

Article

# Analytical Analysis of Power-Constrained Repeaters' Insertion in Large-Scale CMOS Chips

Luigi Gaioni

Department of Engineering and Applied Sciences, University of Bergamo, 24044 Dalmine, BG, Italy;  
luigi.gaioni@unibg.it

**Abstract:** As the die area of CMOS integrated circuits continues to increase, interconnects will become dominant in determining the performance of the circuits from the standpoint of speed and power consumption. Uniform repeater insertion is an effective method used to reduce the propagation delay of a signal in long resistive-capacitive lines. However, non-optimal repeaters' insertion yields non-optimal circuit performance. In this work, we provide a mathematical treatment for optimal repeater insertion with power consumption constraints. In particular, a closed-form expression for the optimum number and size of repeaters is given for a two-stage buffer used as a repeater. The validation of the analytical solution is assessed by means of circuit simulations, by comparing the theoretical optimal number and size of the repeaters to be placed in the long resistive-capacitive line with the simulated values.

**Keywords:** constrained optimization; CMOS buffers; large-scale chip

## 1. Introduction

Over the recent years, CMOS scaling has made it possible to comply with the demands of the computer and information technology industry, leading to very large-scale integrated (VLSI) circuits with increased functionalities and performance. While the device feature size shrinks, the die size of integrated circuits continues to increase. In this scenario, interconnects have become increasingly significant in determining both the speed and the power consumption of the circuits. In large-scale CMOS chips, one of the primary challenges is maintaining signal integrity. As chip technology scales down to smaller nodes, key points such as signal delay and noise become more prevalent. CMOS buffer insertion can mitigate these issues, ensuring that signals are transmitted reliably across the chip, reducing the probability of data corruption or transmission errors. This is particularly crucial in AI (artificial intelligence) applications, where massive amounts of data need to be processed quickly and accurately. Power consumption is also a critical concern, particularly in battery-operated IoT (Internet of Things) devices. Optimum buffer insertion can help minimize power consumption, extending battery life and contribute to the overall energy efficiency of a device. In the framework of AI applications, where high computational power is often required, improving power efficiency helps in managing the heat generated and prevents thermal issues that could otherwise impair performance. By reducing propagation delay, optimal buffer insertion enable faster data transmission and processing speeds. This is crucial for AI applications, which rely on real-time decision making and data processing capabilities, and where faster processing speeds can lead to more responsive and efficient AI systems, which can be crucial in applications ranging from autonomous vehicles to real-time analytics.

In very-large-scale integration design, clock distribution plays a crucial role in the synchronization of the operations of the different components making up the integrated circuit (IC). As the technology advances and the number of components interconnected within an IC increases, the challenges associated with clock distribution become more and more striking. As mentioned, two major concerns in VLSI clocks and, in general,



**Citation:** Gaioni, L. Analytical Analysis of Power-Constrained Repeaters' Insertion in Large-Scale CMOS Chips. *Electronics* **2024**, *13*, 4368. <https://doi.org/10.3390/electronics13224368>

Academic Editor: Hyungjin Kim

Received: 17 September 2024

Revised: 31 October 2024

Accepted: 6 November 2024

Published: 7 November 2024



**Copyright:** © 2024 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

signal distribution are delay, which ultimately affects the speed at which the circuits can be operated, as well as power consumption. The delay, by definition, actually represents the time needed for the clock signal to propagate from the source to a specific component within the IC. As the chip size and the functionalities integrated in the IC increase, and, consequently, the number of clocked elements grows, the delay associated with the clock signal propagation becomes a critical factor. Excessive delay can indeed determine timing violations, when the arrival time of the clock signal in different elements integrated in the chip varies beyond acceptable limits. This may result in functional errors, reduced performance, or even chip failure. Another key point is the power consumption associated with the signal distribution in VLSI circuits, which arises from different sources, including the dynamic component associated with the driving of signals through long interconnect lines, the power associated with buffers, and with clock distribution networks. As a rule of thumb, as chip complexity increases, so does the power associated with the clock distribution. Such a power consumption significantly contributes to overall chip power dissipation and needs to be managed properly in order to avoid thermal issues, reduced battery life—in the case of portable devices—and tough cooling requirements.

