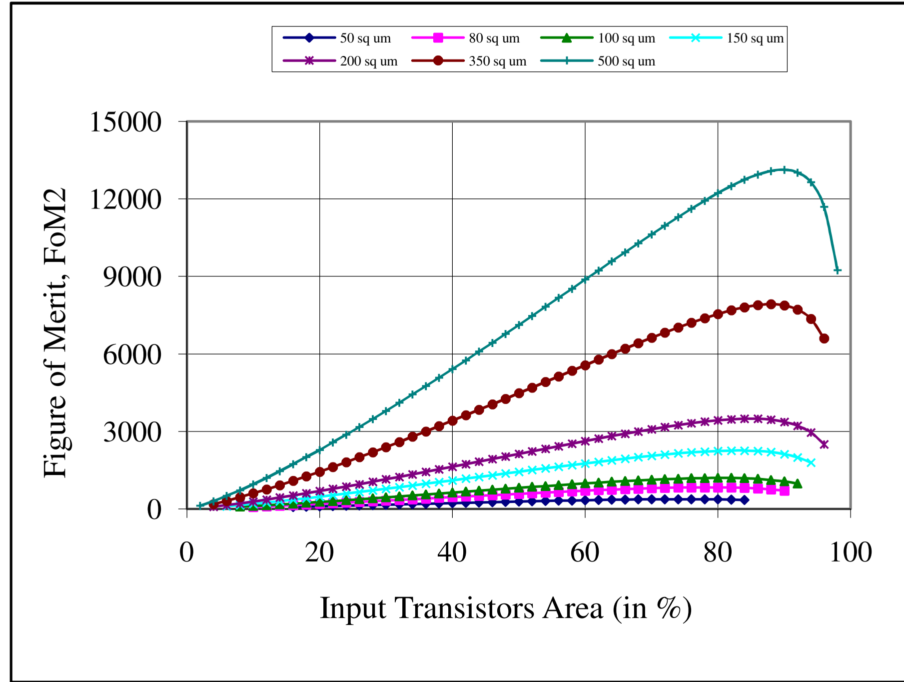


Figure 4.8 The $FoM2$ as a function of input transistor area.

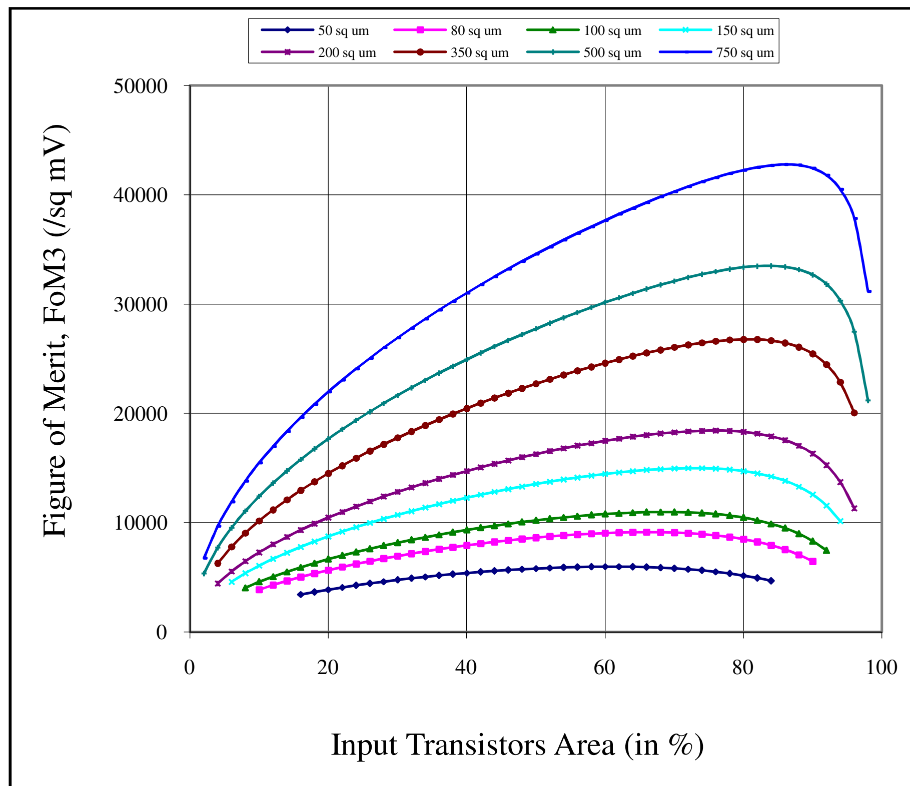


Figure 4.9 The $FoM3$ as a function of input transistor area.

