VLSI circuit reconstruction from mask topology *

N.P. van der Meijs

Laboratory for Network Theory, Department of Electrical Engineering, Delft University of Technology, 2600 GA Delft, The Netherlands

J.T. Fokkema

Section of Technical Geophysics, Department of Mining Engineering, Delft University of Technology, 2600 GA Delft, The Netherlands

Received 12 December 1983
Revised 20 March 1984

Abstract. This paper discusses whether and how parasitic circuit elements must be included in the circuit simulator source file to obtain reliable simulation results. In particular, attention is paid to fabrication tolerances, wire capacitance (including fringing effects), wire resistance (dispersive line effects), coupling capacitances and capacitances associated with contacts and the aspect ratio of (non-rectangular) transistors.

Keywords. Parasitic circuit elements, wire capacitance, fringing capacitance, coupling capacitance, wire resistance, dispersive line effects, transistor aspect ratio.

Introduction

In NMOS integrated circuit design, mask topology is created from circuit schematics. These schematics are usually optimized using a circuit simulation program, for example, SPICE [1]. Together with the fabrication process, the mask layout defines the actual circuit to be fabricated. In most cases, this actual circuit will be quite different from the one intended, due to circuit parasitics. It is now desirable to check circuit performance again, before the very expensive and time consuming fabrication step.

* This research has been partially supported by the 3744/81 Microelectronics Program of the EEC.

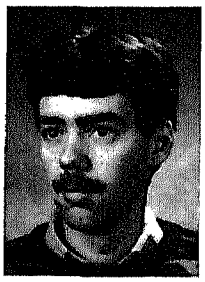
North-Holland
INTEGRATION, the VLSI journal 2 (1984) 85–119

Therefore, it is necessary to reconstruct from mask layout the actual circuit, including parasitics. This can be done by hand for small subcircuits only. For larger circuits, automatic extraction programs are needed [2,3].

This paper now analyses whether and how parasitic elements must be included in the simulator source file. It can be used as a guideline for manual reconstruction of small subcircuits or as a starting point to define an automatic circuit reconstruction program. Although it was written assuming an NMOS process and SPICE as the circuit simulator program, part of the results are in fact applicable to other technologies, while most of the results will do even for other circuit simulator (or higher level) programs.

In Section 1, fabrication tolerances are discussed. This is because it is felt that there is no need to strive for a perfect representation in the simulator source file of all parasitic circuit elements that can be extracted from the mask topology. Instead, an increase in simulation accuracy resulting from the inclusion in the simulator source file of some parasitic element must be compared to the variation of circuit performance due to fabrication tolerances. Further, it is indicated that a first step in the circuit reconstruction procedure must be a correction of mask layout. This is to compensate for a systematic deviation of wire width on the chip from the wire width as actually defined by the fabrication masks.

Next, in Section 2, diffusion capacitance is discussed. It is shown how diffusion capacitances that are hardly considered to be source or drain regions, for example, diffusion bridges connecting two metal wires, can be included in the SPICE source-file.



Nick P. van der Meijs was born in Maasland, The Netherlands, on November 1, 1959. He received his B.S. degree in Electrical Engineering from the Delft University of Technology in 1983. His interests include the whole area of micro-electronics, but are mainly in the area of the electrical and geometrical level of the VLSI circuit design and verification hierarchy. Presently, he is finishing his study at the Delft University of Technology on the subject of parameterized circuit layout.

Jacob T. Fokkema received his M.S. degree in Electrical Engineering from the Delft University of Technology in 1976. From 1976 to 1979 he was employed as a research assistant at the Netherlands Organisation for the Advancement of Pure Research (ZWO.). In 1979 he received his Ph.D. in Electrical Engineering. From 1979 to 1980 he was a staff member of the group of Network Theory, Department of Electrical Engineering of the Delft University of Technology. From 1980 to 1981 he was a research fellow at the Integrated Systems Laboratory at Stanford, U.S.A. In 1981 he rejoined the group of Network Theory at Delft University of Technology. In 1982 he became a staff member at the group of Technical Geophysics, Department of Mining Engineering, Delft University of Technology. His interests are mainly the areas of development of geophysical inversion techniques and possible hardware implementation of the resulting calculation schemes.

In Section 3, poly and metal capacitance are discussed. It is shown that fringing effects are important for an accurate calculation of the capacitance of poly and metal wires. A simple though accurate formula to estimate the fringing capacitances is presented.

In Section 4, wire resistance is discussed. Since both wire resistance and wire capacitance are in fact distributed along the wire, an appropriate method to include these effects in the simulator source file is needed. In particular, conditions are established under which the dispersive character of the wire is of importance or under which the wire can be represented in the simulator source file by a resistor, a capacitor or an ideal electrical node. Thereafter, a dispersive line model using one or more pi-sections is compared to the theoretical dispersive line solution.

In Section 5, capacitances between different wires are discussed. Coupling capacitances associated with crossing or overlapping wires are modeled using the results of Section 3, that is, including fringing capacitances. Also in this section, capacitances associated with contacts between wires on different layers are discussed, because the situation is very much the same as the situation with unconnected overlapping wires.

In Section 6, attention is paid to transistor dimensions. In particular, an analogon method to determine the aspect ratio of non-rectangular transistors is proposed.

1. Processing variations

In integrated circuit technology, the final dimensions of all structures (wires, transistors, capacitors etc.) on finished wafers usually differ from their drawn value (see, for example, Domenik [4] and Cohen [5]). Due to several effects (for example, lateral expansion of local oxidation and imperfect etching) wire width is usually reduced compared to the drawn (intended) value. This reduction of line width is not a mask-dependent constant but it shows some spreading for different wafers. However, it can usually be assumed that the reduction does not vary across a given chip. Realistic minimum, typical, and maximum line-width reduction values (in microns) for a 4-micron process are as follows:

	min	typ	max
poly	0.4	0.8	1.2
diff	0.7	1.1	1.5
metal	1	2	3

It is observed that, for example, a poly wire drawn to be 4 microns wide will typically turn out to be only 3.2 micron wide, with a statistic tolerance as large as about 12%. A diffusion wire drawn to be 4 micron wide would come out to be only 2.9 micron with 25% variation to both sides, and a 6-micron drawn metal wire will turn out to be only 4 micron wide with 25% variation.

1.1. Impact on circuit performance

The typical line-width reduction is in fact a systematic line-width reduction that can introduce relatively large capacitance (and, consequently, timing) variations. Further, since transistor dimensions are determined by the widths of crossing poly and diffusion wires (and by the lateral diffusion of source and drain), transistor length to width ratio and inverter pull up–pull down ratio can vary appreciably from the intended value. The resulting variations in circuit performance can sometimes be larger and more serious than one would expect at first sight. To illustrate this, an example is given where the relative small poly tolerances can still give a delay tolerance of 25%.

Consider the situation where the gate of a certain transistor must be charged or discharged through a pass transistor. These transistors are assumed to be connected by a wire short enough to neglect its resistance and capacitance. The delay in this situation is proportional to the quotient of gate capacitance of the load transistor and the current delivered by the pass transistor, that is proportional to its W/L ratio. Since, to first order, the W/L ratio is inversely proportional to poly width and load transistor capacitance is proportional to poly width, delay for this situation would be proportional to the square of poly width. Consequently, a 12% variation of poly width that can occur as mentioned previously would cause a variation in delay time of about 25%. However, in practice, fringing capacitances and gate-drain and gate-source overlap capacitances provide a stabilizing effect on delay time variations since they are more or less constant. It is clear that one must be aware of these variations during both circuit design (layout design) and analysis (reconstruction and simulation) still.

1.2. Impact on circuit design and analysis

During design it might be necessary not to use minimum size configurations. This might be, for example, the case when delay time must be between two boundaries to avoid undesired hazards. However, since there are many other (not layout related) process parameter variations (like variations in threshold voltage, sheet resistance, oxide thickness etc.), these boundaries cannot be too close to each other. Further, because of mask alignment tolerances, overlaps or small spacings between wires of different levels must sometimes be avoided since relative large variations in the associated coupling capacitances (see Section 5) can occur.

During analysis it is at least mandatory as a first step in circuit reconstruction to correct the dimensions of the drawn layout for the typical line width reductions. For example, neglecting this correction for the configuration as discussed above (delay time proportional to the square of the poly width) would cause an over-estimation of the typical delay time of up to 56%! Implanted undercrossing regions also need a width correction. However, this correction must be applied to an effective undercrossing region obtained from logically anding the undercrossing and the diffusion region. This follows from processing sequences. Now, due to

all remaining tolerances and variations, relative large safety margins must still be used in judging the simulated circuit performance against the design criteria. Statistical methods may possibly be used to make an estimation of the relation between safety margin and the chance for a finished circuit to match the specifications. Also, for some critical designs it might be useful to perform min–max or worst case simulations. A certain margin must then be taken into account still, if only in accordance with simulation accuracy or the impossibility to extract from the layout all parasitic circuit elements.

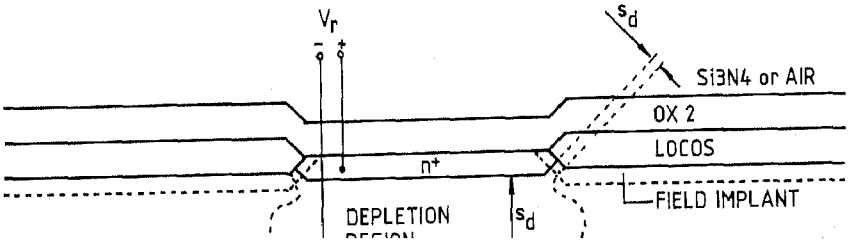
The other way round, however, the large circuit performance variations relieve simulation accuracy requirements. It has no meaning to strive for a circuit simulation within an accuracy of say 3%, when the circuit exhibits performance variations of up to 30%. This is the reason that, for example, modeling of the distributed nature of resistance and capacitance of poly and diffusion wires as proposed in Section 5 is usually satisfying. Moreover, it is emphasized now that the needed simulation accuracy does not automatically become larger with process down-scaling. This is because in all processes, at least while dimensions are still far away from lower bounds imposed by physical laws, minimum dimensions are determined by the relative accuracy with which they can be manufactured.

2. Diffusion capacitance

Consider Fig. 2.1. Diffusion wires in an NMOS process are heavily n-doped regions in a relative lightly p-doped bulk. In case the voltage of a diffused layer is positive relative to the substrate, we have a reverse biased pn + junction while the associating depletion layer, that is formed mainly in the relative lightly p-doped bulk, acts as an isolator between the wire and the substrate. Thus, a capacitance is formed between diffusion and substrate. It is this capacitance that is discussed in this section.

Since the depletion region is surrounding the wire (see Fig. 2.1), the capacitance is conveniently split up in a bottom capacitance and a sidewall capacitance.

Although in this section diffusion capacitance is discussed in particular, the following reasoning is also valid for other interconnects with depletion layer isolation, such as implanted undercrossings [6].



2.1. Bottom capacitance

The width s_d of the depletion layer depends on the voltage applied to the junction. Apart from a small leakage current, the depletion layer acts as an isolator between two conducting regions. So, a capacitor is formed with small signal capacitance per unit area:

$$C_d = \epsilon_0 \epsilon_s / s_d ,$$

where

ϵ_0 = dielectric constant of vacuum (8.85 pF/m),
ϵ_s = relative dielectric constant of silicon (11.4),
s_d = width of depletion layer.

Because the width of the depletion layer depends on the applied voltage, so does the depletion layer capacitance. In fact, one can write for the small signal capacitance per unit area of a reverse-biased pn+ junction [7]:

$$C_d = \sqrt{e\epsilon_0\epsilon_s N_s / 2(V_r + \varnothing_b)} ,$$

where

e = elementary charge ($1.6 * 10^{-16}$ C),
N_s = substrate doping concentration, for example $450 * 10^{12}/\mathrm{cm}^3$,
$\varnothing_b$ = built-in voltage of the junction, approximately 0.7 V in most cases,
V_r = applied reverse voltage between diffused layer and bulk.