It has been shown [1] that a linear increase in interconnect length results in a quadratic increase in propagation delay due to a linear increase in both parasitic capacitance and resistance. Many algorithms have been proposed to set the optimum wire size, minimizing the cost function, such as the delay [2]. As mentioned, the increase in delay and power is a major concern in VLSI clock distribution: In advanced VLSI circuits, the clock signal must be distributed all over the chip with minimum possible skewness. While keeping the skewness at a minimum, the clock distribution network may waste a significant amount of power, up to 50% of the total consumed power in high performance applications [3]. It has to be noticed that delays and skewness in ICs are also strictly related to process variations, which are critical in designing nanoscale CMOS circuits. Process variations inside a specific IC can indeed result in different delays for nominal identical logic gates. As a result, designers typically introduce large timing margins in their designs to avoid timing violations. The so-called clock mesh architecture [4] has emerged as an effective solution, providing multiple paths for clock signals, ultimately making the design less sensitive to process variations. The tolerance to such variations actually comes at the cost of an increased number of interconnects in the IC and, therefore, increased power consumption. To overcome this issue, a number of solutions has been proposed, as the one described in [5], where the authors discussed a hybrid clock network architecture that combines both the rather standard clock tree strategy together with clock meshing. Other approaches leverage the usage of resonant clocks, as described in [6,7]. In general, different methods have been proposed in the literature to reduce power consumption. One of the most effective ways to reduce power consumption is clock gating [8–10], where the clock signal for components or regions in the chip that are not used is turned off. In particular, clock gating is typically implemented by means of specific logic blocks controlled by a so-called gating signal. When such a signal is active, indicating that a specific component or region of the IC is in an *active* state, the clock is allowed to pass through the gating logic, thus enabling the operation for the enabled part of the circuit. On the other hand, when the gating signal is inactive, indicating that the corresponding part of the circuit is in an *idle* state, the clock signal is blocked, and the activities in that portion of the circuit are effectively stopped, resulting in overall power saving. The described technique can lead to a significant reduction in the dynamic power consumption of the IC by preventing the switching activity in circuit blocks, which are not active, and due to a reduction of critical path delay, possibly resulting in an improvement in the overall timing performance of the IC. On the other hand, clock gating typically calls for careful verification, timing analysis and optimization, which are required to avoid timing violations associated with the additional logic implementing the clock gating itself. Other methods deal with the reduction of the dynamic component [11–13], typically obtained by reducing the voltage swing of the signal or implementing charge recycling techniques. Several designs have also been devised to reduce interconnect delay [14–20]: among them, uniform repeater

insertion methods aim at reducing the time for a signal to propagate through a long line by dividing the line itself into equal sections driven by equal size repeaters. However, the insertion of a sub-optimal number of repeaters with an inadequate size yields sub-optimal performance. Moreover, power consumption due to repeater insertion has become higher and higher in modern VLSI chips. Bakoglu provides a solution for the optimum size and number of repeaters to be placed in a long line featuring a specific resistance-capacitance (RC) interconnect impedance [1]. In that work, no constraints are taken into account in the optimization process. In [16], Dhar and Franklin introduce a mathematical treatment for optimal repeater insertion, with and without area constraints. However, no closed-form solution is provided in their work. In [17], the transistor sizing is based on the minimization of the energy–delay product.

Nowadays, buffer insertion leverages modern approaches such as machine learning-based optimization [18,19]. Machine learning-based optimization approaches leverage data-driven techniques to learn patterns and relationships from large datasets of circuit designs and their corresponding performance metrics. These models can then be used to predict the optimal buffer placement and sizing for new designs. While such methods can handle complex circuit structures and multiple optimization objectives, their performance heavily relies on the quality (and quantity) of the training data. In addition, training and deploying machine learning models can be computationally intensive, especially for large-scale problems. On the other hand, analytical methods can often provide exact or, at least, highly accurate solutions, and they can be computationally efficient for a large number of problems, particularly when the circuit structure is relatively simple. By employing closed-form equations, it is possible to eliminate the guesswork and trial-and-error approaches often used in manual designs. Efficiency is another major benefit: a closed-form solution can significantly reduce computational overhead compared to iterative simulation methods. In many cases, a hybrid approach, combining analytical and machine learning techniques, can offer the best solution. For example, analytical methods can be used to generate initial designs, while machine learning can be used to fine-tune the optimization process or handle complex design scenarios.

This work is focused on an analytical treatment for determining the size and the number of repeaters to be uniformly inserted in a long interconnection line, which leads to the minimization of the propagation delay while meeting a given power budget constraint. This paper is organized as follows. In Section 2, a propagation delay model based on discrete resistance and capacitance elements is presented. The constrained delay minimization is outlined in Section 3, where a closed-form expression for the optimum number and size of repeaters is given for a two-stage buffer used as the repeater. A comparison between the analytic model and circuit simulations is presented in Section 4.

## 2. Propagation Delay Model

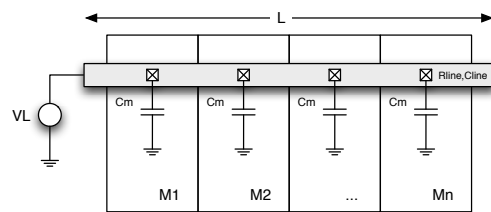
In order to define the propagation delay model, let us first consider the general modular system shown in Figure 1, which is supposed to operate at low frequencies (i.e., not exceeding the MHz regime). Such a system consists of a line distributing the digital signal  $V_L$  to the  $n$  modules, each featuring an input capacitance  $C_m$  loading the line itself. The line is characterized by an overall resistance  $R_{line}$  and an overall capacitance  $C_{line}$ , whose values depend linearly on the line length  $L$  [1]. The delay model is shown in Figure 2 for a uniform line divided into  $k$  sections, each driven by a minimum-size buffer. Each line section,  $L/k$  long, is modeled by a distributed RC circuit, where both  $R_{line}$  and  $C_{line}$  are divided by a factor of  $k$ . The  $C_L$  capacitance, given by the total modules' input capacitance,  $nC_m$ , is divided by the same factor. The input of the buffer is modeled by means of a fixed capacitance  $C_B$ , whereas  $R_B$  models its driving capability. The values of capacitance  $C_B$  and resistance  $R_B$  depend on the buffer input and output transistors' size, respectively.

