A Compact Fully-Differential Distributed Amplifier With Coupled Inductors in 0.18-μm CMOS Technology

Keisuke KAWAHARA††, Student Member, Yohtaro UMEDA††, Member, Kyoya TAKANO††, Member, and Shinsuke HARA†††, Member

SUMMARY This paper presents a compact fully-differential distributed amplifier using a coupled inductor. Differential distributed amplifiers are widely required in optical communication systems. Most of the distributed amplifiers reported in the past are single-ended or pseudo-differential topologies. In addition, the differential distributed amplifiers require many inductors, which increases the silicon cost. In this study, we use differentially coupled inductors to reduce the chip area to less than half and eliminate the difficulties in layout design. The challenge in using coupled inductors is the capacitive parasitic coupling that degrades the flatness of frequency response. To address this challenge, the odd-mode image parameters of a differential artificial transmission line are derived using a simple loss-less model. Based on the analytical results, we optimize the dimensions of the inductor with the gradient descent algorithm to achieve accurate impedance matching and phase matching. The amplifier was fabricated in 0.18-μm CMOS technology. The core area of the amplifier is 0.27 mm², which is 57% smaller than the previous work. Besides, we demonstrated a small group delay variation of ±2.7 ps thanks to the optimization. The amplifier successfully performed 30-Gbps NRZ and PAM4 transmissions with superior jitter performance. The proposed technique will promote the high-density integration of differential traveling wave devices.

key words: Distributed amplifier, Differential amplifier, Coupled inductor, CMOS integrated circuit, Optimization, Optical fiber communication.

1. Introduction

The demand for information and communication technologies is increasing year by year due to the spread of cloud services. Recent communications are based on digital coherent optical transmission systems. These systems require analog circuits with high operating speeds of more than 100 Gbaud. Since advanced CMOS processes such as 22 nm have maximum oscillation frequencies $f_{\text{max}}$ around 200 GHz, a bandwidth wider than 0.25$x f_{\text{max}}$ is required for 100 Gbaud transmissions. Distributed amplifiers have broad bandwidth, high output power, and excellent input-output matching.

These characteristics meet the requirements. Some distributed modulator drivers reached bandwidths of nearly 100 GHz [1]–[3]. Recently, distributed drivers were also developed for quantum key transfer systems [4]. Another application of distributed amplifiers is frequency interleaving. In that case, distributed amplifiers are used for signal combining and frequency conversion [5]–[6]. Distributed amplifiers are, thus, an essential component in the development of future communication technologies. On the other hand, most of the circuits used in optical transceivers have differential topology. Differential circuits have advantages in common-mode noise rejection and DC coupled cascading. Nevertheless, most of the distributed amplifiers reported in the past are single-ended [7]–[10] or pseudo-differential topologies [11]–[14]. This is because distributed amplifiers require many inductors or transmission lines to form artificial transmission lines at the input and output. In differential topology, a total of four artificial transmission lines are required, which not only consumes a large chip area but also makes layout design difficult. Although the pseudo-differential topology can simplify the layout, it does not provide the aforementioned advantages. Besides, the previous differential distributed amplifiers consume core areas of more than 0.5 mm² [3], [15], which raises an issue of manufacturing cost. Inductive coupling is an effective way to reduce the number and areas of inductors. For example, small-area laser drivers using coupled inductors have been reported [16]–[17]. For distributed circuits, some transversal filters and mixers adopt coupled inductors [18]–[20]. In this study, we developed a small-area differential distributed amplifier using coupled inductors. We set target specifications to bandwidth $>0.25$x $f_{\text{max}}$ and core area $<0.5$ mm² for applying to state-of-the-art optical fiber communication systems. Fig. 1 shows a schematic of the proposed amplifier. Coupled inductors reduce the area to about half that of conventional amplifiers and enable a simple layout design. A challenge to using coupled inductors is that capacitive parasitic coupling between the two inductors affects the characteristics of the lines. Undesirable capacitive coupling lowers the impedance of the artificial transmission lines and causes reflections at the termination. This not only deteriorates input-output matching but also leads to increased group delay and gain variations at low frequencies [21]. The parasitic coupling also changes the phase constant of the lines, which causes a roll-off in the
gain response due to a phase mismatch between the gate and drain lines. This study derives odd-mode image parameters of the differential artificial transmission lines. The lines are optimized by a gradient descent algorithm based on the analytical solution. In recent circuit design, simulation-based optimization is widely used. Although simulation-based optimization works well when accurate device models are available, it requires significant effort for measurement and modeling. We used a simple loss-less model of the inductors extracted by electromagnetic field analysis. This reduces design time and effort, and it ensures the explainability of the circuit operation. This paper is an extended version of [22]. We added an explanation of circuit analysis with image parameters and a detailed procedure for modeling and optimization. The effectiveness of the proposed technique was demonstrated by experiments and comparison with state-of-the-art research.