To get an idea of the voltage dependency, let

$$C_0 = C_d(V_r = 0) = \sqrt{e\epsilon_0\epsilon_s N_s / 2\varnothing_b} .$$

Hence,

$$C_d = C_0 / \sqrt{1 + (V_r / \varnothing_b)}$$

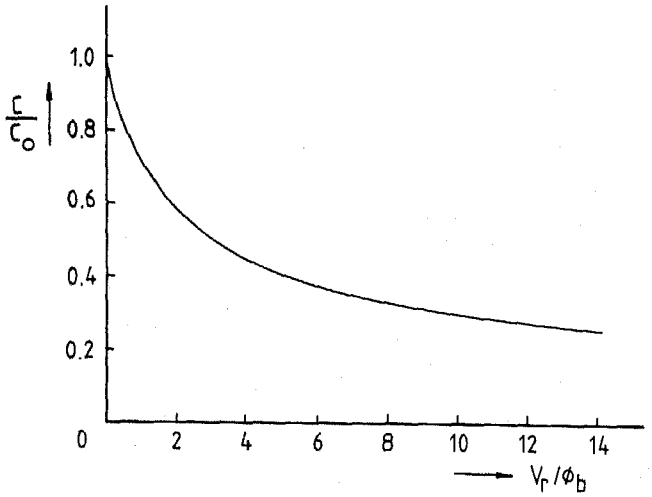


Fig. 2.2. Normalized capacitance-voltage characteristics for a reverse biased pn+ junction.

and this relationship is shown in Fig. 2.2. A large voltage dependency is clear now. This must be accounted for in circuit simulation.

2.2. Sidewall capacitance

The capacitance associated with the sidewall depletion layer of the wire again consists of two parts. First, there is the sidewall capacitance introduced by the bending of the depletion region around the wire. Second, there is a relative large contribution from the region where the doping concentration is defined by the field or stopper implant. Although the area is small, the capacitance per unit area is very high because of the high field implant concentration and the associating small width of the depletion layer. (For example, the field implant concentration can be $5*10^{16}/\text{cm}^3$.) (See Fig. 2.1 where the width of the depletion layer is indicated for both regions.) It follows that the sidewall capacitance is important for (narrow) diffusion wires. However, since the doping profiles are well defined, the junctions have well-defined and uniform shapes. Thus, it is appropriate to account for the sidewall capacitance in Farad/unit length. Only the variation of sidewall capacitance with junction voltage can be a little different compared to the voltage variation of bottom capacitance. Usually, it is modeled as follows:

$$C_d = C_0/(1 + V_r/\varnothing_b)^p, \quad 0 < p < 1,$$

where C_d and C_0 are now capacitances per unit length. Finally, there is another (small) capacitance component. It is the capacitance to substrate of the lateral diffusion of source and drain under the gate of the transistor. Since the lateral diffusion length is a constant, the capacitance is proportional to the width of the transistor. Again, this capacitance is voltage-dependent.

2.3. SPICE modeling of diffusion capacitance

To account for the nonlinear diffusion capacitances in circuit simulation, the SPICE program has included a diffusion capacitance model for transistor source and drain terminals. This model includes both bottom and sidewall capacitance and accounts for junction leakage current and forward biasing, too. For every transistor one can specify the area and perimeter of both source and drain. In this way, all diffused regions must be modeled as a source or drain region.

This is not a very straightforward method. For example, a diffusion region can be a source and/or drain of several transistors together, e.g., in a NOR-gate. It is not obvious now what the source and drain dimensions are for every particular transistor. Also, a diffusion region can be used to connect two metal wires, and it might be artificial to model the diffusion region as a source or drain region of a particular transistor. However, straightforward methods are usually most advantageous since they are easiest to implement in an automatic extraction algorithm and they reduce the chance to make a mistake during manual extraction. Further, one must still include the third capacitance component (proportional to the width of the transistor), but neglecting this capacitance does not introduce large errors in most cases.

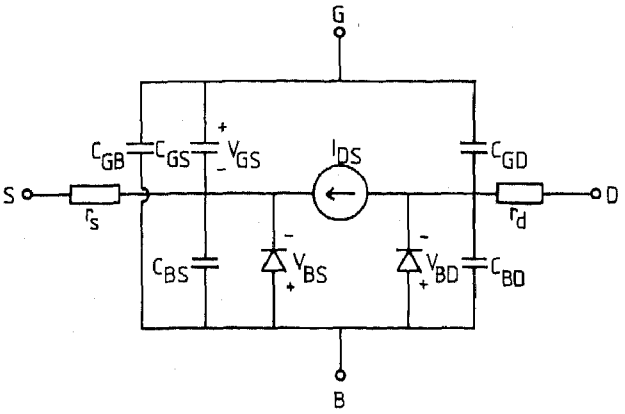


Fig. 2.3. SPICE2 MOSFET model.

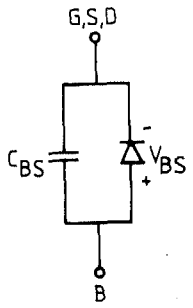


Fig. 2.4. Diffusion capacitance model derived from the model of Fig. 2.3.

To circumvent the problem of modeling every diffusion area as a source or drain region, one can use two SPICE diode models. One diode must then be scaled to the area and one to the perimeter of the diffusion region.

An alternative possibility is to use a 'stripped' transistor model. In Fig. 2.3 the complete SPICE MOSFET model as described in [8] is given. In this figure, r_s and r_d are the series resistance of source and drain respectively. Further, C_{GB}, C_{GS} and C_{GD} are voltage-independent overlap capacitors. C_{BS} and C_{BD} are the substrate–source and substrate–drain junction capacitances that are scaled to the appropriate source and drain area and perimeter. Now, one can specify in the transistor model r_s, r_d, C_{GB}, C_{GS} and C_{GD} to be zero. Further, connecting gate, source and drain together and also setting the area and perimeter of drain (or source) equal to zero, we have our desired diffusion model (see Fig. 2.4).

In case one of these methods is applied to all diffusion areas, it is very easy and straightforward to account for the lateral diffusion junction component under the gate of the transistor. This capacitance component can now be included in the transistor model, as a true, physical, source or drain region.

3. Poly and metal capacitance

A poly or metal wire is a conductive layer with dielectric isolation from substrate and from other wires. The dielectric is often SiO2 (see Fig. 3.1). The

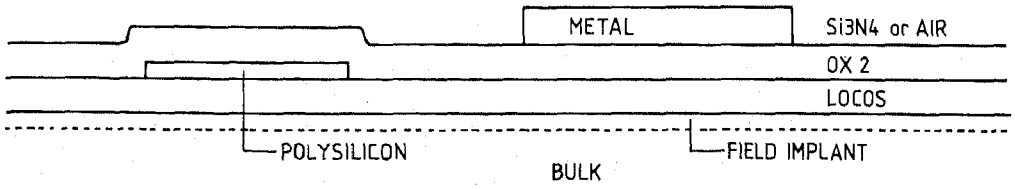


Fig. 3.1. Poly and metal interconnect geometry.

wire has capacitance to substrate, that is to be discussed in this section, and to other wires that is to be discussed in Section 5.

However, simple parallel-plate calculations are shown not to be accurate enough for today's and future's integrated circuits. This has already been reported on many occasions [10–18]. Instead, a fringing capacitance term must be added to the parallel-plate capacitance term. A simple though accurate formula to estimate fringing capacitance when the geometry is known is given.

3.1. Parallel-plate capacitance

With zero voltage on a poly or metal interconnection layer (see Fig. 3.1) relative to substrate, the capacitance of this layer to substrate when neglecting fringing effects is equal to the capacitance of a parallel-plate capacitor with the oxide SiO2 as dielectric and separation between the plates equal to the oxide thickness:

$$C_{pp} = A * (\epsilon_0 \epsilon_r / h_{ox}) = A * C_a,$$

where

$\epsilon_0 \epsilon_r$ = dielectric constant, for $SiO_2 = 3.9 * 8.85 * 10^{-12}$,
h_{ox} = height of conductor above substrate, equal to the oxide thickness,
A = area of the conductor.

Now, $C_a = (\epsilon_0 \epsilon_r / h_{ox})$ is a capacitance per unit area and is characteristic for the layer under consideration, poly and metal both have their own area capacitance.

But, with increasing voltage a depletion layer will arise under the oxide just as under the gate of a transistor (see, for example Mead and Conway [9, p. 55]). The voltage-dependent depletion capacitance effectively adds in series to the oxide capacitance, but because of the high oxide thickness and field-implant doping level, its influence is only very small. In fact, it can be calculated from [7] that for $h_{ox} = 0.6$ micron and field implant concentration equal to $5 * 10^{16}/cm^3$ a small-signal capacitance variation of only 2% will result from a 5-volt voltage variation. Comparing with processing tolerances it is clear that the nonlinearity of poly and metal capacitances is relatively unimportant. Therefore, poly and metal capacitance (parallel-plate as well as fringing capacitance) is considered in the following to be constant with voltage.

3.2. Fringing capacitance, two-dimensional

For accurate capacitance calculations it is necessary to add to the parallel-plate capacitance the fringing capacitance of ground plane to top side and sidewalls (see Fig. 3.2). We now have

$$C = C_{pp} + C_{fring}.$$

where C_{fring} can form a large part of C. In fact, in the two-dimensional case, the fringing capacitance is a function of the ratios t/h and w/h, where t is the conductor thickness, h its height above substrate and w its width (see Fig. 3.2).

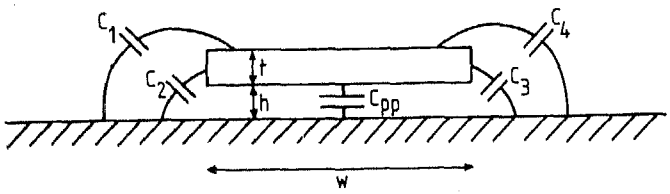


Fig. 3.2. Parallel plate capacitance C_{pp} and fringing capacitances C_1, C_2, C_3 and C_4.

For example, from the very accurate formula of conductor capacitance given by Chang [12] one can construct Fig. 3.3 where fringing capacitance is given per unit wire length. It must be noted that with Chang's formula a homogeneous dielectric is assumed to fill the upper half-space, which is different from practice in IC technologies (see Fig. 3.1). Also, conductor cross-section might be more or less trapezoidal in practice while Chang's formula deals with rectangular cross-sections only. So, it is possibly the best to extract capacitances from test chip characterization.