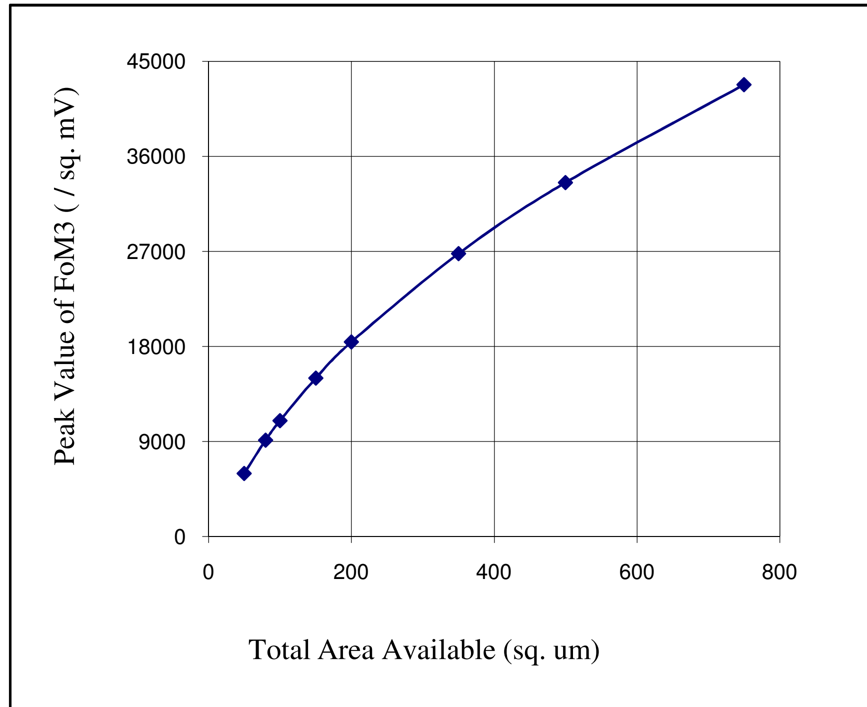


Figure 4.10 Peak value of $FoM3$ as a function of total area.

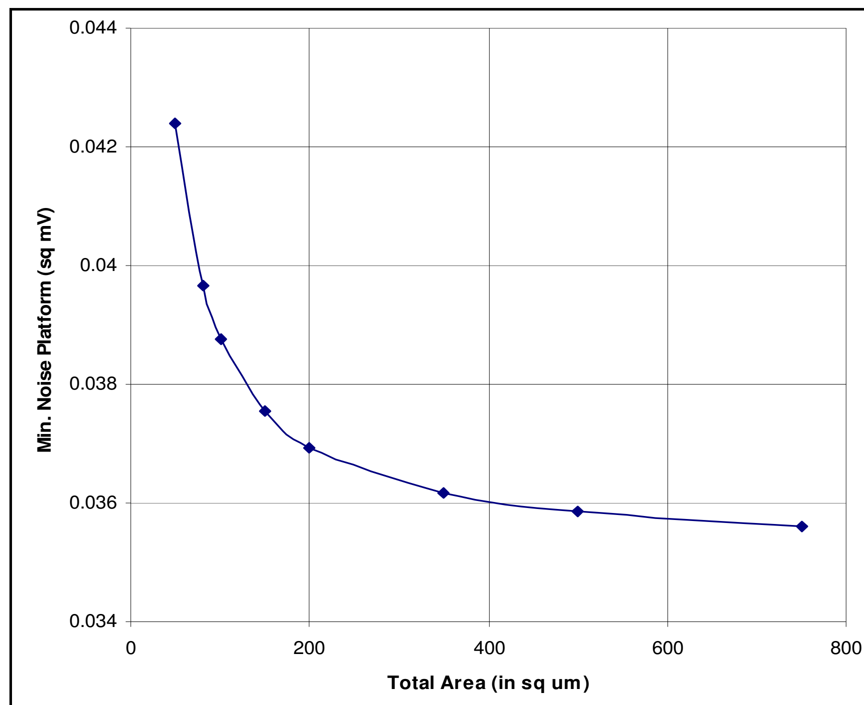


Figure 4.11 Minimum noise platform as a function of total area.

Fig. 4.11 indicates that for a fixed value of total area, there is a minimum value of band noise present in the circuit. It implies that if the total area of the circuit is reduced the noise platform may increase exponentially beyond a point. It is also clear that one cannot lower the noise platform beyond a limit, whatever large area is assigned to the circuit.

4.7 Simulation Results

Circuit simulations using SPICE, also validated the analytical results. To perform this task, a value of total area was chosen. The total area was divided between input and load transistors in a predefined ratio. Then for this distribution of areas a large number of combinations of aspect ratio of input and load transistors were simulated to obtain the differential dc voltage gain, unity-gain bandwidth and input-referred thermal noise.