In this work,  $R_B$  is referred to as the on-resistance of the output MOS transistors biased with a gate-to-source voltage equal to  $V_{DD}$  (i.e., the maximum allowed supply voltage), and with a drain-to-source voltage equal to  $V_{DD}/2$ . A geometrical interpretation of  $R_B$  is given in Figure 3, where the value of such a resistance is visualized by the angle  $\alpha$  with

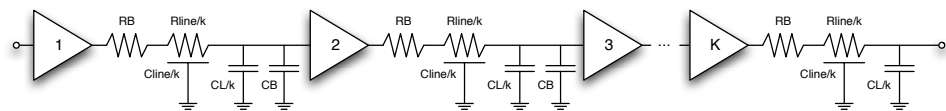
respect to the y-axis of the  $I_D - V_{DS}$  characteristic of the output transistor. The propagation delay of the line with  $k$  minimum-size buffers used as repeaters can be expressed as

$$T_{D,msize} = k \left[ R_B \left( \frac{C_{line} + C_L}{k} + C_B \right) + \frac{R_{line}}{k} \left( \frac{C_{line}}{2} + C_L + C_B \right) + D_B \right], \quad (1)$$

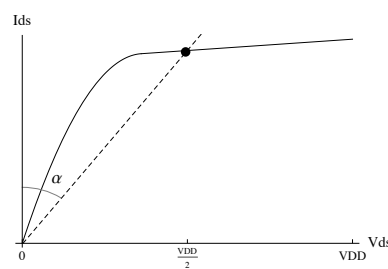
where the parameter  $D_B$  takes into account the intrinsic delay of the minimum-size buffer, mainly depending on the number of the stages making up the buffer. It is worth noting that the concept of a “minimum-size buffer” used throughout this work is somehow arbitrary: we define it as the buffer able to properly drive a segment  $L/n$  long of the line. For a two-stage buffer in the 65 nm commercial technology used in circuit simulations reported in Section 4,  $D_B$  is close to 60 ps. Its value can be obtained by simulating an unloaded buffer or via a calculation as reported in [21]. The value of such a parameter does not change with the buffer dimension as the increase in input capacitance of the  $i$ th stage in the buffer is compensated by the increase in the driving capability of the  $(i - 1)$ th stage.



**Figure 1.** A general modular system with  $n$  elements. The digital signal  $V_L$  is distributed to the modules by means of a common line with length  $L$ .



**Figure 2.** Delay model for a line divided into  $k$  sections each driven by a minimum-size buffer.



**Figure 3.** The  $R_B$  resistance of the repeater equals the ratio  $V_{ds}/I_{ds}$  of the output transistors and is represented by the angle  $\alpha$  between the dotted line and the y-axis.

Delay can be reduced by optimizing the buffers’ size (and thus increasing their driving capability). When a buffer’s size is increased by a factor of  $h$ , its input capacitance and output resistance become  $hC_B$  and  $R_B/h$ , respectively, and the delay expression takes the form of

$$T_D = k \left[ \frac{R_B}{h} \left( \frac{C_{line} + C_L}{k} + hC_B \right) + \frac{R_{line}}{k} \left( \frac{C_{line}}{2} + C_L + hC_B \right) + D_B \right]. \quad (2)$$

Note that  $D_B$  has not scaled with  $h$ , as previously mentioned. Optimal values for  $h$  and  $k$ , that is, the values which guarantee the minimum propagation delay, can easily be obtained by setting  $\frac{\partial T_D}{\partial h} = 0$  and  $\frac{\partial T_D}{\partial k} = 0$ :

$$h_{opt} = \sqrt{\frac{R_B(C_{line} + C_L)}{R_{line}C_B}}, \quad (3)$$

$$k_{opt} = \sqrt{\frac{R_{line}(C_{line} + 2C_L)}{2(R_B C_B + D_B)}}. \quad (4)$$

The resulting delay expression becomes

$$T_{D,opt} = \frac{2\sqrt{R_{line}(C_{line} + C_L)R_B C_B} + \sqrt{2R_{line}(C_{line} + 2C_L)(R_B C_B + D_B)}}{2}, \quad (5)$$

which is smaller than the optimum delay achievable with minimum-size buffers under the reasonable assumption that  $R_{line}C_B < 2R_B C_{line}$ . To ensure that the previously computed optimum values are indeed global optima,  $T_D$  must be a convex function, which can be demonstrated by showing that the Hessian matrix for  $T_D$  is positive semi-definite. The Hessian matrix,  $H$ , for  $T_D$  is

$$H(T_D) = \begin{bmatrix} \frac{\partial^2 T_D}{\partial h^2} & \frac{\partial^2 T_D}{\partial h \partial k} \\ \frac{\partial^2 T_D}{\partial k \partial h} & \frac{\partial^2 T_D}{\partial k^2} \end{bmatrix}, \quad (6)$$

where it can be easily shown that

$$\begin{aligned} \frac{\partial^2 T_D}{\partial h^2} &= \frac{2R_B(C_L + C_{line})}{h^3}, \\ \frac{\partial^2 T_D}{\partial k^2} &= \frac{R_{line}(2C_L + C_{line})}{k^3}. \end{aligned} \quad (7)$$

By considering only positive values for the variables  $h$  and  $k$  (i.e., only physically meaningful values) and that all the parameters in (7) are positive real constants, it is possible to conclude that the second derivatives are positive. Hence,  $H(T_D)$  is positive semi-definite and, in turn,  $T_D$  is convex.

