Elektronische Signaalbewerking (deel 2)

ET2405-d2

Lecture Notes

A. van Staveren and W.A. Serdijn

February 2, 2009

with corrections in line with the kind suggestions made by Evert de Haan, EE B.Sc. class of 2006

Contents

Introduction

Nowadays, electronics can be found in mostly any application. The main function for these electronics is *signal processing*, i.e. amplification, filtering, ...etc.

Taking as an example the filter that can be found preceding an analog-to-digital converter, its function is to suppress signals outside the nyquist band (half the sampling frequency) to prevent the converter from folding those relatively high-frequency signals into the base band. An other example is a filter that is used in a class-F power amplifier for mobile applications. Class-F power amplifiers are based on switching a power source on and off. The rate of switching is determined by the input signal. In combination with a resonant circuit (filter) the block-wave like signal is changed into a sine-wave like signal. This to prevent a lot of additional harmonics to be transmitted.

The task for an electronic designer is to design for a given set of specification, which are set by the application, the electronic solution that is able to perform the required signal processing function. Electronic design has been for a long time a kind of *art of electronics*. One needed to have a mental catalogue of a lot of circuit diagrams and their basic performance, such that in the case of a design task, one could select the known circuit diagram that was closest to the solution. In order to make it fulfill the specifications completely, it often needed to be tuned (or tweaked). Luckily, more systematic design approaches are developed which do not rely on this kind of expertise knowledge (heuristics). In contrast, those design methodologies rely on a synthesis type of approach. For this, one must know the behavior of the basic building blocks (electronic components) and the procedure for synthesizing the specific function. No a-priori knowledge of complete circuit solutions is required.

In these lecture notes a design methodology is presented to synthesize electronic filters. The intention of the methodology is that for a given filter function, expressed as *differential equation*, one is able to synthesize (design) an electronic implementation that can perform that function. The methodology is hierarchical, i.e. starting with ideal building blocks and subsequently implementing them gradually. The first three chapters pay attention to the block diagram level, whereas the last three chapters pay attention to the actual implementation at transistor level.

Chapter 1 starts with describing the history of filter design. As many types of filters exist, subsequently, a classification is developed that puts all these filter types in an overall structure. It enables the designer to select the correct class of filter for his/her application. These lectures notes focus on the design of *analog active time-continuous filters*. This chapter describes the first step of the design methodology: transforming the differential equation describing the required filter function into a block diagram comprising ideal basic building blocks (integrators and scalers).

Chapter 2 addresses the *quality* of the filter implementation. The quality of a filter is mainly determined by its *dynamic range* (DR). This chapter describes how the filter topology relates to the DR and how the DR can me maximized. For this, the noise and distortion of filters is treated.

Chapter 3 focusses on the effect *parameter spread* and the choice of a certain level of *model complexity* has on the *accuracy* of filters. Design solutions are treated to be able to design accurate filters with relative inaccurate components.

Chapter 4 continues with the implementation of the filter. In chapter 1 it is concluded that one of the basic building blocks is the integrator. In this chapter the design of basic integrator implementations is treated. For this the small-signal models of the transistors (amplifying stages), that are to be used to design the active part (opamp) of the integrator, are treated. These small-signal models are required to analyze the performance of a designed active part.

Chapter 5, subsequently, describes how an active part can be synthesized by using transistors in such a way that maximum quality is obtained.

Chapter 6 describes how the performance of a feedback loop comprising an active part can be analyzed. In this way the performance of a designed integrator can be determined.

Chapter 1

From differential equation to block diagram

1.1 Introduction

The first filters ever made consisted of coils, capacitors and resistors. However, coils cannot be applied in integrated filters, thus the "conventional" design of continuous time filters does not fully cover the necessary design theory, especially not for active inductorless filters. Still, as a starting point of the design of active integrated filters we give an historical introduction (section 1.2 in order to highlight the development of 20th century filter design, because design theory developed for passive filters can be partially used for active filters.

For nowadays high-performance of electronics, designing filters without a well-defined design trajectory is not easy. A lot of design aspects may pose additional conditions on the design and when it is not clear at what level of hierarchy this should be taken into account, it becomes difficult to end up with an optimal filter. Therefore, section 1.3 shows in short a possible design trajectory for filter design.

One of the key issues in filter design is to obtain the required dynamic range, i.e. what minimum and maximum signal the filter can handle at the same time. The dynamic range is therefore an important *quality* aspect for filters. Section 1.4 shows how dynamic range relates to signal-processing theory by which it can be shown that it indeed is a fundamental quality aspect.

From the historical overview but also when scanning the modern literature on filter design, one probably notices that a huge amount of different filter implementations and structures are used. To get insight in this field, section 1.5 describes a classification of filter implementations. This helps the designer in structuring the design from specification to realization, i.e. what would be the optimal filter type/implementation for my application.

Section 1.6 shows that the state-space description is well suited to describe

the filter to be designed. The description should comprise on the one hand the information about the filter transfer and on the other hand the information about the topology. Topology information is important, as will be seen in next chapters, because of the dependency of the dynamic range on the topology for a given transfer.