However, Chang's formula can provide a useful estimation of capacitance in case t, h and w are known. For example, the results obtained by Dang and Shigyo [13] indicate that the dependency of conductor capacitance to a slight trapezoidality of conductor cross-section is not very spectacular. Also, for the geometry of Fig. 3.4 (without a homogeneous dielectric but more resembling practice), Dierking and Bastian [14] numerically found fringing capacitance to be 101 pF/m. This

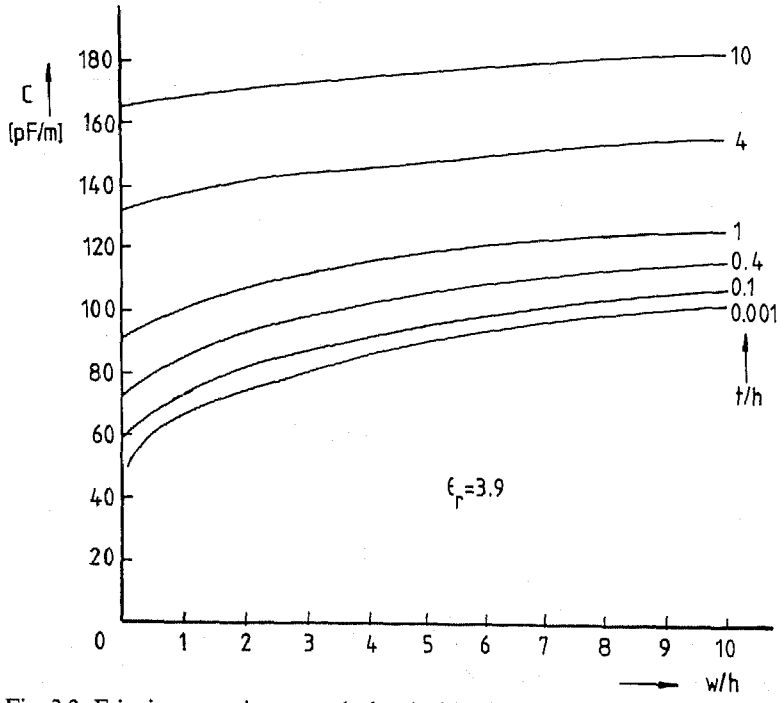


Fig. 3.3. Fringing capacitances calculated with Chang's formula.

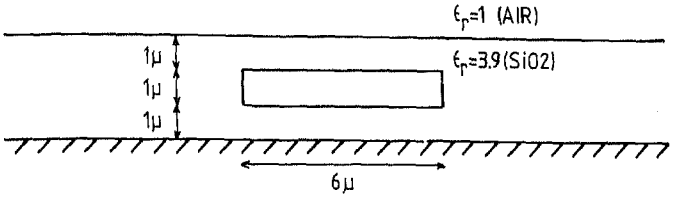


Fig. 3.4. Interconnect geometry more resembling practice.

is an acceptably small difference compared to Chang's formula (assuming a homogeneous dielectric) that would give $C = 120$ pF/m for an overestimation of wire capacitance of 6%. The reason for this is that only a small portion of fringing field lines originate in the top side of the conductor. This is also the reason that (above a certain minimum) fringing capacitance is relatively independent of conductor width, as can be seen from Fig. 3.3.

3.2.1. Simple wire capacitance approximation formula

Unfortunately, Chang's formula, that was obtained by conformal mapping methods, is mathematically very complex. It is not very well suited for preliminary IC design or implementation in CAD programs. Therefore, several authors [15,16,17] presented simpler empirical formulas to approximate conductor capacitance. The simplicity of these formulas is always paid for with a certain loss of accuracy. A relative small loss of accuracy goes with a formula that was recently proposed by Sakurai et al. [16]. It is claimed to be accurate within 6% for $0.3 \leqslant w/h \leqslant 30$ and $0.2 \leqslant t/h \leqslant 30$:

$$C/\epsilon_0\epsilon_r = 1.15*(w/h) + 2.80*(t/h)^{0.222}. \tag{3.1}$$

Although the accuracy will be satisfactory for most purposes, it must be noted that a t/h ratio of 30 will not be encountered among IC wires. With this in mind, an IC wire capacitance approximation formula was developed from curve-fitting on results obtained from Chang's formula for a reduced range of t/h ratios. This approximation matches Chang's formula within 2% for $w/h \geqslant 1$ and $0.1 \leqslant t/h \leqslant 4$, and within 6% for all $w/h \geqslant 0.3$ and $t/h \leqslant 10$:

$$C/\epsilon_0\epsilon_r = w/h + 0.77 + 1.06*(w/h)^{0.25} + 1.06*(t/h)^{0.5}. \tag{3.2}$$

In this formula the first term at the right-hand side accounts for the parallel-plate capacitance and the three other terms account for the fringing capacitance. In Table 3.1 a comparison is made between Sakurai's formula (3.1) and the improved formula (3.2) with numerical computer calculations available from Chang and (the last two lines) from Ruehli and Brennan [11]. The contrast is clear.

As was mentioned before, and as can also be seen from (3.2), fringing capacitance does not depend very strongly on wire width. Advantage can be taken from this by substituting $w/h = 3$ in the formula to obtain a very simple capacitance formula, where the fringing capacitance is independent of wire width,

Table 3.1
Comparison of Sakurai's formula (3.1) and improved formula (3.2) with computer calculations according to Chang [12] and Ruehli and Brennan [11] (last 2 lines)

w/h	t/h	Computer calculation	Improved formula (3.2)	Relative error of (3.2)	Relative error of (3.1)
1.12	0.318	3.591	3.578	−0.4%	−3.7%
2.01	0.485	4.793	4.780	−0.3%	−2.0%
2.52	0.318	5.242	5.223	−0.4%	−3.3%
2.72	0.802	5.815	5.801	−0.2%	−0.4%
3.18	0.936	6.394	6.391	< 0.05%	−0.4%
3.63	0.802	6.820	6.812	−0.1%	+0.3%
3.68	0.485	6.661	6.656	−0.1%	−0.7%
3.79	1.116	7.159	7.159	< 0.05%	+1.0%
4.24	0.936	7.544	7.557	+0.2%	+1.2%
5.44	1.200	8.978	8.990	+0.1%	+2.2%
6.36	0.802	9.754	9.763	+0.1%	+2.3%
7.42	0.936	10.935	10.965	+0.3%	+3.3%
8.21	1.805	12.140	12.198	+0.5%	+4.1%
9.85	1.805	13.842	13.922	+0.6%	+5.0%
11.90	1.453	15.847	15.916	+0.4%	+5.6%
14.78	1.805	18.974	19.052	+0.4%	+6.4%
22.22	4.070	27.077	27.430	+1.3%	+8.5%
58.25	7.113	63.943	64.775	+1.3%	+11.5%
6.25	0.625	9.548	9.543	−0.1%	+1.7%
2.50	0.500	5.367	5.352	−0.1%	+1.7%

Note: From the error percentages in Table 3.1 of Sakurai's formula for large w/h ratio's, it can be seen that the reference used by Sakurai to estimate the accurateness of his formula is not the same as the reference used by Chang, that was also used by us.

with an accuracy of 5% for $w/h \geqslant 2$ and $0.1 \leqslant t/h \leqslant 5$:

$$C/\epsilon_0\epsilon_r = w/h + 2.16 + 1.06 * (t/h)^{0.5}. \tag{3.3}$$

Since the t/h ratio is characteristic for the interconnection layer under consideration, the fringing capacitance per unit wire length can be specified for each wiring layer, independently of wire width above a certain minimum ($w/h \geqslant 2$). Thus, capacitances of long (so that the end capacitances can be neglected) and not too narrow wires could easily be calculated using

$$C = w * l * C_a + l * C_f,$$

where

C_a = the area capacitance, the parallel-plate capacitance per unit area for the wiring level under consideration,

C_f = the fringing capacitance per unit wire length for the wiring level under consideration.

3.3. *Fringing capacitance, three-dimensional*

For wires that are not very long, a three-dimensional approach must be followed for high accuracy. But, since small wires can only have a small influence

on delay time and voltage waveforms, and because they show relative large capacitance variations due to line-width and alignment tolerances, a reduced accuracy usually suffices. Therefore, a pseudo three-dimensional approximation of fringing capacitance usually satisfies for all conductors. These conductors need not necessarily be rectangular. The pseudo three-dimensional approximation is obtained by defining an edge capacitance C_e per meter equal to half the two-dimensional fringing capacitance per unit wire length C_f and by multiplying this edge capacitance through the perimeter of the conductor. The total capacitance of a conductor can then be written as

$$C = A * C_a + P * C_e,$$

where

A = the conductor area,
P = the conductor perimeter,
C_a = the parallel-plate capacitance per unit area,
C_e = the edge capacitance per unit length,
$C_e = 0.5 * C_f = \epsilon_0 \epsilon_r (1.08 + 0.53\ t/h^{0.5})$ [F/m] (or measured).

3.3.1. Comparison with numerical results

Consider a square conductor, 4 micron each side, thickness 1 micron and height above substrate 1 micron. The dielectric is SiO_2 ($\epsilon_r = 3.9$). We now have

$$C_a = \epsilon_0 \epsilon_r / h_{ox} = 34.5\ \mu\text{F/m}^2,$$

$$C_e = \epsilon_0 \epsilon_r (1.08 + 0.53 t/h^{0.5}) = 55.6\ \text{pF/m}.$$

Hence,

$$C = 16 * 10^{-12} * C_a + 16 * 10^{-6} * C_e = 1.44\ \text{fF}.$$

Now, Ruehli and Brennan [10] numerically found fringing capacitance for this geometry to be two times the parallel-plate capacitance, being 0.552 fF. Therefore, C would be 1.66 fF. Thus, our pseudo three-dimensional approximation underestimates the capacitance of this geometry with only 13%. This is a considerable and useful improvement compared to parallel-plate (for an underestimation of capacitance of 67%) or two-dimensional calculations (underestimation of 40%).

3.4. Some remarks

It must be clear now that fringing capacitances are important. In particular, one must realize that these fringing capacitances (and coupling capacitances, to be discussed in Section 6) will relatively increase with process down-scaling. The reason for this is that the height of the wires above the substrate will preferably be held as high as possible for small capacitances and to maintain a high field inversion voltage. Moreover, the wires must be as thick as possible for low sheet resistance (poly) and high current capability (metal) [13,18,19]. Thus, in future it will become even more important to account for fringing fields.

For example, consider two processes with metal and poly interconnection; see below for some realistic dimensions belonging to a current process, and a future scaled-down process (the dimensions are in microns):

	poly		metal	
	present	future	present	future
t	0.25	0.25	0.9	0.9
h	0.6	0.6	1.2	1.2
w	4.0	2.0	6.0	3.0

Let us now calculate the capacitances per unit length for the geometries given above using (3.2) to accurately show dependence on conductor width (see below for the results ($\epsilon_r = 3.9$)) (capacitances are in pF/m):

	poly		metal	
	present	future	present	future
parallel-plate-term	230	115	172	86
fringing-term	109	100	113	104
total	339	215	285	190

From these results the (increasing) importancy of fringing capacitances is very obvious.

4. Wire resistance

Apart from capacitance, a wire that is formed in metal, poly or diffusion has resistance, too. It is common practice to express this resistance as

$$R = (l/w) * R_s,$$

where R_s (sheet resistance) is the resistance of a square piece of material for a direction of current orthogonal on one of the sides. Realistic values of sheet resistance in present IC technologies are as follows:

diff $30\Omega/\square$, poly $30\Omega/\square$, metal $30\text{m}\Omega/\square$.

While it can be seen that metal resistance is relatively small, poly and diffusion resistance can become important when used for long interconnects [18,19].

For further calculations it will appear to be convenient to define for a wire the capacitance and resistance per unit length, supposing current flows in the length direction, as follows:

$$c = C_a * w + 2C_e \quad [\text{F/m}], \qquad r = R_{\text{sheet}}/w \quad [\Omega/\text{m}].$$

Here, w is the wire width, C_a and C_e are defined in Section 3 and R_s is defined above. Lower case symbols c and r are used to indicate that the resistance and capacitance are distributed along the wire. This is characteristic for IC technolo-

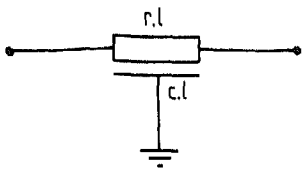


Fig. 4.1. Dispersive interconnect schematic.

gies, but special attention is needed for circuit simulation. A simple though accurate way of modeling the dispersive character of the wire is proposed.

For convenience, we note that a (uniform) *rc*-line will often be depicted as shown in Fig. 4.1.