### 3. Constrained Delay Optimization

In order to minimize the delay while meeting a given power budget constraint, a simple power model for the circuit shown in Figure 2 is now introduced. Henceforth, we will consider an  $N$ -stage tapered buffer with a tapering factor equal to  $F$  as the minimum-size repeater. The size of the buffer can be scaled by the same factor  $h > 1$  introduced in Equation (2). The total power consumption accounted for in the model can be split into two components:

- Dynamic power consumption,  $P_{dl}$ , related to the line capacitance  $C_{line}$  and the total module capacitance  $C_L$ ; it can be expressed as

$$P_{dl} = fV_{DD}^2(C_{line} + C_L), \quad (8)$$

where  $f$  is the frequency of the transmitted signal.

- Dynamic power consumption,  $P_{db}$ , related to the input capacitance of the  $N$  stages making up the tapered buffer; it can be expressed as

$$P_{db} = kfV_{DD}^2 \left( hC_B \sum_{i=0}^{N-1} F^i \right). \quad (9)$$

The total power consumption is actually also affected by the contribution of the short circuit,  $P_{sc}$ , relevant to the  $k$ -tapered buffers placed along the line. This contribution is due to the direct current flowing from  $V_{DD}$  to the ground for a short time during switching. An expression of  $P_{sc}$ , based on the Sakurai alpha power model, is given in [22]. This expression underestimates the short-circuit power consumption for interconnects featuring a large RC. In this situation, the first stage of a multi-stage buffer may dissipate a significant amount of short-circuit power due to the degraded waveform at its input originating from the large RC loading the former driver. A simple expression of  $P_{sc}$  for repeaters driving large RC loads is given in [23]. However, it is worth noting that the  $P_{sc}$  contribution to the total power consumption is generally much smaller than the contribution coming from  $P_{dl} + P_{db}$ . The total power  $P_T$  can thus be given by the sum of the following two components:

$$P_T = P_{dl} + P_{db} \quad (10)$$

The problem of minimizing the delay while meeting the power consumption constraint can be formulated as follows:

$$\begin{aligned} & \text{minimize} && T_D = f_D(\mathbf{x}) \\ & \text{subject to} && P_T = f_P(\mathbf{x}) < P_{MAX}, \end{aligned} \quad (11)$$

where the vector  $\mathbf{x} = (h, k)$  is the optimization variable of the problem, the function  $f_D : \mathbb{R}^2 \rightarrow \mathbb{R}$  is the objective function, the function  $f_P : \mathbb{R}^2 \rightarrow \mathbb{R}$  is the inequality constraint function, and  $P_{MAX}$ , the maximum allowed power consumption, is the bound for the constraint. Both the delay  $T_D$  and the total power consumption  $P_T$  are functions of the size  $h$  and number  $k$  of the repeaters placed in the line. In particular,  $f_D(h, k)$  and  $f_P(h, k)$  are expressed by means of Equation (2) and Equation (10), respectively. A vector  $\mathbf{x}^*$  is called optimal if it has the smallest objective value among all vectors that satisfy the constraint. It is possible to solve the minimization problem considering two cases. In the first one, the unconstrained minimum for the delay occurs within the feasible region, as shown in Figure 4a. In this situation,  $\mathbf{x}^* = (h_{opt}, k_{opt})$ , where  $h_{opt}$  is given by Equation (3) and  $k_{opt}$  is given by Equation (4). In the second case, the unconstrained local minimum lies outside the feasible region, as shown in Figure 4b. In this situation, the inequality constraint in Problem (11) can be contracted into an equality constraint and the minimization problem, which, written in the standard form, becomes

$$\begin{aligned} & \text{minimize} && f_D(\mathbf{x}) \\ & \text{subject to} && f_P(\mathbf{x}) - P_{MAX} = 0. \end{aligned} \quad (12)$$

The constrained minimum occurs at  $\mathbf{x}^*$  when  $\nabla_{\mathbf{x}} f_D(\mathbf{x}^*)$  and  $\nabla_{\mathbf{x}} f_P(\mathbf{x}^*)$  are parallel, as follows:

$$\nabla_{\mathbf{x}} f_D(\mathbf{x}^*) = \lambda \nabla_{\mathbf{x}} f_P(\mathbf{x}^*), \quad (13)$$

for some  $\lambda$ . Hence, the extrema of the function  $f_D(\mathbf{x})$  which satisfy the constraint are given by the solution of the following system of equations:

$$\begin{aligned} \nabla_{\mathbf{x}} f_D(\mathbf{x}) &= \lambda \nabla_{\mathbf{x}} f_P(\mathbf{x}) \\ f_P(\mathbf{x}) - P_{MAX} &= 0. \end{aligned} \quad (14)$$

It is convenient to introduce the Lagrangian  $L(\mathbf{x}, \lambda)$  associated with the constrained problem, defined as

$$L(\mathbf{x}, \lambda) = f_D(\mathbf{x}) + \lambda(f_P(\mathbf{x}) - P_{MAX}), \quad (15)$$

where  $\lambda$  is known as the Lagrange multiplier. Setting  $\nabla_{\mathbf{x}, \lambda} L = 0$  yields the same system (14) of nonlinear equations that, in general, can be solved by means of numerical techniques. The solution  $(\mathbf{x}^*, \lambda^*) = (h^*, k^*, \lambda^*)$ , with  $h^* > 0$  and  $k^* > 0$  (i.e., the physically meaningful values corresponding to the extrema of the function  $f_D(\mathbf{x})$ ), gives us the minimum delay

$T_D$  meeting the power budget requirement. A closed-form expression for the solution  $(h^*, k^*)$  of the constrained problem (12) is given in Equations (16) and (17), in the case of a two-stage tapered buffer with a tapering factor equal to  $F$  used as the repeater.