Finally, section 1.7 shows how from a transfer function a block diagram can be derived for a filter.

1.2 History of filter design

Around 1890, several people were involved in improving the quality of transmission lines by adding coils. Only in 1899, did M.I. Pupin succeeded in improving the attenuation characteristic of telephone and telegraph wires by inserting coils. His success resulted in a world-wide use of "Pupin Lines", see figure 1.1. The

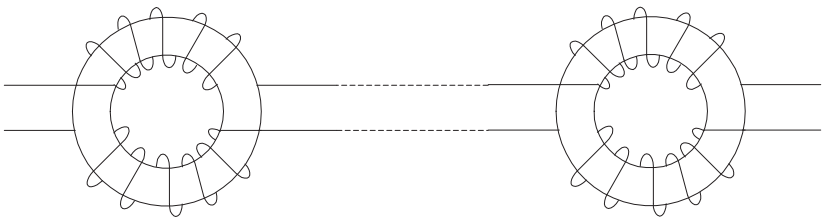

Figure 1.1: Adding coils in a transmission line to improve the performance.

.

behavior of these lines was more elaborately researched by G.A. Campbell, who, in 1903, published an article which described the frequency behavior of the lines. He invented the low-pass characteristic of the cable. He also realized the use of the cable as a band-pass filter, by replacing the coils by a combination of coils and capacitors. The problem of making filters with bulky cables led Campbell and K.W. Wagner to the simulation of the cable by a ladder construction of impedances. This was also indirectly suggested by Pupin, and the resulting filter was called the "electrical wave filter". The year 1915 may be considered the day of birth of the first electrical filter.

Design methods were invented by many people, amongst others by O.J. Zobel. The design method he developed was the beginning of the transmission line theory, that spoke in terms of characteristic impedance and wave propagation to describe the attenuation of the filter. He introduced a method to design filters with an infinite number of coils and capacitors. More filter theory was developed by S. Darlington and S. Butterworth. Butterworth made fourth-order filter sections, that were intercoupled by amplifiers (realized by tubes). Thus he was the

first person to design active filters. Also from his hand are the well-known Maximally Flat Magnitude (MFM) attenuation characteristics (1930). Around the same time, W.R. Bennett solved the problem of realizing passive maximally flat transfer functions for filters of any order. W. Cauer also designed passive filters, but he used Tchebysheff approximations to describe the transfer function. Between 1930 and 1940, Cauer published several articles on the design of filters with some desired attenuation curve. In 1939, Darlington published an article in which he used Tchebysheff approximations to design transfer functions. The impact of Darlington's and Cauer's work was great, although the computing power in those days was too small to make full use of the theory.

Current monolithic technology does not allow the use of coils. The drawback of the generation of only poles on the real axis when making filters with only resistors and capacitors can be circumvented by using active components. Sallen and Key delivered a general design method to construct active R-C filters. It was based on cascading second-order stages. This method was not very popular in those days, because of the use of tubes. The emerging silicon technology, though, made it very attractive.

In 1977, the first switched-capacitor filter was applied. These filters still use a continuous signal amplitude, but process the signal at discrete time events. In 1979, Tan and Gray found solutions to tuning filters by placing automatic tuning circuits on chip. Tunability was realized by applying JFETs. Further important research was carried out by Moulding, Voorman, Tsividis, Nauta and Groenewold (tuning method, use of Gilbert Gain Cell for constructing integrators, MOSFET-C filters, high-frequency filters and dynamic range optimization, respectively).

In the nineties, dynamic translinear circuits and filters were introduced. These filters, making efficient use of the exponential behavior of bipolar transistors, were a generalization of the first translinear filter introduced by adams in 1979. The dynamic translinear filters are based on the principle depicted in figure 1.2. For this

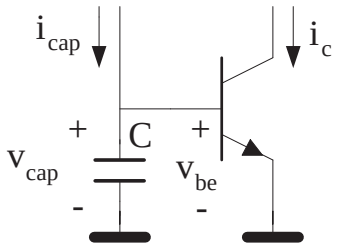

Figure 1.2: The basic circuit in a dynamic translinear filter.

.

circuit the relation between the capacitor current (i_{cap}) and the collector current

(i_c) is given by:

$$CV_T \frac{di_c}{dt} = i_c i_{cap} \tag{1.1}$$

in which V_T is the thermal voltage (kT/q).

1.3 Overview of filter design trajectory

Filter design consists of several steps. The trajectory from specification to circuit is here discussed. Some of the steps will only be mentioned, because they are treated more thoroughly in following sections, others are merely mentioned to give a complete overview, but are actually beyond the scope of this book.

1.3.1 Application

Filters are used in many applications. Roughly speaking the applications can be divided into two groups:

- selection;
- shaping.

The main difference between selection and shaping is that selection acts on wanted and unwanted signals in the frequency domain, whereas shaping is used for the wanted signals only.