The total band noise and the figures of merit, $FoM2$ and $FoM3$ were computed from these parameters. The total band-noise and figures of merit as a function of input transistor area in percent (ratio of area allocated to input transistors to the total available area, A) for various values of total area are plotted as shown in Figs. 4.12, 4.13 & 4.14, respectively. These plots match well with the analytical results shown in previous section.

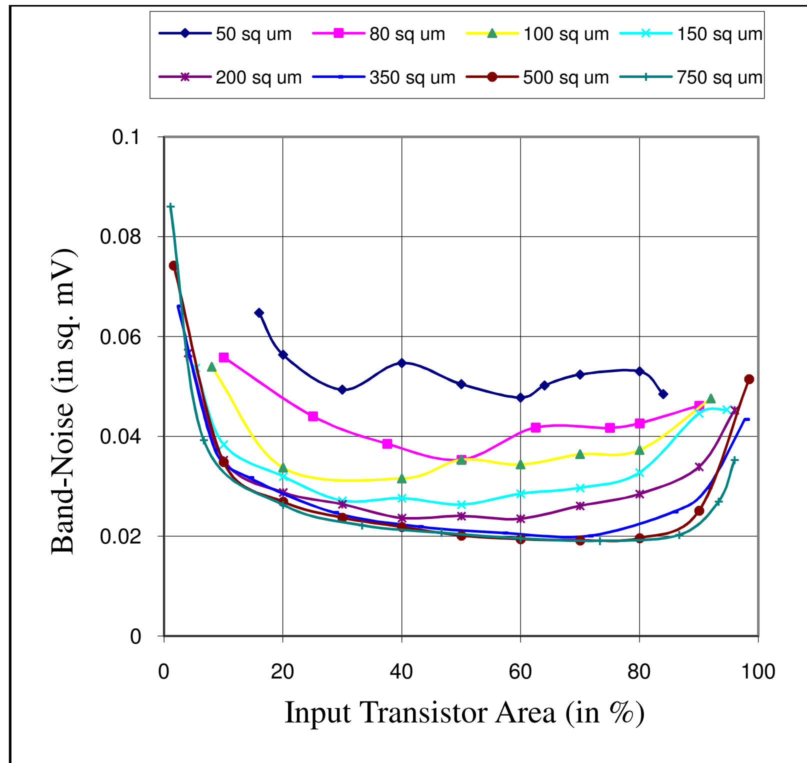


Figure 4.12 Simulated plot of total band-noise as a function of area.

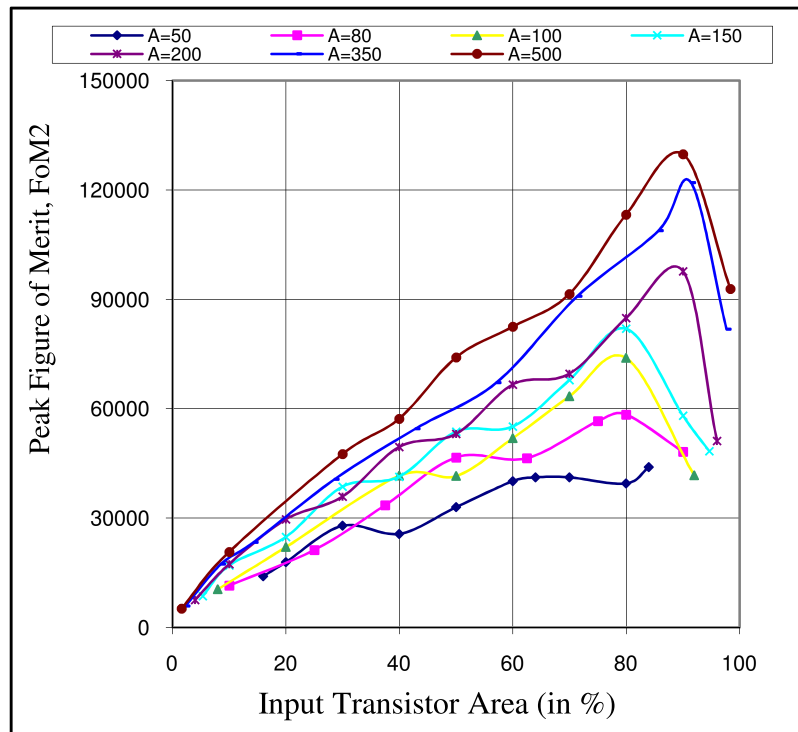


Figure 4.13 Simulated plot of figure of merit, $FoM2$ as a function of area.