4.1. Dispersive line theory

Consider a rectangular wire of length l, and with resistance r in Ohm/unit length and capacitance c in Farad/unit length (see Fig. 4.2). Suppose, at time $t = 0$ a voltage step is applied to one end of the wire. What is then the response $v(x, t)$ for $t > 0$ and $0 \leqslant x \leqslant l$?

Neglecting any nonlinearities of capacitance (and resistance) it is easy to verify using Ohm's law and the definition of capacitance, that

$$\partial v/\partial x = -ri, \qquad \partial i/\partial x = -c\partial v/\partial t,$$

and eliminating i we get

$$\partial^2 v/\partial x^2 = rc\partial v/\partial t.$$

This is the well-known diffusion equation. Its solution for a voltage step input (step magnitude: V_{in}) and with $v(x, t) = 0$ for $t < 0$ and $i(l, t) = 0$ is a standard textbook problem (see, for example, Boyce and Diprima [20]). Now, $v(x, t)$ for $0 \leqslant x \leqslant l$ and $t > 0$ is given by

$$\frac{v(x,t)}{V_{\text{in}}} = 1 - \sum_{n=1}^{\infty} \frac{4}{(2n-1)} \exp\left\{\frac{-(2n-1)^2\pi^2 t}{4l^2 rc}\right\} \sin\left\{\frac{(2n-1)\pi x}{2l}\right\}. \tag{4.1}$$

Fig. 4.3 shows normalized responses for $x/l = 0.1$, 0.5 and 1 respectively, together

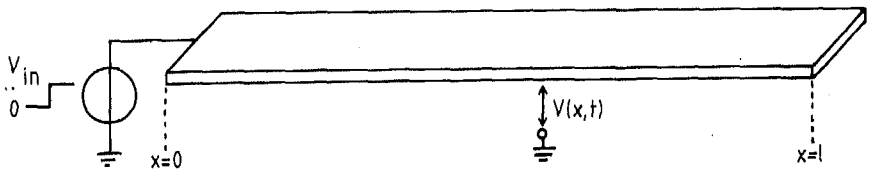


Fig. 4.2. Dispersive interconnect geometry.

with the response of a concentrated RC stage with $R = rl$ and $C = cl$. For $x = l$ and t large enough, say $t > rcl^2$, the following approximation is very accurate:

$$\frac{v(l,t)}{V_{in}} = 1 - \frac{4}{\pi}\exp\left\{\frac{-\pi^2 t}{4l^2 rc}\right\}. \tag{4.2}$$

4.2. Dispersive line connecting source and load

From the above theory, and also from Fig. 4.3, we might conclude that a poly or diffusion wire introduces in the circuit a propagation delay proportional to the square of its length. This was already mentioned by several authors (see, for example, Mead and Conway [9]). However, in practice the situation is somewhat different because of the finite output resistance R_o of the driving stage and the gate capacitance C_o of the load transistor (see also Bilardi et al. [19]):

(1) With large C_o compared to cl (and no source resistance R_o) the wire capacitance can be neglected. On the other hand, wire resistance is important since it is the only resistance that is present and nearly all the charge delivered by the driving stage must pass this resistance. Now, the wire behaves as a concentrated resistor connected between the driving stage and C_o.

(2) With large R_o compared to rl (and no load capacitance C_o) the resistive effect of the wire is a minor effect and can be neglected. However, the charge storage effect of the wire is fully manifested because wire capacitance is the only capacitance that is present. Thus, the wire behaves as a concentrated capacitor connected between R_o and ground.

(3) With both $R_o \gg rl$ and $C_o \gg cl$, wire resistance as well as wire capacitance can be neglected. In this case, the wire behaves as an ideal electrical node.

(4) Only in case neither $R_o \gg rl$ nor $C_o \gg cl$, the dispersive character of the wire is of importance, and must be accounted for in the simulator source file.

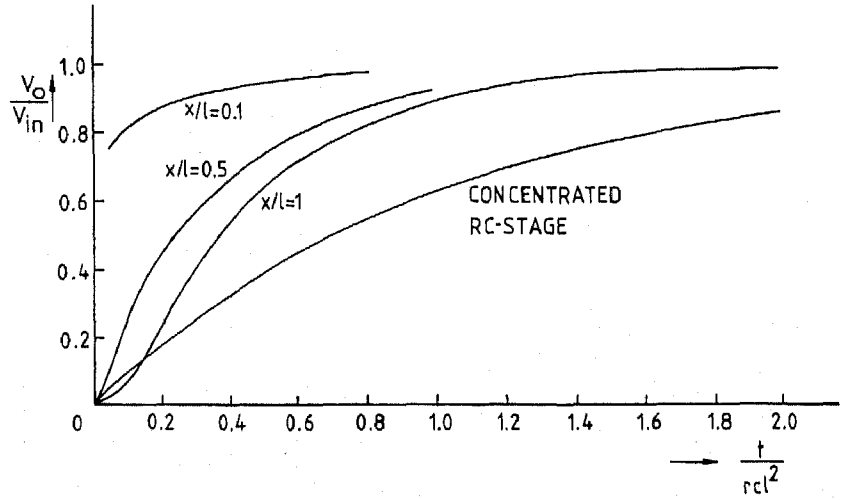


Fig. 4.3. Voltage step responses for a dispersive line and a concentrated RC-stage.

4.2.1. Error introduced by neglecting dispersive effects

We can now easily estimate the upper bound of the error that arises in neglecting wire resistance or capacitance. This maximum error will appear to depend on the ratio rl/R_o and cl/C_o. Consider therefore a voltage step applied to the input of the wire and define the response time t_r as the time needed for the output voltage to reach 90% of the final value. For a simple RC stage (concentrated R and C), this response time is approximately given by $t_r = 2.3\ RC$. Furter, ϵ_t is the relative error in the response time of a certain approximation compared to the actual response time. Now suppose the wire is driven from a source with impedance R_o and there is no load capacitance (see Fig. 4.4(a)). It is clear, that the response is faster than the response of the circuit shown in Fig. 4.4(b):

$$V_o = 1 - \exp\{-t/(R_o + rl) * cl\}, \qquad t_{r_1} = 2.3(R_o + rl) * cl,$$

but slower than the response of the circuit of Fig. 4.4(c):

$$V_o = 1 - \exp\{-t/(R_o * cl)\}, \qquad t_{r_2} = 2.3(R_o * cl).$$

Because

$$t_{r_2} < t_r < t_{r_1} \quad \text{and} \quad t_{r_1} = t_{r_2}(1 + rl/R_o),$$

we have

$$\epsilon_t < rl/R_o.$$

In case the wire is loaded with a capacitance C_o (and there is no source resistance) (see Fig. 4.5(a)), the response will be faster than the response of the circuit shown in Fig. 4.5(b):

$$V_o = 1 - \exp\{-t/rl(C_o + cl)\},$$

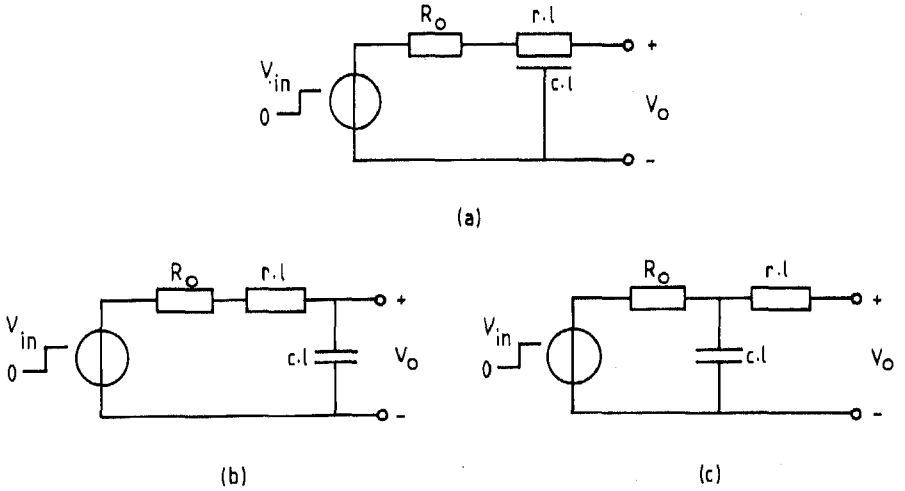


Fig. 4.4. Estimating the influence of the dispersive character of a wire with zero load capacitance.

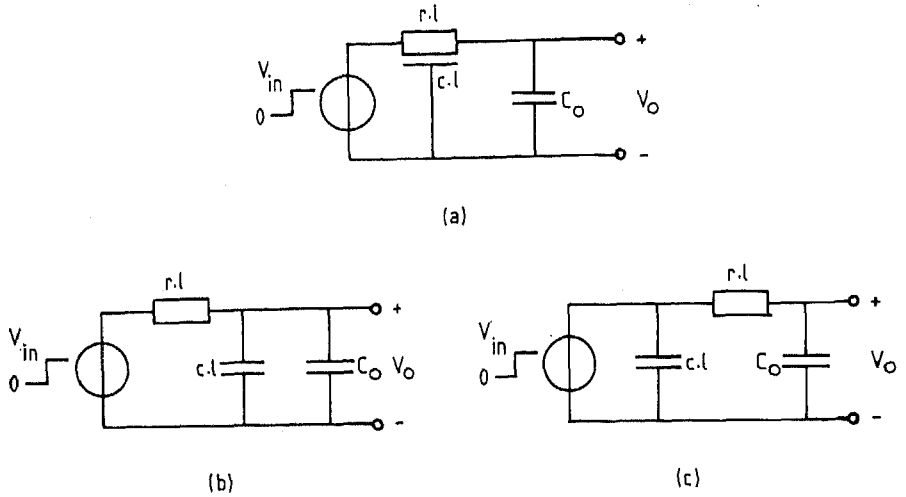


Fig. 4.5. Estimating the influence of the dispersive character of a wire with zero source resitance.

but slower than the response of the circuit of Fig. 4.5(c):

$$V_o = 1 - \exp\{-t/rl * C_o\}.$$

Hence,

$$\epsilon_t < cl/C_o.$$

From this analysis one can conclude that, for a given admissible response time error (denote this admissible ϵ_t by ϵ), it is permitted to model the wire for circuit simulation as follows:
- a single (concentrated) resistor ($R = rl$) if $cl/C_o < \epsilon$;
- a single (concentrated) capacitor ($C = cl$) if $rl/R_o < \epsilon$.

Further, one can neglect both wire resistance and capacitance in case $rl/R_o < \epsilon$ together with $cl/C_o < \epsilon$. Only in case $rl/R_o > \epsilon$ together with $cl/C_o > \epsilon$, the dispersive character of the wire must be modeled (see Fig. 4.6 where an ϵ of about 0.05 is considered to be in accordance with process tolerancies, as discussed in Section 1).

4.2.2. Practical driving stages

So far we dealt with a source with resistive output impedance R_o. This, however, is not exactly the case in practical circuits. Looking at the I_{DS} versus V_{DS} curves of a MOS transistor (see Fig. 4.7) it can be seen that there is only a limited region (region I), that can be considered to be resistive. To determine whether wire resistance has some influence on response time, we note that in region I the transistor output resistance can be approximated by

$$R_o = 1/\beta(V_{GS} - V_T).$$

This R_o-value corresponds to the straight diagonal line in Fig. 4.7. Thus, in the

linear region, wire resistance can be omitted whenever

$$rl < \epsilon/\beta(V_{GS} - V_T).$$

This formula must be applied to all transistors that can drive the line, eventually accounting for the fact that (for example, in an OR-gate) more transistors can drive the line simultaneously.