$$h^* = \sqrt{2} \sqrt{\frac{A_1(D_B A_1 + R_B C_B A_3)}{(F + 1) f V_{DD}^2 R_{line} C_B^2 A_2'}} \tag{16}$$

$$k^* = \frac{1}{\sqrt{2}} \sqrt{\frac{R_{line} A_1 A_2}{(F + 1) f V_{DD}^2 (D_B A_1 + R_B C_B A_3)'}} \tag{17}$$

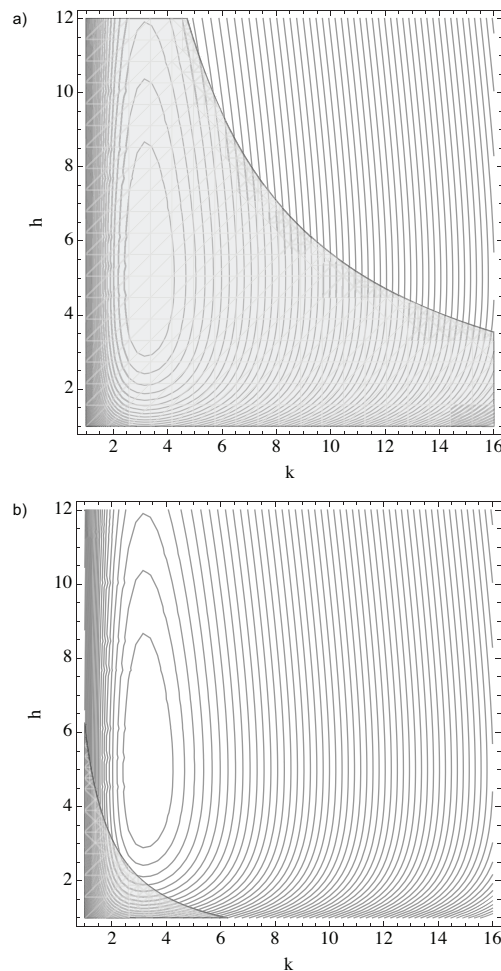
where

$$A_1 = P_{MAX} - (C_{line} + C_L) f V_{DD}^2, \tag{18}$$

$$A_2 = 2P_{MAX} + ((F - 1) C_{line} + 2F C_L) f V_{DD}^2, \tag{19}$$

$$A_3 = P_{MAX} - F(C_{line} + C_L) f V_{DD}^2 \tag{20}$$

have the dimensions of power. It is worth recalling here that the parameters shown in the previous equations can be derived from one sub-module of the modular system shown in Figure 1. A practical evaluation of these parameters may require the usage of CAD parasitic extraction tools, and it is not expected to be critical, as the size of the sub-module should be limited with respect to the overall system.



**Figure 4.** Contour lines of the propagation delay as a function of the number  $k$  and size  $h$  of the repeaters placed in the distribution line. The shaded area corresponds to the feasible region, where the power constraint is met. In part (a), the unconstrained minimum occurs within the feasible region, whereas in part (b), the unconstrained minimum delay lies outside the feasible area.

#### 4. Model Validation

For the purpose of validating the delay model and the constrained optimization process illustrated in Sections 2 and 3, a comparison between circuit simulations and analytical computation of the propagation delay and of the optimum  $h$  and  $k$  parameters will be discussed in this section.

In particular, post-layout simulations have been carried out for a test structure inspired to the so-called Macro Pixel ASIC (MPA) [24], which was designed for the Phase II upgrade of the Compact Muon Solenoid (CMS) Outer Tracker detector at the High Luminosity Large Hadron Collider (HL-LHC), at CERN. The outer tracker of the CMS at the HL-LHC will include a number of modules, referred to as Pixel-Strip (PS) modules, integrating both pixelated and strip sensors. The latter will be read out by means of a dedicated chip called Strip Sensor ASIC (SSA), which will perform the analog to digital conversion of the data and transmit them to the Macro Pixel ASIC every 25 ns. On the other hand, the MPA will be exploited for the readout of the pixel layer of the PS module. The Macro Pixel ASIC will be bump bonded to the pixel detector, as in standard hybrid sensors adopted in high-energy physics experiments. The MPA consists of a matrix of  $16 \times 128$  pixels featuring a size of  $1440 \mu\text{m} \times 100 \mu\text{m}$ , and it is designed using low-power 65 nm CMOS technology. Power consumption minimization is one of the key challenges in the design of the Macro Pixel ASIC. Indeed, applications at the outer tracker of the CMS experiments require that the power density for the whole chip, including the analog front-end, the digital logic, and the I/O blocks, be smaller than  $90 \mu\text{W}/\text{cm}^2$ . As a result, the Macro Pixel ASIC leverages a Multi-Supply Voltage (MSV) architecture to significantly reduce digital power consumption without compromising the performance of the analog front-end, particularly from the stand point of noise.