2. Analysis and Design

2.1 Analysis

Fig. 2(a) shows a small signal equivalent circuit of the amplifier cell, where $C_{gs}$, $C_{gd}$, and $C_{ds}$ are gate-source capacitance, gate-drain capacitance, and drain-source capacitance of the MOSFETs. $C_X$ is a cross-coupled capacitor that makes the amplifier unilateral for a differential signal [23]. Assuming a differential input signal, we can consider an odd-mode half circuit shown in Fig. 2(b). The voltage at node $Z$ equals zero, and it can be regarded as a virtual ground. The half circuit is a paralleled connection of a hybrid-pi model of the MOSFETs and cross-coupled capacitors. The admittance matrix of the half circuit is given by

$$Y_{\text{half}} = \begin{bmatrix} j\omega(C_{gs} + C_{gd} + C_X) & j\omega(−C_{gd} + C_X) \\ g_m + j\omega(−C_{gd} + C_X) & j\omega(C_{ds} + C_{gd} + C_X) \end{bmatrix}. \tag{1}$$

When $C_X = C_{gd}$, then

$$Y_{\text{half}} = \begin{bmatrix} j\omega(C_{gs} + 2C_{gd}) & 0 \\ g_m & j\omega(C_{ds} + 2C_{gd}) \end{bmatrix}. \tag{2}$$

The amplifier becomes unilateral due to the cross-coupled capacitors. The input miller capacitance without $C_X$ is $(1 + A_v)C_{gs}$, with a voltage gain of $A_v$ [24], and thus, $C_X$ decreases the input capacitance when $A_v > 1$. On the other hand, the output miller capacitance is given by $(1 + 1/A_v)C_{gd}$, and thus, $C_X$ slightly increases the output capacitance. However, it is almost no problem, because the bandwidth of the distributed amplifier is generally limited by the gate line, not the drain line. The important advantage of using $C_X$ is that the drain line and gate line can be separately designed. It allows a designer to optimize the inductors as discussed later.

Fig. 3 shows a layout and loss-less model of a coupled inductor. $L$, $M$, $C_{OX}$, and $C_M$ are self-inductance, mutual inductance, the capacitance between the winding wire and substrate, and mutual capacitance, respectively. This loss-less model is consistent with ignoring all the parasitic resistances in the model used in [25]. The same was also used in [26]. The sign of the mutual inductance is typically chosen to be negative so that the magnetic flux is enhanced by the differential currents. Fig. 4(a) shows an equivalent circuit of the gate line. We assumed that the input signal is in odd-mode and the amplifier cell can be regarded as a shunt capacitor $C_C = C_{gs} + 2C_{gd}$. This circuit is similar to the segmented model of a differential transmission line, and its odd-
The phase delay per section is given by

\[ \tau_G = \frac{2}{\omega_G} = \sqrt{(L - M)(C_G + 2C_{OX} + 4C_M)}. \]  

The same holds for the drain line by replacing \( C_G \) to \( C_D \), and \( Z_D \) and \( \tau_D \) are obtained. To minimize linear distortion, impedance and phase matching must be ensured. In other words, \( Z_G = Z_D = R_T \) and \( \tau_G = \tau_D \) must hold.