In the saturation region, the transistor effectively acts as a current source. However, we will not make a detailed analysis of the case of a current source driving a dispersive wire as was done by Bilardi et al. [19]. Instead, it is simply noted that the current flowing into the wire is smaller than the current that would flow in case the driving transistor was ideally resistive for the whole V_{DS} region. (A constant drive for the transistor is assumed.) With smaller current flowing into the wire, its resistance plays a less significant role. Thus, the linear region obviously serves as a worst case situation for determining whether wire resistance can be neglected.

At the moment of extracting the circuit, however, it is not yet known what the value of $V_{GS} - V_T$ is that we must use the formula above. In principle, this formula must be evaluated for the highest $V_{GS} - V_T$ value that could possibly occur, because a high $V_{GS} - V_T$ gives a low output resistance and when wire resistance can be neglected for this high $V_{GS} - V_T$, it can be neglected for all $V_{GS} - V_T$ values. (Note that we consider $V_{GS} - V_T$ instead of V_{GS} alone, since V_T is not a constant but increases with increasing source-bulk voltage V_{SB}.)

Determining the highest $V_{GS} - V_T$ value can, however, sometimes be tricky. It is not always equal to $V_{DD} - V_T$, as would be the case for the pull-down transistors in a chain of E/D inverters. (Note that in this case it depends, due to the ratio factor, on the pull-down transistor whether wire resistance can be neglected. Also in this case, we could consider V_{GS} alone since V_{SB} and thus V_T is

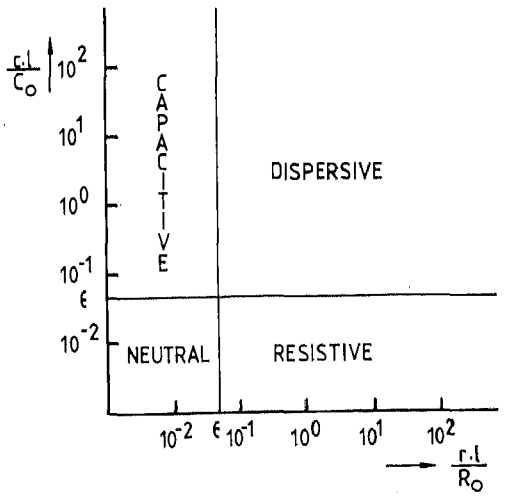


Fig. 4.6. Characterizing a dispersive line connecting source and load.

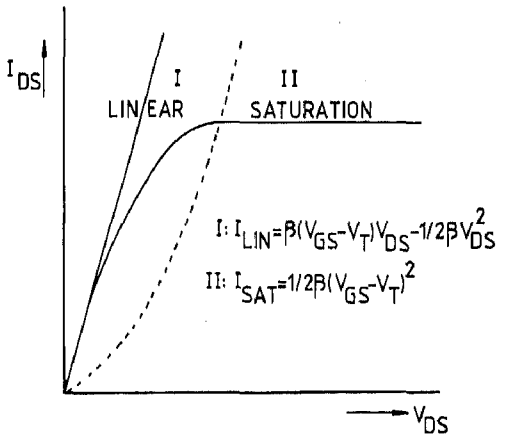


Fig. 4.7. Current–voltage characteristics of a MOS transistor.

a constant.) For example, the gate can be driven above V_{DD} by means of bootstrapping, or the source can be connected to substrate in stead of V_{SS} [21].

In fact, instead of trying to extract the maximum $V_{GS} - V_T$ directly from the layout (or from the electrical circuit to which the layout corresponds), it may be wiser and more accurate to use some information available from a preliminary circuit simulation run. This run eventually uses estimated/predicted values for wire capacitances. Now, one can use the voltage information available from the simulation or the timing information. In the latter case, the output resistance of a transistor (or a more complicated driving stage) is estimated through the quotient of rise time (fall time) and load capacitance.

4.3. Modeling of dispersive effects

For modeling the dispersive region it seems appropriate to do this with an RC ladder network. A cascade of n pi-sections was chosen (see Fig. 4.8(a)). A cascade of T-sections (Fig. 4.8(b)) would be equally possible, but a cascade of stages as shown in Fig. 4.8(c) is not symmetric and this is inherently a disadvantage for modeling a symmetric structure. This is reflected in the number of stages needed for a good accuracy. Fig. 4.9 gives the voltage step response curves for a network with n pi-sections ($n = 1$, 2 and 10) and a concentrated RC stage with $RC = rcl^2$, together with the exact response of a distributed rc line for the case $R_o = C_o = 0$, i.e., the most critical case. It is observed that it is sufficient to set in circuit modeling $n = 1$ or $n = 2$, while for simplicity it is very attractive to set $n = 1$.

Of course, when the switching frequency f becomes comparable to $1/rcl^2$, this time–domain analysis is not valid anymore. It was, however, found by Antinone and Brown [22], that, for $f < 1/2rcl^2$, a T-section approximation with $n = 1$ gives magnitude- and phase-errors (in the frequency domain) of less than 5%. For $f > 1/2rcl^2$, more stages are needed for accurate modeling. In fact, from [22] it

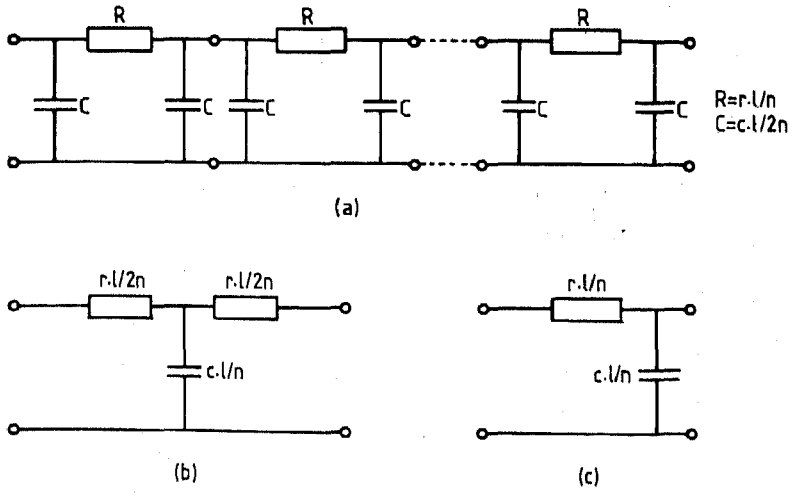


Fig. 4.8. Modeling a dispersive interconnect.

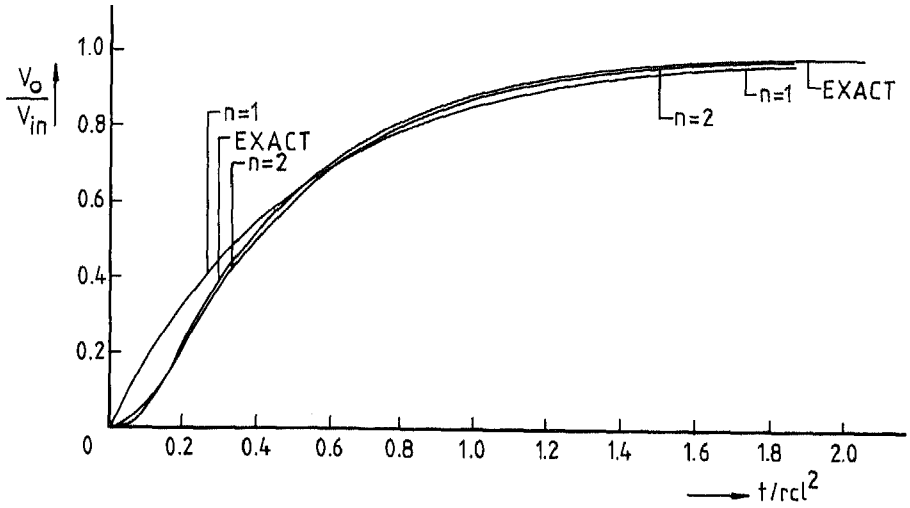


Fig. 4.9. Dispersive line approximations compared to the theoretical solution.

follows that five T-stages will generally yield satisfactory results for switching frequencies as high as ten times $1/rcl^2$. Although Antinone and Brown referred to T-section approximations, these figures will approximately be equal for our pi-section approximation.

4.4. Some remarks

Two remarks have to be made. Firstly, we state that is advantageous to neglect in circuit simulation wire resistance whenever possible. As this is favorable to the number of nodes in the equivalent circuit, this is favorable to the size of the MNA-matrix of our circuit simulation program. (Also, the pi-section approximation is therefore preferred in stead of the T-section approximation.) Neglecting wire resistance is nearly always possible in small subcircuits. This is because wires are short then, and the product of r and l will be small compared to the output impedance of the driving transistor. On the contrary, for these wires it is by far not always true that wire capacitance is much smaller than the load capacitance because the gate capacitance per unit area is only of the order of about ten times the interconnect capacitance per unit area. Since interconnect capacitance might even be doubled due to fringing effects, they must practically always be included in the simulator source file.

Secondly, as a note to the circuit designer, it is mentioned that he/she must be careful when entering the dispersive region (see Fig. 4.6). In the resistive or capacitive region, wire delay is proportional to the length of the wire. This is because either the resistive or the capacitive effect of the wire determines the delay time. Also, as long as wire resistance and wire capacitance are both small compared to R_0 and C_0 respectively, delay is independent of wire length. However, in the dispersive region, delay is proportional to the square of the wire length. This can be seen from Fig. 4.3 as well as from Fig. 4.9. Now, delay time

can easily explode with wire length and the designer must be aware of this fact. A solution might be to adjust the driving transistor aspect ratio or to insert buffer stages in the line (see Mead and Conway [9, p. 23]).

5. Coupling capacitances and contact capacitances

In a typical NMOS process, a wire on each interconnection layer has its own characteristic capacitance to substrate (bottom as well as edge capacitance) as discussed in Sections 2 and 3. In addition, capacitances between different wires are present, too. These capacitances are particularly important when two (or eventually three in case of an implanted undercrossing under poly and metal) wires overlap or cross each other. Even in the case they only come close to each other, coupling capacitance is noticeable.

Firstly, the situation of neighboring wires is discussed in Section 5.1. It will there be made clear that, except for relative small spacing, the net wire capacitance does depend only very little on wire spacing. Therefore, and also because the actual magnitude of the coupling capacitance of neighboring wires is relatively small and very unpredictable due to line width and alignment tolerances, coupling capacitances for neighboring wires are not modeled.

Next, overlap capacitances are discussed in Section 5.2. An approximation method for these capacitances that includes fringing capacitances is proposed. It clearly shows how important these coupling effects can be. Also, it can be used for reconstruction purposes when more exact measurement data are not available.

Finally, capacitances associated with the contacts between different wires are discussed in Section 5.3, since the situation is very much like the situation of Section 5.2. The capacitances are in fact approximated using the results of Section 5.2.

5.1. Neighboring wires

In this section, capacitance between neighboring wires will be discussed. However, metal or poly coupling is different from the coupling between two neighboring diffusion wires, and these two cases will be discussed separately in subsequent subsections.

5.1.1. Neighboring metal or poly wires

Consider three parallel conductors above a conducting ground plane (the substrate), as shown in Fig. 5.1(a). The equivalent circuit is shown in Fig. 5.1(b). In [11], Ruehli and Brennan gave numerical values for these capacitances as a function of the center-to-center conductor spacing s for two sets of wire thickness t and height above substrate h. Also, they gave for each conductor the sum of its capacitances to the other conductors and to substrate. Their results are depicted in Fig. 5.2. In both cases, the dielectric ($\epsilon_r = 4$) is assumed to fill the upper half space. Further, wire width is 5 microns and wire length (not clear in the original paper) is 100 microns.