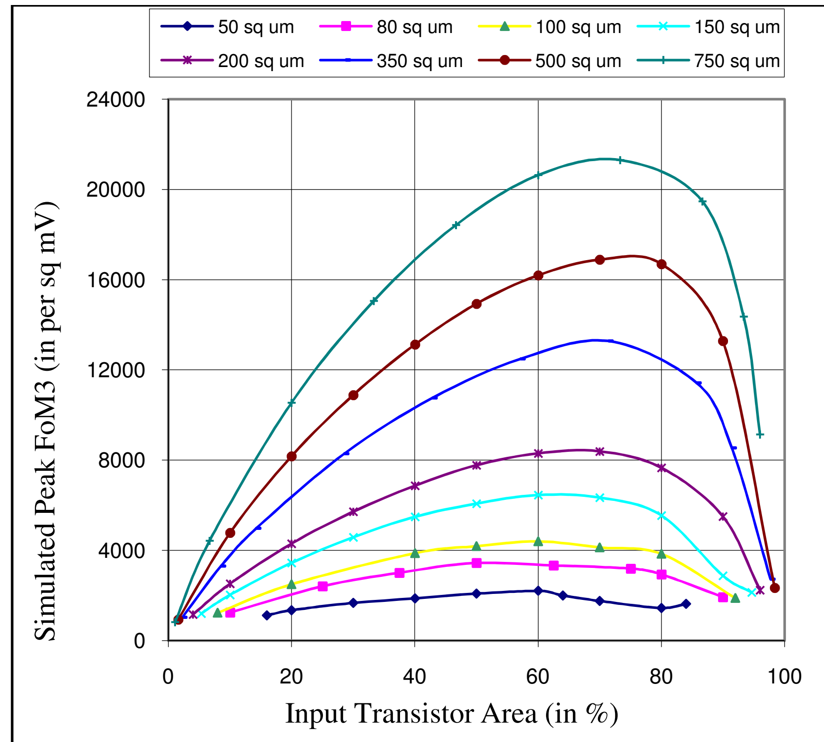


Figure 4.14 Simulated plot of figure of merit, $FoM3$ as a function of area.

In order to demonstrate the utility of figure of merit as a tool to optimize the design, we define $BN\%$ -- Band Noise obtained at the peak figure of merit as a percentage of minimum band-noise achievable for a given area in addition to previously defined parameters $Ad\%$, $UGB\%$ and $IRN\%$ (in chapter 3).

Table 4.1 compares the analytical and simulated values of peak figure of merit $FoM2$, $Ad\%$, $UGB\%$ and $IRN\%$ for constant area. Table 4.2 compares the analytical and simulated values of peak figure of merit, $FoM3$, $Ad\%$, $BN\%$ for constant area. The analytical and simulated values of voltage gain, unity-gain bandwidth and input-referred noise at the respective peak values of figure of merit are close to the best achievable values at that area.

This validates that the proposed methodology can be deployed in CAD tools for the automatic synthesis of optimal circuit of differential amplifier under the constraint of area.

Table 4.1 Comparison Between Analytical and Simulated Performance at Peak Figure of Merit, $FoM2$.

Area (μm^2)	Peak Figure of Merit, FoM2		Ad%		UGB%		IRN%	
	Analytical	Simulated	Analytical	Simulated	Analytical	Simulated	Analytical	Simulated
50	380	44036	96.1	100	93.86	100	101.34	100
80	838	58404	96.64	100	93.1	96.38	101.57	100
100	1202	73893	96.72	96.87	93.25	91.62	101.1	100
150	2263	81954	97.86	94.74	93.4	86.3	101.2	100
200	3496	97730	98.11	99.94	93.55	89.95	101.15	100
350	7924	121967	98.07	99.2	95.74	80.54	100.86	100
500	13131	129797	98.16	93.84	95.83	71.78	102.31	100.57
750	23033	181063	98.21	100	96.89	59.7	100.59	100

Table 4.2 Comparison Between Analytical and Simulated Performance at Peak Figure of Merit, $FoM3$.

Area (μm^2)	Peak Figure of Merit, FoM3 (per sq. mV)		Ad%		BN%	
	Analytical	Simulated	Analytical	Simulated	Analytical	Simulated
50	5973.92	2202	99.84	95.69	100.31	100
80	9136.22	3438.2	99.92	77.93	100.63	100
100	10981.2	4397.7	99.88	83.5	100.67	100
150	14990.2	6454.6	99.85	79.97	100.71	100
200	18423.9	8385.4	99.99	82.72	100.84	100
350	26776.3	13279	99.9	83.51	100.72	115.3
500	33501.5	16885	100	87.45	100.74	120.7
750	42781.4	21303	100	87.0	100.60	116.7