### 2.2 Modeling of the Coupled Inductor

The values of \( L, M, C_{OX}, \) and \( C_M \) vary depending on the dimensions of the coupled inductor such as a diameter \( d \), line width \( w \), and line space \( s \). These relationship needs to be clarified for design optimization. The four-port \( Y \)-parameters of the inductor are obtained by

\[ Y_{11} = j \left( \frac{-L}{\omega(L^2 - M^2)} + \omega(C_{OX} + C_M) \right), \]

\[ Y_{21} = j \left( \frac{M}{\omega(L^2 - M^2)} - \omega C_M \right), \]

\[ Y_{31} = j \left( \frac{L}{\omega(L^2 - M^2)} \right), \]

\[ Y_{41} = j \left( \frac{-M}{\omega(L^2 - M^2)} \right), \]

where \( Y_{11} = Y_{22} = Y_{33} = Y_{44} \), \( Y_{21} = Y_{12} = Y_{34} = Y_{43} \), \( Y_{31} = Y_{13} = Y_{42} = Y_{24} \), and \( Y_{41} = Y_{14} = Y_{32} = Y_{23} \). These are obtained by solving the simultaneous equations from (1) to (4). The following equations are obtained by

\[ L = \frac{\text{Im}(Y_{31})}{\omega \text{Im}(Y_{31})^2 - \omega \text{Im}(Y_{41})^2}, \]

\[ M = \frac{-\text{Im}(Y_{41})}{\omega \text{Im}(Y_{31})^2 - \omega \text{Im}(Y_{41})^2}, \]

\[ C_{OX} = \frac{\text{Im}(Y_{11}) + \text{Im}(Y_{31}) + \text{Im}(Y_{41}) + \text{Im}(Y_{21})}{\omega}. \]
Thus, the values of $L$, $M$, $C_{OX}$, and $C_M$ can be extracted from four-port $Y$-parameters that are obtained by electromagnetic analysis or measurement. In the following, a modeling procedure is explained with self-inductance as a representative, and the same holds for $M(d, s, w)$, $C_{OX}(d, s, w)$, and $C_M(d, s, w)$. The model formula for self-inductance is defined as a following quadratic polynomial.

$$L(d, s, w) \equiv \alpha_0 d^2 + \alpha_1 s^2 + \alpha_2 w^2 + \alpha_3 ds + \alpha_4 dw + \alpha_5 sw + \alpha_6 d + \alpha_7 s + \alpha_8 w + \alpha_9,$$  \hspace{1cm} (19)

where $\alpha_i (i \in \mathbb{N}, i < 10)$ are coefficients to be determined. We analyzed various inductors listed in Table 1 with Keysight’s Momentum electromagnetic simulator, and a vector of the self-inductance was obtained as follows.

$$L = \begin{bmatrix} L_1(d_1, s_1, w_1) \\ L_2(d_2, s_2, w_2) \\ \vdots \\ L_N(d_N, s_N, w_N) \end{bmatrix},$$  \hspace{1cm} (20)

where $N$ is the number of inductors. The model parameters minimizing the squared error between the model and extracted values can be determined by [28]

$$\min \begin{bmatrix} \hat{a}_0 \\ \hat{a}_1 \\ \vdots \\ \hat{a}_9 \end{bmatrix}^T = (D^T D)^{-1} D^T L,$$  \hspace{1cm} (21)

where $D$ is a matrix of the dimensions defined as

$$D = \begin{bmatrix} d_1^2 & s_1^2 & w_1^2 & d_1 s_1 & \ldots & w_1 \\ d_2^2 & s_2^2 & w_2^2 & d_2 s_2 & \ldots & w_2 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ d_N^2 & s_N^2 & w_N^2 & d_N s_N & \ldots & w_N \end{bmatrix}.$$  \hspace{1cm} (22)

The fitted model formula $\hat{L}(d, s, w)$ is obtained by substituting the model parameters $\hat{a}_i$ into (19). Fig. 5 shows a modeling result at $d = 120 \, \mu m$, $s = 6 \, \mu m$, and $w = 5 \, \mu m$. This dimension was not used for modeling. The broken lines and solid lines represent $Y$-parameters obtained from the electromagnetic analysis (EM) and model. The model agreed well with the EM from 10 GHz to 30 GHz.