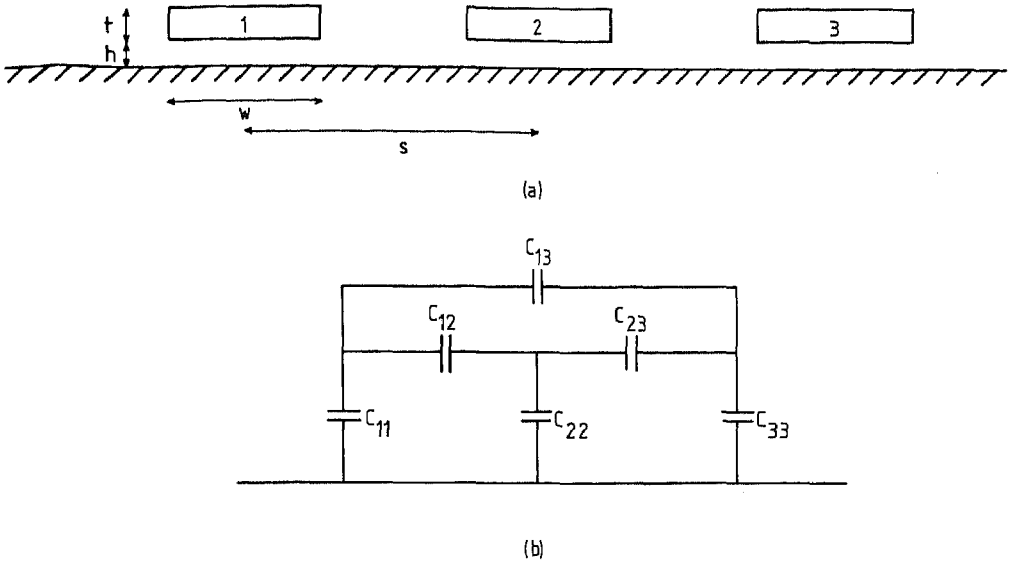


Fig. 5.1. (a) Three parallel wires above a conducting groundplane. (b) Equivalent circuit.

From Fig. 5.2 it can be seen that, even for relative large spacing, some coupling is present. However, except for relative small spacing, the increase of coupling capacitance is approximately balanced by a decrease of substrate capacitance. This can be seen from C_T in case the neighboring wires are floating, while C_S applies in case the neighboring wires are shunted to ground. For example, in [11], $C_{T_{22}}$ defines the so-called total capacitance of conductor 2 and is given by

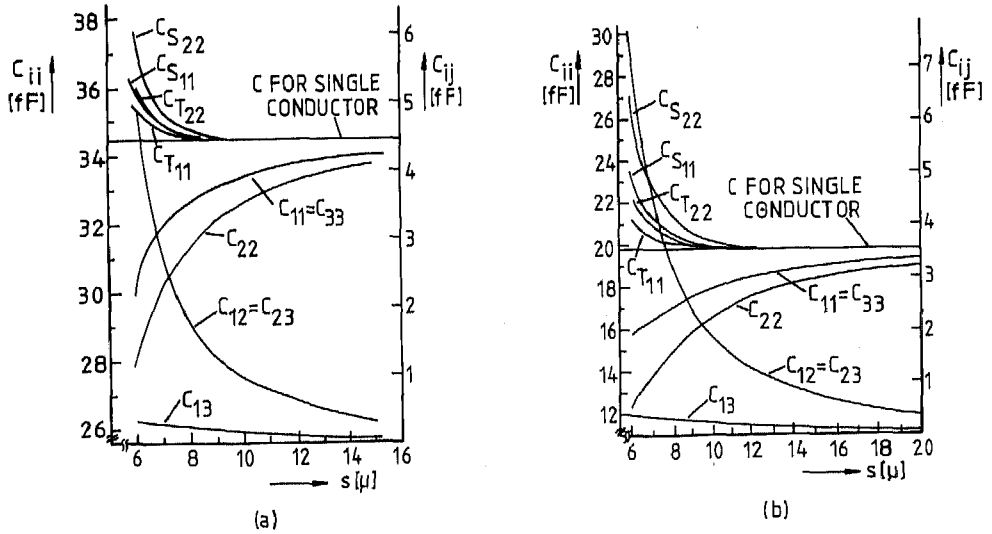


Fig. 5.2. Capacitances vs. conductor spacing after Ruehli and Brennan [11]. (a) $h = 0.8$ μ and $t = 0.5$ μ. (b) $h = 2$ μ and $t = 1$ μ.

(assuming symmetry, so that C_{13} can be neglected):

$$C_{T_{22}} = C_{22} + 2 * C_{23} * C_{33}/(C_{23} + C_{33}).$$

Further, $C_{S_{22}}$, the short-circuit capacitance of conductor 2 is given by

$$C_{S_{22}} = C_{22} + C_{12} + C_{23}$$

and the short circuit capacitance of conductor 1 is given by

$$C_{S_{11}} = C_{11} + C_{12} + C_{13}.$$

Fig. 5.2(b) shows a larger interaction than Fig. 5.2(a) since the ground plane is further away.

However, the following can safely be stated:

(1) Except for very small wire spacing the net capacitance of a wire does virtually not depend on the proximity of other conductors.
(2) The coupling capacitance is only of the order of one tenth or less of the net wire capacitance for integrated circuits made in present technology.

Keeping this in mind, in most cases it is not a large drawback when coupling capacitances between neighboring conductors are not modeled and included in the circuit simulator source file. This simplifies the circuit reconstruction procedure because each conductor can be treated separately as if there are no other conductors nearby. Only the small voltage spikes resulting from the coupling (they are also in the order of one tenth or less of the actual signal transition) are then lost from the simulator output.

Moreover, even the in case we do compute coupling capacitances for neighboring wires and include them in the simulator source file, we can hardly expect a better match of simulator output and actual circuit performance. This is because line width and alignment tolerances of the fabrication process will cause a drastic relative variation of these capacitances. It thus remains a task of the circuit designer still to design the circuit to function properly independent of the exact magnitude of these (relative small) capacitances. An investigation of the upper and lower bounds of capacitances between neighboring wires might, however, be the subject of future work.

5.1.2. Neighboring diffusion wires

A diffusion wire in an NMOS process, being a heavily n-doped region in a relative lightly p-doped substrate (see Section 2), is isolated from the substrate by a depletion region. The thickness of this space charge region increases with increasing reverse voltage applied to the junction. For a discussion of coupling capacitance, consider two parallel wires and assume that the spacing between the wires is large enough (and the voltages small enough) to prevent the depletion layers from touching each other (punch trough). Then, only a small coupling capacitance originating from the top side of the wires exists.

While generally a capacitance can be present between two space charge regions, this is not the case here. This can, for example, be seen from the fact that the 'outside' of all depletion regions is at bulk potential. So, no potential difference

can exist between the boundaries of different space charge regions and, consequently, no capacitance. Thus, since the remaining coupling capacitance originating from the top side of the wires is even smaller than the metal/poly coupling capacitance and equally subjected to processing variations, we will further neglect this capacitance.

5.2. Overlap capacitances

Unlike the situation with neighboring wires, coupling capacitances of overlapping wires will be modeled. This is not only because they are much more significant, but they are more predictable, too. Moreover, the total wire capacitance (especially of the upper wire) increases compared to the case without overlap.

Firstly, we will consider the case of a (narrow) wire in a specific layer, running under or over a very large conductor in another layer. The parallel plate (coupling) capacitances being trivial, the edge capacitances of the narrow wire are estimated. Depending on which layers are under consideration, the edge capacitance of the narrow wire is now a capacitance to the other layer, to the other layer as well as to substrate or remains unchanged as a capacitance to substrate (for example, in case the narrow wire is a diffusion wire). These estimations are done repeatedly for all possible combinations of narrow wire and large conductor, yielding a complete coupling capacitance model for the narrow wire–large conductor case.

Secondly, this model is applied to the case with two (narrow) crossing or overlapping wires, thereby neglecting the effects associated with the crossing of the edges. This is justified following a similar reasoning as in Section 5.1 to neglect coupling capacitances of neighboring wires. The model is compared to known results in Section 5.2.6.

5.2.1. Metal wire above a large diffusion or poly region

Consider a metal wire running above a large diffusion region (Fig. 5.3(a)) or above a large poly region (Fig. 5.3(b)). Now, the capacitance of the metal wire to the underlaying poly or diffusion can obviously again be specified in terms of area and edge capacitance as proposed in Section 3. In both cases, the dielectric is the second field-oxide layer (OX2). Thus the capacitance of a metal wire to a large underlying poly or diffusion region can be approximated by

$$C = A * C_a + P * C_e,$$

where

A = the metal overlap area,

P = the total length of the metal edges running above poly or diffusion.

C_a and C_e follow from the geometry as discussed in Section 3.

5.2.2. Diffusion wire under a large metal region

This situation is depicted in Fig. 5.3(c). Now, it can safely be assumed that the

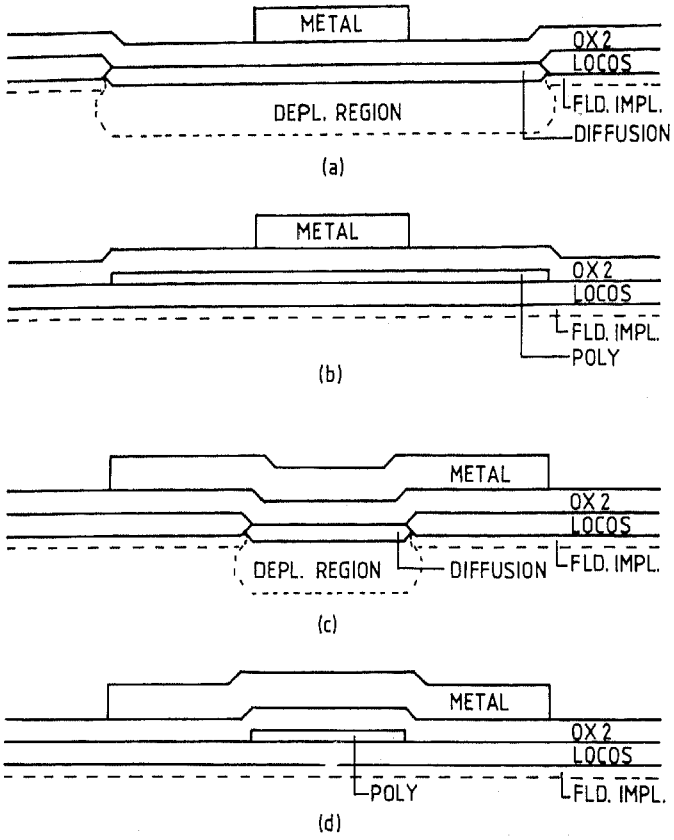


Fig. 5.3. Illustration of overlap capacitances.

field implant region prevents (most of) the diffusion fringing field lines from reaching the covering metal. Therefore, although some fringing effects must be present, the diffusion sidewall capacitance to the covering metal is assumed to be zero, and the diffusion sidewall (and area) capacitance to substrate is considered not to change from the metal covering. The only capacitance of a diffusion wire to a large covering metal region is thus a capacitance proportional to the area of overlap as discussed above.

5.2.3. Poly wire under a large metal region

This situation is depicted in Fig. 5.3(d). While the poly area capacitance to substrate does not change from the metal covering and the poly area capacitance to metal is discussed above, poly fringing capacitance is discussed here. It is obvious that the poly now has fringing capacitance to substrate as well as to metal. Although curve fitting to Chang's [12] capacitance formula for a wire between two ground planes would perhaps also give an accurate approximation formula of capacitance for this case as it did for wire capacitance in Section 3, this is left for future work.

Instead, it seems a reasonable assumption that the total fringing capacitance to substrate as well as to metal is approximately equal to the fringing capacitance without covering. Since the spacings poly-substrate and poly-metal are (approximately) equal, half the fringing capacitance is considered to be capacitance to metal and half is considered to be capacitance to substrate.