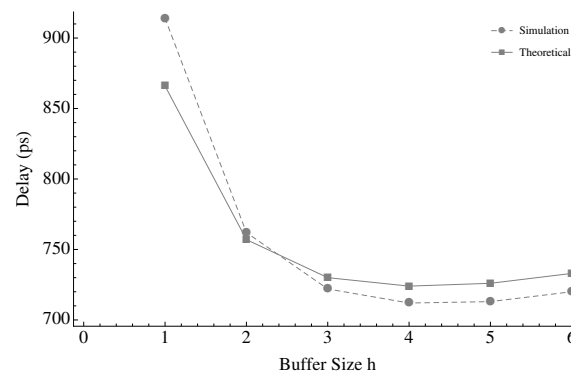
Memory and clock gating are also widely used in the design of the Macro Pixel ASIC, whose description is provided in [25]. Inspired to the MPA design, a test structure was simulated, leveraging a commercial 65 nm process with nine metal layers (together with an additional redistribution layer), where a 40 MHz clock signal is distributed to the 16 rows of a matrix through a vertical, 2.3 cm long line, laid out with the nine metal layer with a width of  $0.5 \mu\text{m}$ , featuring a parasitic line capacitance  $C_{line}$  equal to 6 pF and a parasitic line resistance  $R_{line}$  of  $220 \Omega$ . A shield connected to the ground potential was designed in the nine metal layers underneath the vertical line. Each row of the matrix is driven by a dedicated buffer featuring an input capacitance of 25 fF. The simulated structure can thus be modeled with the aid of the system shown in Figure 1, where  $n$ , the number of the modules of the system, is equal to 16 (the number of the row of the MPA) and  $C_m$  is equal to 25 fF, namely the input capacitance of the buffer driving the row of the matrix. The main parameters, as extracted from the layout and used in the simulations, are gathered in Table 1. The test bed was simulated by means of the Spectre simulator with BSIM4 (V4.5) models. It is worth noticing that the simulated test structure consists of a fully custom design (no standard cell libraries were used) meant to check the theoretical results obtained in the previous sections. The repeaters used in the vertical clock line of the simulated structure are two-stage CMOS buffers supplied with  $V_{DD} = 0.8 \text{ V}$ . As mentioned, the use of a reduced supplied voltage (with respect to the nominal core voltage of 1.2 V for the 65 nm CMOS technology) leads to a quadratic reduction in power dissipation, which is a major concern in the MPA chip.

**Table 1.** Parameters used in the simulations.

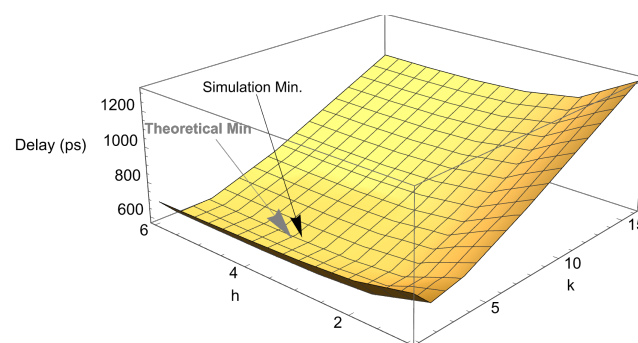
$D_B$	25 ps
$R_B$	$35 \Omega$
$R_{line}$	$220 \Omega$
$C_B$	67 fF
$C_{line}$	6 pF
$C_m$	25 fF



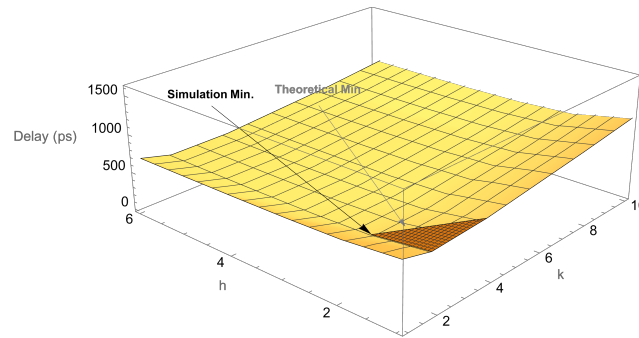
Figure 5 shows a comparison between simulation results and theoretical values, computed by means of Equation (2), of the propagation delay when four repeaters are uniformly placed along the line. In the figure, the delay is plotted as a function of the size  $h$  of the repeaters. Simulation and theoretical results are in fairly good agreement, with the model replicating the optimum delay occurring for  $h \approx 4$ . In order to set the optimum number and size of repeaters that minimize the propagation delay without power budget constraints, a number of simulations were carried out by varying the  $k$  parameters, ranging from 1 to 16, and  $h$  parameters, ranging from 1 to 6. The optimum values obtained by means of such simulations were compared with the values  $h_{opt}$  and  $k_{opt}$  foreseen by Equations (3) and (4). Figure 6 shows the propagation delay as a function of  $h$  and  $k$  for the simulated structure. A black arrow points to the optimum coordinate  $(h_{sim}, k_{sim})$  obtained by means of circuit simulations, whereas a gray arrow points to the pair  $(h_{opt}, k_{opt})$  obtained by calculations. As far as the constrained minimization process is concerned, it is possible to outline the feasible region and to search for the minimum delay within such a region by means of circuit simulations. In Figure 7, the shaded area represents the feasible region meeting the power budget constraint  $P_{MAX} \leq 230 \mu\text{W}$ . Simulations for the constrained problem led to the optimum, which is pointed out by the black arrow, while the theoretical optimum values of  $h$  and  $k$  satisfying the power consumption constraint were obtained by means of Equations (16) and (17). Once again, a gray arrow points to the theoretical constrained minimum. Table 2 shows simulations and theoretical results relevant to the analysis carried out on the test circuits: it has to be pointed out that  $(h_{opt}, k_{opt}) \in \mathbb{N}^2$  for the simulation results, whereas theoretical analysis leads to a couple of values belonging to real numbers.