### 2.3 Optimization

Impedance and phase matching must be ensured to design a differential distributed amplifier with superior performance. The goal of the optimization is to find dimensions that satisfy the impedance and phase matching conditions. Since these conditions are primarily determined by the imaginary part of the impedances, the use of a loss-less model is appropriate. The mismatch between the nominal impedance and terminating resistance causes reflection waves. These reflections degrade not only input and output matching but also the flatness of frequency response of gain and group delay [21]. That is to say, the following equation needs to be satisfied.

$$C_M = \frac{-\text{Im}(Y_{41}) - \text{Im}(Y_{21})}{\omega},$$  \hspace{1cm} (18)

On the other hand, a mismatch between $\tau_D$ and $\tau_G$ causes a roll-off like a transversal filter. The delay matching is corresponded to the cutoff matching by considering (10). Thus, the following equation needs to be satisfied.

$$\min |\tau_G - \tau_D| \Leftrightarrow \min |\omega_G - \omega_D|.$$  \hspace{1cm} (24)

The other requirement is maximizing the cutoff frequency. The artificial transmission lines have a significantly large peak in group delay response near the cutoff frequency, which degrades the signal quality. The cutoff frequency is limited by the gate line because $C_G$ is typically larger than $C_D$. This requirement is given by

$$\max \omega_G.$$  \hspace{1cm} (25)

The image impedance and cutoff frequency are determined by the dimensions of the inductors. The design of the inductors is a multi-objective and multi-variable optimization problem, and it is difficult to find the optimum dimensions by repeating the electromagnetic analysis. In this research, we determined the dimensions by numerical optimization based on the analytical solutions. We defined an evaluation function for the gate line as follows.

$$E_G(d, s, w) = \zeta_0 (Z_G - R_T)^2 + \zeta_1 (-\omega_G),$$  \hspace{1cm} (26)

where $\zeta_0$ and $\zeta_1$ are weighting factors that are positive constants. $Z_G$ and $\omega_G$ are given by (5) and (6), and $L$, $M$, $C_{OX}$, and $C_M$ are functions of $d$, $s$, and $w$ as described in
Fig. 6 Optimization result of the gate line. The solid lines and gradation represent $Z_G$ and $f_G$ at $w = 4 \mu m$, respectively.

Fig. 7 Optimization result of the drain line. The solid lines and gradation represent $Z_D$ and $(f_D - f_0)$ at $w = 6 \mu m$, respectively.

(19). $E_G$ is, therefore, a function of $d$, $s$, and $w$. This captures the trade-off between $L$, $M$, $C_{OX}$, and $C_M$. The optimum dimensions of the gate inductors can be fined by minimizing $E_G$. It satisfies the impedance matching condition and maximizes the cutoff frequency. Similarly, the evaluation function for the drain line was defined as follows.

$$E_D(d, s, w) = \eta_0 (Z_D - R_T)^2 + \eta_1 (\omega_G - \omega_D)^2,$$

where $\eta_0$ and $\eta_1$ are weighting factors. The first term and second term represent the requirements for impedance matching and phase matching. The dimensions of the drain line inductors are also determined by minimizing $E_D$.