5.2.4. Implanted undercrossings

In some NMOS processes there is an additional interconnection layer intended to undercross a poly wire to connect two diffusion wires. It is implemented as a pre-implanted arsenic tunnel under the first field-oxide layer [6]. From the resulting geometry, we can state the following:

(1) The capacitance of a narrow poly (or metal) wire to a large undercrossing region is the same as the capacitance of poly (or metal) to substrate. (The dielectrics are the same.)

(2) The edge capacitance of an implanted undercrossing does not depend on covering poly (or metal) wires. (As was found for diffusion wires.)

5.2.5. Two narrow wires

Above, we have specified a coupling capacitance approximation method for the case that one wire is much wider than the other. However, we will also apply this model to the case of narrow crossing wires or small regions of overlap. Although a three-dimensional approximation method must then be used for high accuracy, a reduced accuracy is justified by the fact that small regions of overlap will result in small coupling capacitances with (generally) a small influence on circuit performance. Also, alignment and line width tolerances will again cause coupling capacitance to vary appreciably. This variation is relative strong for narrow wires or small regions of overlap and this justifies again the statement that there is no need to model small geometries as accurate as larger geometries.

5.2.6. Comparison with numerical results

Consider Fig. 5.4, where a dielectric with $\epsilon_r = 4$ is assumed. For this geometry we will calculate the coupling capacitance C_{12} (denoting the lower conductor as number 1 and the upper conductor as number 2) and compare it with known results. Moreover, from the considerations above, we can find that compared to the single conductor case the capacitances to substrate C_{11} and C_{22} of the lower conductor and the upper conductor respectively, are reduced. We will compute these reductions ΔC_{11} and ΔC_{22} and compare them with known results. In fact, the known results are available from a three-dimensional computer calculation by Ruehli and Brennan [11], and are given by

$$C_{12} = 2.25 \text{ fF}, \qquad \Delta C_{11} = 0.57 \text{ fF}, \qquad \Delta C_{22} = 1.11 \text{ fF}.$$

Now, from the results obtained in Section 3 and using the above proposed approximation for the (coupling) capacitances associated with the 2 wiring levels of Fig. 5.4 (and Fig. 5.2), we can compute Table 5.1.

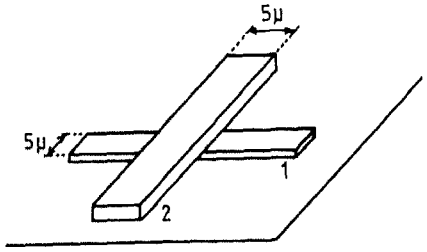


Fig. 5.4. Two crossing wires at $h = 0.8$ μ (with $t = 0.5$ μ) and at $h = 2$ μ (with $t = 1$ μ).

Next, for the geometry of Fig. 5.4 we find, using this table,

$$C_{12} = 2.22 \text{ fF}, \qquad \Delta C_{11} = 0.29 \text{ fF}, \qquad \Delta C_{22} = 0.95 \text{ fF}.$$

It can be seen from comparing these results with the results obtained by Ruehli and Brennan that the approximation can use some improvement. More information is needed then. However, this approximation is a useful improvement compared to simple parallel plate calculations. They would result in

$$C_{12} = 1.23 \text{ fF}, \qquad \Delta C_{11} = 0.00 \text{ fF}, \qquad \Delta C_{22} = 0.43 \text{ fF}.$$

5.3. Capacitances associated with contacts

In this section, capacitances associated with the contacts between wires on different wiring levels will be discussed. Again an approximate way to evaluate these capacitances will do. This is because these capacitances will show relative large variations, because they strongly depend on the alignment of the masks and on the line width tolerances. The proposed approximation, however, becomes more accurate for larger contacts.

In fact, the actual electrical connection of the two wires is achieved by etching a small hole in the dielectric that normally isolates the two wires. Fig. 5.5 gives schematically a cross-sectional view of a metal–polysilicon contact. Similar figures may be drawn for other contacts (metal–diffusion and poly–diffusion).

We can learn from these figures that the center conductor, which forms the actual electrical connection between the wires, has little influence on capacitance.

Table 5.1
Capacitances associated with the wiring levels of Fig. 5.2

	Area capacitances [μF/m²]	Edge capacitances [pF/m]
Layer 1 to substrate	44.3	58.1
Layer 2 to substrate	17.7	50.5
Layer 1 to layer 2	50.6	29.1
Layer 2 to layer 1	50.6	66.7
Layer 1 to substrate with covering layer 2	44.3	29.1

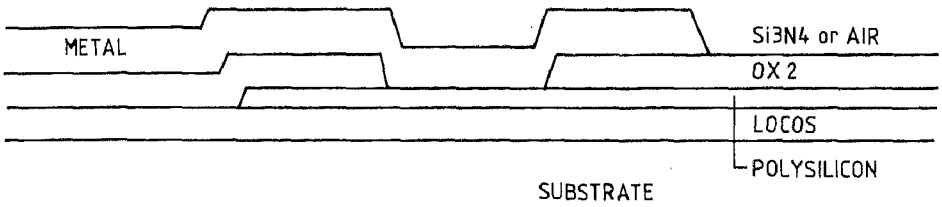


Fig. 5.5. Illustration of a metal–polysilicon contact.

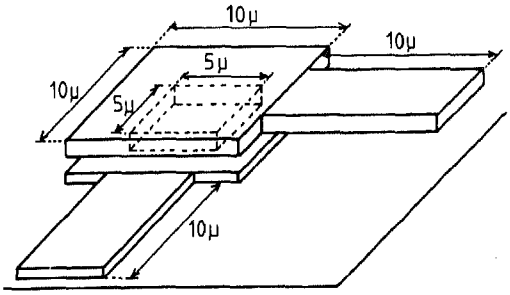


Fig. 5.6. Geometry of a contact between the wiring levels of Fig. 5.2 [11].

This is because it is shielded by the two contact surfaces. This conclusion had already been made by Ruehli and Brennan [11] after numerical computations.

Therefore, the same rules that were given for approximating the capacitances associated with overlapping conductors can be used. Capacitances of both layers to substrate must be calculated as if these wires are only overlapping and not connected. Then, the sum of these capacitances to substrate is taken to be the capacitance to substrate of the actual configuration of connected wires.

5.3.1. Comparison with numerical results

Again, we will compare this approximation with numerical data available from Ruehli and Brennan [11]. For a contact between two wires on the wiring levels of Fig. 5.4 and as shown in Fig. 5.6 they found the capacitance to be given by

$$C = 12.44 \text{ fF}.$$

The above proposed approximation, using the capacitance values of Table 5.1 leads to

$$C = 12.78 \text{ fF}.$$

This is also a considerable improvement compared to parallel plate calculations. They would result in

$$C = 7.53 \text{ fF}.$$

6. Transistor dimensions

In a circuit simulation program as SPICE there are two groups of parameters describing a transistor. (Or describing a capacitor implemented as a transistor

with source and drain connected together. This section also applies to such capacitors.)

The first group of parameters are the model parameters. They are related to the type of the transistor (enhancement, depletion, intrinsic etc.) and to the process. They do not play any role in circuit reconstruction, except that they must be chosen according to the type of transistor that is recognized in the layout. These parameters have to be specified only once in the simulator source file. Examples are gate oxide thickness, zero bias treshold voltage and bulk treshold parameter. Also, the lateral diffusion parameter of source and drain is generally considered as a model parameter, although a small dependency on the size of the diffusion window (source and drain region) is present.

The second group are the device parameters. They depend on the geometry (they describe the geometry) of each particular device. Unlike the model parameters, these parameters have to be specified for every transistor. In particular, there are the dimensions of source, drain and gate. These dimensions must be extracted from the layout. As Section 2 points out that source and drain diffusion regions are more conveniently considered as interconnections and shows how to model them, the dimensions of the active region (channel) will be discussed in this section. Special attention is paid to the aspect ratio of non-rectangular devices.

6.1. Rectangular devices

For a rectangular device (see Fig. 6.1) determination of transistor length and width is simple. However, as mentioned in Section 1, they are not equal to their layout. In stead, poly and diffusion line width reduction cause transistor length and width reduction, respectively (see Fig. 6.2(a), (b)). In this context, the lateral expansion of local oxidation that causes diffusion wire narrowing and thus reduces transistor width, is usually termed as a 'birds-beak' [23,24] (see Fig. 6.2(a)). Ideally, these corrections to transistor dimensions have already been performed with the correction to the width of poly and diffusion that must preferably be done before all other circuit reconstruction steps, as discussed in Section 1.

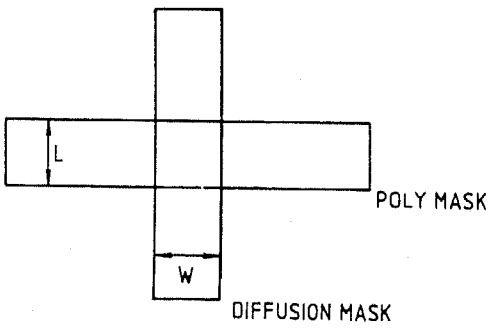


Fig. 6.1. Self-aligned MOS transistor.

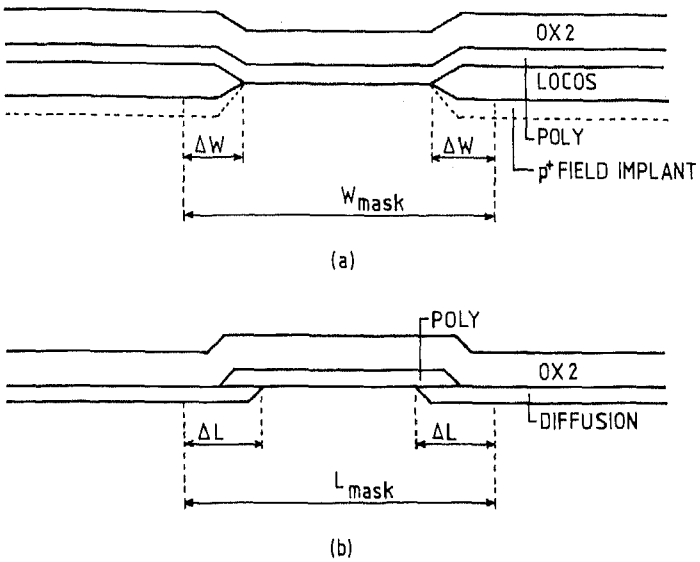


Fig. 6.2. (a) Channel narrowing due to 'birds beak'. (b) Channel shortening due to poly line width reduction and lateral diffusion.

However, in Fig. 6.2(b) there is shown another effect causing transistor length reduction. This effect is the lateral diffusion of source and drain under the gate and must separately be corrected for. Now, as mentioned above, this correction is (to first order) not layout-dependent. A model parameter can be specified in SPICE to correct dimensions of all transistors for lateral diffusion effects.

6.2. Non-rectangular devices

Less obvious are the width and length dimensions of a transistor with a different geometry as rectangular. For example, to obtain a compact layout, a 'long' transistor will very often be folded as shown in Fig. 6.3(b). A 'wide' transistor as encountered in driver and buffer circuits, often has the meander form as drawn in Fig. 6.3(a). To calculate the width and length of such transistors, the effects of the bends must be known. These effects are certainly not negligible.