**Figure 5.** Simulated and theoretical values of the propagation delay with four repeaters uniformly placed along a distribution line featuring  $C_{line} = 6 \text{ pF}$  and  $R_{line} = 220 \Omega$ .



**Figure 6.** Propagation delay as a function of  $h$  and  $k$  for a distribution line featuring  $C_{line} = 6 \text{ pF}$  and  $R_{line} = 220 \Omega$ . The black arrow points to the optimum coordinate obtained by means of circuit simulations, whereas the gray arrow points to the coordinate obtained by calculations.



**Figure 7.** Propagation delay as a function of  $h$  and  $k$  for a distribution line featuring  $C_{line} = 6$  pF and  $R_{line} = 220 \Omega$ . The black arrow points to the optimum coordinate obtained by means of circuit simulations, whereas the gray arrow points to the coordinate obtained by calculations. The shaded area represents the feasible region meeting the power budget constraint of  $230 \mu\text{W}$  associated with the clock distribution circuit.

**Table 2.** Simulation and theoretical results.

	Unconstrained		Constrained	
	$h_{opt}$	$k_{opt}$	$h_{opt}$	$k_{opt}$
Simulation ( $h, k$ ) $\in \mathbb{N}^2$	4	4	2	2
Theoretical ( $h, k$ ) $\in \mathbb{R}^2$	4.09	3.65	1.8	2.8

Both in the case of unconstrained and constrained problems, simulation results are shown to be in good agreement with theoretical data. While the outcomes are generally consistent, slight discrepancies were found. Such differences can be attributed to the exclusion of crowbar current power consumption in the model described in Section 3. Indeed, the current model focuses only on the dynamic power consumption required to drive a capacitive line and does not account for the crowbar current associated with digital buffers during switching. Despite these minor differences, the overall results remain remarkably similar. To maintain simplicity and usability, we chose not to include the crowbar current component in this initial model.

## 5. Conclusions

Repeater insertion is a well-known solution for driving long interconnects in very-large-scale integration circuits. On the one hand, such a method allows for overcoming the quadratic dependence of the propagation delay on the interconnect length; on the other hand, power consumption related to buffer insertion can have a significant impact in modern CMOS VLSI design. In this work, an analytical method was presented for determining the number and size of the repeaters to be uniformly inserted into an RC line in order to minimize the delay while meeting a given power budget. Simple yet accurate models were used in the analysis, both for propagation delay and power consumption. A closed-form solution for optimum insertion was given for a two-stage buffer used as a repeater. A comparison between the power-constrained optimum number and size of repeaters obtained by means of circuit simulations, from analytical estimates, was carried out for a specific test circuit inspired by the so-called Macro Pixel ASIC (MPA), a large area readout chip which will be employed in the Compact Muon Solenoid tracker at the High Luminosity Large Hadron Collider. The comparison between circuit simulations and analytical calculations confirms the validity of the proposed models and methodology. In particular, a test case was investigated where a set of buffers drove a line featuring a distributed resistance of  $220 \Omega$ , together with a load capacitance of 6 pF. Simulations for the unconstrained problem led to an optimum number of buffers equal to 4, with a size

equal to 4, whereas the theoretical analysis led to 4.09 and 3.65, respectively. Regarding constrained minimization, simulation results led to an optima for two buffers that were of size 2. Alternatively, using a theoretical computation, the minimum was found for the values of 1.8 and 2.8. The results this analysis can be reasonably extended for a generic modular system in which a signal is transmitted through a long resistive-capacitive line. It has to be stated that the results drawn in this work are relevant to systems operating in the low-frequency domain. Extending the results for the high-frequency domain would imply incorporating the inductive behavior of the signal lines and treating them from a transmission line perspective, which is out of the scope of this work.

**Funding:** This research received no external funding.

**Data Availability Statement:** Data are contained within the article.

**Acknowledgments:** The author wishes to thank Paolo Lazzaroni (INFN Pavia) for their useful comments and support in writing this work.