The optimum point was searched using the gradient descent algorithm [29]. The ranges of the dimensions were constrained in $90 \mu m \leq d \leq 150 \mu m$, $2 \mu m \leq s \leq 10 \mu m$ and $4 \mu m \leq w \leq 8 \mu m$ so that the quality factor of the inductors would not become too small. The number of steps was 10000, and the dimensions were randomly changed every 100 steps to prevent convergence to the local optimum solutions. Fig. 6 shows an optimization result for the gate line, where the solid lines and gradation represent $Z_G$ and $f_G = \omega_G / 2\pi$, respectively. The optimization converged to a point where $d = 116 \mu m$, $s = 7 \mu m$, and $w = 4 \mu m$. $Z_G$ and $f_G$ converged 50.0 $\Omega$ and 24.5 GHz at the optimum point. The $f_G$ was maximized on the 50-$\Omega$ contour line, which satisfied impedance matching and maximized the cutoff frequency. Fig. 7 shows an optimization result for the drain line. The dimensions converged to $d = 120 \mu m$, $s = 4 \mu m$, and $w = 6 \mu m$. $Z_D$ and $f_D = \omega_D / 2\pi$ were 50.1 $\Omega$ and 24.9 GHz at the optimum point. The error between $f_G$ and $f_D$ was only 2%, which satisfied the phase matching condition. This error was mainly caused by rounding the dimensions to the integer. The optimization took only 4.14 seconds on an author’s laptop.

2.4 Prototyping

The amplifier was fabricated in Rohm’s one-poly five-metal 0.18-$\mu m$ standard CMOS process. The measured $f_{max}$ of the n-MOSFET was 65 GHz [30]. The top and bottom metal layers were used for the inductor winding and ground wiring, respectively. The bottom metal under the inductors was removed, and the silicon substrate was exposed. A higher resistance silicon block area was specified on the exposed silicon substrate to prevent eddy current losses. The number of turns of the edge inductors was set to one and their diameters were varied to be close to $L_g/2$ and $L_d/2$ while maintaining the width and spacing. The self-inductances of the gate and drain edge inductors were 270 pH and 250 pH at 20 GHz, respectively. $C_X$ was implemented with metal-insulator-metal capacitors, and $R_T$ was implemented with poly resistors. The poly resistors were placed on a thick insulator for shallow trench isolation, which provided a small parasitic capacitance of 4.8 fF. Fig. 8 shows a micrograph of the fabricated chip. The coupled inductors reduced the complexity of wiring. The core area was 0.35 mm $\times$ 0.76 mm = 0.27 mm$^2$, and the whole chip area including all the test pads was 1.1 mm $\times$ 0.66 mm = 0.73 mm$^2$. The bias voltages were $V_{PD} = 2.5$ V and $V_{GG} = 1.4$ V. The value of $V_T$ was adjusted so that the tail current of each stage was 7.5 mA in both simulation and measurement. The amplifier drew a current of 30 mA and consumed a power of 75 mW.

3. Measurement and Discussion

The four-port S-parameters were measured using Keysight’s PNA-X N5247A vector network analyzer. The measurement frequency range was 0.1 GHz to 30 GHz with a frequency
The bandwidth was 18.5 GHz. The PP signal was a differential pseudo signal connected to the input and output of the amplifier. The input generator and UXR1104A real transmission line indicates that the impedance is less than the Nyquist frequency. The gain variation was ±1.1 dB below the Nyquist frequency of 15 GHz. The gain variation was mainly due to the propagation losses of the artificial transmission lines. A steep roll-off was observed at 25 GHz. This was consistent with the theoretical cutoff frequency of the artificial transmission lines. The input and output reflections of $S_{dd11}$ and $S_{dd22}$ were less than −15 dB and −13 dB, respectively. A small impedance mismatch was achieved by the optimization. Fig. 9(b) shows common-mode gain and isolation. The common-mode gain was nearly −10 dB that is smaller than the differential gain thanks to the fully-differential topology. $S_{dd12}$ was smaller than −20 dB because the amplifier cell was unilateralized by $C_X$. Fig. 9(c) shows group delay responses. To eliminate the variation due to the measurement noise, a moving average of 10 nearby points is shown as a solid line. We demonstrated a small group delay variation of ±2.7 ps below the Nyquist frequency. This flat group delay response indicates that the impedance and phase matching of the artificial transmission lines is properly ensured. Then, we performed a time domain measurement. Keysight’s M8195A arbitrary generator and UXR1104A real-time oscilloscope were connected to the input and output of the amplifier. The input signal was a differential pseudo-random signal with a length of $2^9$−1. The differential outputs were connected to Ch.1 and Ch.3 of the oscilloscope, and the Ch.3 signal was observed. The linear distortion due to the RF cables and probes was removed by the channel de-embedding before the measurement. Fig. 10(a) shows an eye diagram for a 30-Gbps NRZ signal. The inset shows the through measurement with a calibration standard. The output swing was 100 mVpp and 160 mVpp for the through and amplifier. The estimated timing margin at a bit error rate of $10^{-12}$ for the throughput and amplifier was 0.88 UI and 0.75 UI, and the root-mean-square jitter was 0.34 ps and 0.42 ps, respectively. The optimization reduced the group delay variation, resulting in small jitter degradation. Fig. 10(b) shows an eye diagram for a 30-Gbps PAM4 signal. The output swing was 130 mVpp and 220 mVpp. The eye width was 0.5 UI and 0.4 UI for the through and amplifier, achieving 30-Gbps PAM4 transmission.