To investigate these effects, a three-dimensional device analysis program could

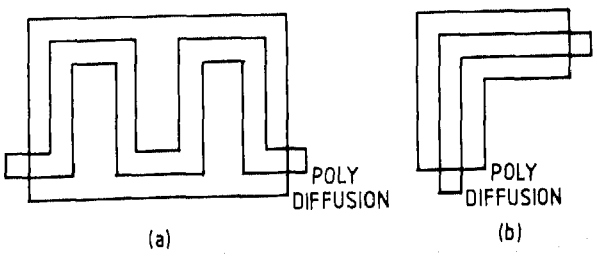


Fig. 6.3. Common non-rectangular transistor geometries.

possibly be used. Such a program, however, must then be available but its complexity would anyway form a large drawback. Moreover, in most cases these programs do not output external quantities as source and drain currents at all. Also, investigation of these effects using test chips is not very attractive. It is a very expensive and time-consuming alternative. Fortunately, it seems that a much simpler method can also provide the needed information. This method is based upon the assumption that the transistor model parameters do not change as a result of the bend of the channel region. Thus, only an appropriate effective channel width and length must be found. But, since the product of the effective width and length must be equal to the area of the active region for a correct value of the gate channel capacitance, an effective L/W ratio provides already enough information.

6.2.1. L/W analogen method

Consider an arbitrary shaped transistor. For example, the transistor may have a geometry as shown in Fig. 6.3(a) or (b) but more complex geometries are also allowed. Let the transistor be biased in the linear region ($V_{GS} \gg V_T$ and $V_{DS} \ll (V_{GS} - V_T)$ for an NMOS device). Now a surface inversion layer exists, forming the conductive channel between source and drain [7]. With very small drain source voltage the mobile charge forming the channel can be considered to be distributed uniformly under the gate. Due to the finite mobility of this mobile charge, the active region now behaves as a two-dimensional uniform distributed sheet resistance. (More exactly, while μ is the mobility or velocity per unit electric field and q is the density of the mobile charge in Coulombs per unit area, the sheet resistance is given by $R_s = 1/\mu q$.)

Accordingly, the effective L/W ratio of a non-rectangular transistor is equal to the L/W ratio of a uniform distributed sheet resistance with the same shape and good conducting connections at places corresponding to the transistor source and

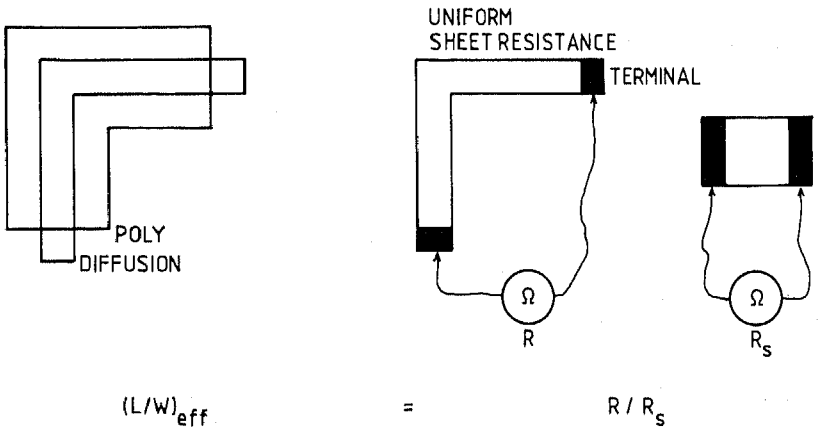


Fig. 6.4. Illustration of the proposed analogen method to determine the aspect ratio of non-rectangular FET's.

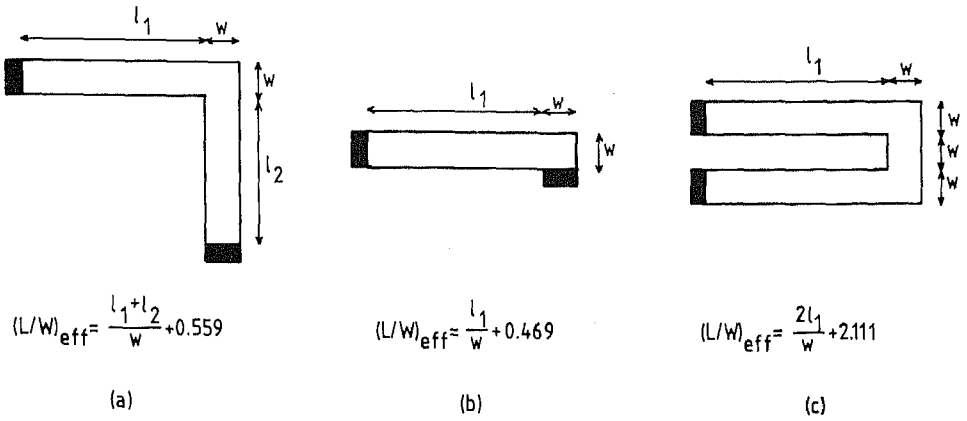


Fig. 6.5. Effective aspect ratios according to Hall [25].

drain terminals. (Geometry as well as resistance may be scaled.) The L/W ratio of the sheet resistance is here defined as the ratio R/R_s. Fig. 6.4 gives an illustration of the followed train of thought.

Now, the ratio R/R_s may be determined experimentally in a way similar to Fig. 6.4, but it may also be determined using a suitable numerical computer program. Such programs are much simpler than the programs previously mentioned. They can be run on much smaller computers and require less computing time still. It might even be possible to implement such a program as a subprogram in an automatic circuit reconstructor. Also, conformal mapping techniques can form a very powerful tool. These techniques were employed by Hall [25] in relation to thin film resistor geometries.

However, for manual reconstruction as well as circuit layout design it might be useful to make a list of effective L/W ratios for some common non-rectangular geometries. For example, Fig. 6.5 gives some results according to Hall [25]. For 'wide' transistors, similar figures might be given. It is then, however, more convenient to speak in terms of conduction and W/L ratios in stead of resistance and L/W ratios.

Further, it might be estimated from Fig. 6.5 as a useful rule of thumb, that a corner square in such non-rectangular geometries has an effective contribution of approximately 0.5 to the aspect ratio.

As an example, we have for the geometry of Fig. 6.5(a) with $l_1 = 3w$ and $l_2 = 2w$, an effective L/W ratio of 5.559. The area equals 6, and we can compute L_{eff} and W_{eff} to be input in SPICE according to

$$W_{\text{eff}} = \sqrt{(L_{\text{eff}} * W_{\text{eff}})/(L/W)_{\text{eff}}} = 1.039,$$

$$L_{\text{eff}} = \sqrt{(L_{\text{eff}} * W_{\text{eff}}) * (L/W)_{\text{eff}}} = 5.775.$$

6.2.2. *Some remarks*

It is noted now that a conflict may arise concerning the gate–source and gate–drain overlap capacitances. These are usually included in the transistor

model as capacitances per unit length and are multiplied by the device width to produce the appropriate overlap capacitance. However, for the example above an overestimation of 4% would result. Much larger deviations may also occur for some (unusual) geometries. When this is not allowed, the overlap capacitances must be specified explicitly. Eventually, for non-rectangular transistors only, they must then be set equal to zero in the transistor model.

As mentioned previously, this method is derived under the assumption that the transistor model parameters do not change. But, it is obvious that for non-rectangular geometries small-size effects may also play a pronounced role and they will probably also need separate attention. This is, however, omitted here since it would be only a correction to a correction.

Finally, it is noted that a computer program to determine R/R_s for non-rectangular geometries may also be used to calculate interconnection resistances, although it is doubtful whether such accuracy is needed. This accuracy must be seen in view of Section 1.

References

[1] Nagel, L. and D. Pederson, Simulation program with integrated circuit emphasis (SPICE), *16th Midwest Symp. on Circuit Theory*, Waterloo, Ontario, 1973.

[2] Akino, T., M. Shimode, Y. Kurashige and T. Negishi, Circuit simulation and timing verification based on MOS-LSI mask information, *Proc. 16th Design Automation Conf.*, San Diego, CA, 1979.

[3] Tucker, M.G. and W.J. Haydamack, VLSI design and artwork verification, *Hewlett Packard J.* **32** (6) (1981) 25–29.

[4] Domenik, S., Parasitic load and device parameter calculations, *Lambda* **2** (2) (1981) 68–69.

[5] Cohen, P.B., Layout considerations in predicting VLSI performance, *VLSI Design* **4** (1) (1983) 64–65.

[6] Benschop, N.F. and L.C.M. Pfennings, Compact NMOS array multipliers with inverting full adders, *Philips J. Res.* **36** (3) (1981) 173–194.

[7] Grove, A.A., *Physics and Technology of Semiconductor Devices* (Wiley, New York, 1967).

[8] Alexander, D.R., R.J. Antinone and G.W. Brown, SPICE2 MOS modeling handbook, Rept. BDM/A-77-071-TR, BDM Corporation, Albuquerque, NM, 1977.

[9] Mead, C. and L. Conway, *Introduction to VLSI Systems* (Addison–Wesley, Reading, MA, 1980).

[10] Ruehli, A.E. and P.A. Brennan, Accurate metallization capacitances for integrated circuits and packages, *IEEE J. Solid-State Circuits* **SC-8** (4) (1973) 289–290.

[11] Ruehli, A.E. and P.A. Brennan, Capacitance models for integrated circuit metallization wires, *IEEE J. Solid-State Circuits* **SC-10** (6) (1975) 530–536.

[12] Chang, W.H., Analytical IC metal-line capacitance formulas, *IEEE Trans. Microwave Theory Tech.* **MTT-24** (9) (1976) 608–611; **MTT-25** (8) (1977) 712.

[13] Dang, R.L.M. and N. Shigyo, Coupling capacitances for two-dimensional wires, *IEEE Electron. Device Letters* **EDL-2** (8) (1981) 196–197.

[14] Dierking, W.H. and J.D. Bastian, VLSI parasitic capacitance determination by flux tubes, *IEEE Circuits and Systems Magazine* **CSM-4** (1) (1982) 11–18.

[15] Elmasry, M.I., Capacitance calculations in MOSFET VLSI, *IEEE Electron. Device Lett.* **EDL-3** (1) (1982) 6–7.

[16] Sakurai, T. and K. Tamaru, Simple formulas for two- and three-dimensional capacitances, *IEEE Trans. Electron. Devices* **ED-30** (2) (1983) 183–185.

[17] Yuan, C.P. and T.N. Trick, A simple formula for the estimation of the capacitance of two-dimensional interconnects in VLSI circuits, *IEEE Electron Device Lett.* **EDL-3** (12) (1982) 391–393.

[18] Sinha, A.K., J.A. Cooper, Jr. and H.J. Levinstein, Speed limitations due to interconnect time constants in VLSI integrated circuits, *IEEE Electron Device Lett.* **EDL-3** (4) (1982) 90–92.

[19] Bilardi, G., M. Pracci and F.P. Preparata, A critique of network speed in VLSI models of computation, *IEEE J. Solid-State Circuits* **SC-17** (4) (1982) 696–702.

[20] Boyce, W.E. and R.C. Diprima, *Elementary Differential Equations and Boundary-Value Problems* (Wiley, New York, 1977).

[21] Knepper, R.W., Dynamic depletion mode, an *E/D* MOSFET circuit method for improved performance, *IEEE J. Solid State Circuits* **SC-13** (5) (1978) 542–548.

[22] Antinone, R.J. and G.W. Brown, The modeling of resistive interconnects for integrated circuits, *IEEE J. Solid-State Circuits* **SC-18** (2) (1983) 200–203.

[23] Chiu, K.Y., J.L. Moll and J. Manoliu, A birds beak free local oxidation technology feasible for VLSI circuits fabrication, *IEEE J. Solid-State Circuits* **SC-17** (2) (1982).

[24] Kooi, E., J.G. van Lierop and J.A. Appels, Formation of silicon nitride at a Si-SiO2 interface during local oxidation of silicon and during heat-treatment of oxidized silicon in NH3 gas, *J. Electrochemical Society* **123** (7) (1976) 1117–1120.

[25] Hall, P.M., Resistance calculations for thin film patterns, *Thin Solid Films* **1** (1967) 277–295.