**Conflicts of Interest:** The author declares no conflicts of interest.

## References

1. Bakoglu, H.B.; Meindl, J.D. Optimal Interconnection Circuits for VLSI. *IEEE Trans. Electron. Devices* **1985**, *32*, 903–909. [[CrossRef](#)]
2. Cong, J.J.; Leung, K.-S. Optimal wiresizing under Elmore delay model. *IEEE Trans. Comput. Aided Des. Integr. Circuits Syst.* **1995**, *14*, 321–336. [[CrossRef](#)]
3. Zhang, H.; George, V.; Rabaey, J.M. Low-swing on-chip signaling techniques: Effectiveness and robustness. *IEEE Trans. VLSI Syst.* **2000**, *8*, 264–272. [[CrossRef](#)]
4. Jung, J.; Lee, D.; Shin, Y. Design and optimization of multiple-mesh clock network. In *VLSI-SoC: Internet of Things Foundations. VLSI-SoC 2014. IFIP Advances in Information and Communication Technology*; Springer: Cham, Switzerland, 2016.
5. Cheng, W.K.; Yeh, Z.M.; Kao, H.Y.; Huang, S.H. Cross-Mesh Clock Network Synthesis. *Electronics* **2023**, *12*, 3410. [[CrossRef](#)]
6. Chou, C.H.; Lai, Y.T.; Chang, Y.C.; Wang, C.Y.; Cheng, L.C.; Huang, S.H.; Chang, S.C. Ping-Pong Mesh: A New Resonant Clock Design for Surge Current and Area Overhead Reduction. *IEEE Trans. Comput. Aided Des. Integr. Circuits Syst.* **2017**, *36*, 146–155. [[CrossRef](#)]
7. Challagundla, D.; Bezzam, I.; Islam, R. Design Automation of Series Resonance Clocking in 14-nm FinFETs. *Circuits Syst. Signal Process.* **2023**, *42*, 7549–7579. [[CrossRef](#)]
8. Wu, Q.; Pedram, M.; Wu, X. Clock-gating and its application to low power design of sequential circuits. *IEEE Trans. Circuits Syst. I Regul. Pap.* **2000**, *47*, 415–420.
9. Kim, C.Y.; Lee, H.C. Low-Power, High-Sensitivity Readout Integrated Circuit With Clock-Gating, Double-Edge-Triggered Flip-Flop for Mid-Wavelength Infrared Focal-Plane Arrays. *IEEE Sensors Lett.* **2019**, *3*, 3501404. [[CrossRef](#)]
10. Giustolisi, G.; Mita, R.; Palumbo, G.; Scotti, G. A Novel Clock Gating Approach for the Design of Low-Power Linear Feedback Shift Registers. *IEEE Access* **2022**, *10*, 99702–99708. [[CrossRef](#)]
11. Asgari, F.H.A.; Sachdev, M. A low-power reduced swing global clocking methodology. *IEEE Trans. VLSI Syst.* **2004**, *12*, 538–545. [[CrossRef](#)]
12. Wu, F.; Jia, S.; Wang, Y.; Zhang, G. Low swing drivers based on charge redistribution. *Sci. China Inf. Sci.* **2010**, *53*, 2377–2388. [[CrossRef](#)]
13. Prasanthkumar, B.; Fayaz, D.B.; Nishad, A.K. A Power-Efficient Clock Distribution Network with Novel Repeater. In Proceedings of the International Conference on Smart Electronics and Communication, Trichy, India, 10–12 September 2020.
14. Nekili, M.; Savaria, Y. Parallel regeneration of interconnections in VLSI & ULSI circuits. *Proc. IEEE Int. Symp. Circuits Syst.* **1993**, *3*, 2023–2026.
15. VADler, EGFriedman, Repeater Design to Reduce Delay and Power in Resistive Interconnect. *IEEE Trans. Circuits Syst. II* **1998**, *45*, 607–616.
16. Dhar, S.; Franklin, M.A. Optimum buffer circuits for driving long uniform lines. *IEEE J. Solid-State Circuits* **1991**, *26*, 32–40. [[CrossRef](#)]
17. Tretz, C.; Zukowski, C. CMOS transistor sizing for minimization of energy-delay product. In Proceedings of the Sixth Great Lakes Symposium on VLSI, Ames, IA, USA, 22–23 March 1996; pp. 168–173.
18. Khetarpal, V.; Gupta, L.; Dhand, R.; Sharma, P. Machine Learning Techniques for VLSI Circuit Design: A Review. In *Intelligent Systems Design and Applications. ISDA 2023. Lecture Notes in Networks and System*; Springer: Cham, Switzerland, 2024; pp. 191–199.
19. Trincherro, R.; Bradde, T.; Telescu, M.; Stievano, I.S. Modeling of IC Buffers from Channel Responses via Machine Learning Kernel Regression. *IEEE Electromagn. Compat. Mag.* **2023**, *13*, 84–87. [[CrossRef](#)]
20. Frankel, B.; Sarfati, E.; Wimer, S.; Birk, Y. Post-Silicon Analysis of Shielded Interconnect Delays for Useful Skew Clock Design. *IEEE Trans. Electron. Devices* **2019**, *66*, 4875–4882. [[CrossRef](#)]

21. Cherkauer, B.S.; Friedman, E.G. A unified design methodology for CMOS tapered buffers. *IEEE Trans. VLSI Syst.* **1995**, *3*, 99–111. [[CrossRef](#)]
22. Sakurai, T.; Newton, R. Alpha-power law MOSFET model and its applications to CMOS inverter delay and other formulas. *IEEE J. Solid-State Circuits* **1990**, *25*, 584–594. [[CrossRef](#)]
23. Adler, V.; Friedman, E.G. Delay and power expressions for a CMOS inverter driving a resistive-capacitive load. *Proc. IEEE Int. Symp. Circuits Syst.* **1996**, *4*, 101–104.
24. Ceresa, D.; Haranko, M.; Kloukinas, K.; Kaplon, J.; Caratelli, A.; Giovinazzo, D.; Mykyta, H.; Kaplon, J.; Konstantinos, K.; Murdzek, J.; et al. Characterization of the MPA prototype, a 65 nm pixel readout ASIC with on-chip quick transverse momentum discrimination capabilities. In Proceedings of the Topical Workshop on Electronics for Particle Physics, KU Leuven—Campus Carolus, Antwerpen, Belgium, 17–21 September 2018; Volume 166.
25. Ceresa, D.; Marchioro, A.; Kloukinas, K.; Kaplon, J.; Bialas, W.; Re, V.; Traversi, G.; Gaioni, L.; Ratti, L. Macro Pixel ASIC (MPA): The readout ASIC for the pixel-strip (PS) module of the CMS outer tracker at HL-LHC. *J. Instrum.* **2014**, *4*, C11012. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.