Table 2 lists a performance summary and comparison with the distributed amplifiers previously reported. Rito et al. [3] reported a distributed modulator driver using SiGe:C BiCMOS technology with $f_{\text{max}} = 500$ GHz. Taking advantage of the excellent device performance, 90-Gbps PAM4 transmission was demonstrated. However, the bandwidth normalized by $f_{\text{max}}$ is 0.18, which is not a particularly wide bandwidth from the point of view of the circuit. Wu et al. [7] introduced a frequency interleaving technique and reported excellent bandwidth. However, this technique causes rapid group delay variations at the sub-band junctions. Lee et al. [8] proposed a distributed amplifier using deeply doped channel technology, which has very low power consumption.

![Fig. 9](image_url) Measured S-parameters. The solid lines and dashed lines represent the measurement and simulation. Unless otherwise indicated, the simulation results are slow corner. (a) Differential gain and reflection. (b) Common-mode gain and in-out isolation. (c) Group delay response.

![Fig. 10](image_url) Measured eye diagram. (a) 30-Gbps NRZ. (b) 30-Gbps PAM4. Inset: eye diagram with a through structure.
The amplifiers reported in [11]–[14] used a pseudo-differential topology. Pseudo-differential does not provide the advantages of differential circuits. The distributed amplifier reported in this study has a small core area of 0.27 mm² and a fully-differential topology. Compared to the most similar circuit [15], the coupled inductors allow a 57% reduction in silicon area. Although [4]–[8] used the single-ended topology, they have a larger core area than this work. The developed amplifier has a small power consumption of 75 mW at the cost of the gain. The gain can be increased by cascading amplifiers [7]. A future challenge is to develop a high-gain distributed amplifier that takes advantage of the fully-differential topology that allows true DC coupled cascading. The bandwidth normalized by \( f_{\text{max}} \) was 0.28, achieving the target specification. By applying the proposed circuit to advanced processes with \( f_{\text{max}} > 200 \) GHz, a bandwidth of 56 GHz or higher is expected and enables 100 Gbaud transmissions. In addition, loss compensation can be used to further extend the bandwidth reached to the theoretical bandwidth of 25 GHz. Besides, PAM4 transmission was demonstrated only in this study and [3]. These studies reported small group delay variations of less than ±5 ps. Therefore, it is important to strictly satisfy the impedance and phase matching conditions for multi-level transmission.

4. Conclusion

We presented a compact fully-differential distributed amplifier that reduces a silicon area by more than 50%. To achieve a flat frequency response, we analyzed the differential artificial transmission lines using a simple loss-less model. The lines were optimized by a gradient descent algorithm based on the derived odd-mode image parameters. The phase and impedance matching were strictly satisfied by the optimization, and we demonstrated a small group delay variation of ±2.7 ps and 30-Gbps NRZ and PAM4 transmission. Therefore, these results will accelerate the high-density integration of differential distributed circuits.

Acknowledgments

This work was supported through the activities of VDEC, The University of Tokyo, in collaboration with Rohm Corporation, Toppan Printing Corporation, Cadence Design Systems, Mentor Graphics, and Keysight Technologies.

References