Selection by means of a filter is depicted in figure 1.3. A well-known example

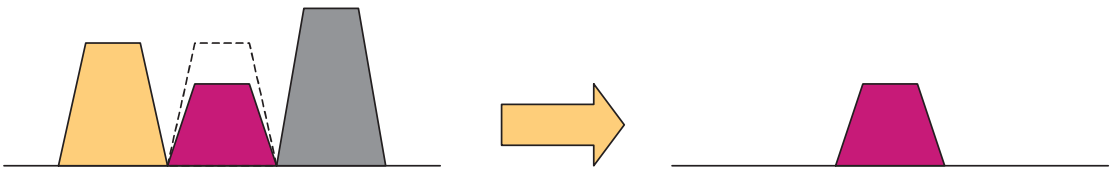

Figure 1.3: A filter used for selection.

of a filter used for selection is the radio receiver in its basic function. From a large number of radio stations broadcasting at different channels the radio receivers selects one of them. Roughly speaking the filter selects which part of the frequency spectrum is passed and which part is rejected. Passing a channel should be done with no losses, ideally, and rejection should be done such that 100% of the power is dissipated or reflected by the filter (ideally).

Shaping by means of a filter is depicted in figure 1.4. Shaping is used when the power level in the spectrum of interest is changed in order to adapt the signal

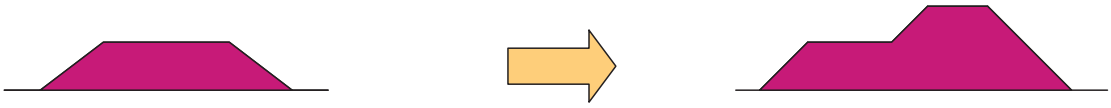

Figure 1.4: A filter used for shaping.

for further processing. An example of shaping can be found in a tape deck. When retrieving signals from a tape, the signal will be contaminated by noise. This noise is, besides due to the electronics, due to the principal of magnetic recording itself. This noise of the tape is mainly found in the higher part of the frequency spectrum. Directly recording the signal on the tape would deteriorate the signal for the higher frequencies. By first amplifying the higher frequencies of the signal to be recorded, the relative effect of the high frequency noise of the tape is reduced.

1.3.2 Specification

For filters in general specification can be derived for from the application. The specification of the filter are several, for example: attenuation curve, dynamic range requirements, power consumption and supply voltage. The highest level specification must be handled first, which is the attenuation curve. In many applications, this attenuation behavior is given as depicted in fig.1.5. A filter curve

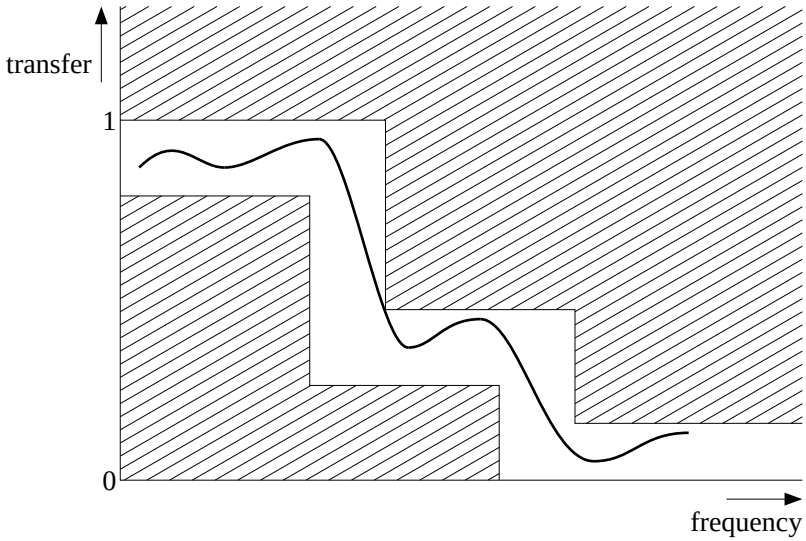

Figure 1.5: Example of attenuation curve

must be designed such that the required attenuation curve is reached. It is usually attractive to use "standard" filter transfer functions. Well-known types are: Butterworth, Chebyshev and elliptic or Cauer. The design of these filter types is

usually based on a frequency normalized low-pass filter curve. Transformations are applied to the low-pass curve to turn it into, for example, a bandpass filter.

1.3.3 Mapping onto topologies

When the filter polynomial is known, or the pole and zero positions, which is actually the same, a topology should be found to map these information on. A topology is an idealization of a filter. It consists of ideal integrators (branches valued $1/s$) and an interconnection circuit. There are many ways to realize a filter function. Some possible topologies are shown in one of the following sections. The choice of a topology is mostly based on experience of "good behavior" in stead of on clear and objective methods. There are some topologies that are known for their low sensitivity and good dynamic range behavior.

1.3.4 Implementation

The steps discussed so far only resulted in abstract filter structures. The structures consist of ideal integrators and interconnection circuitry. They are still implementation independent. The last step is to find an implementation of the integrators, in bipolar, MOS or BiMOS technology.

1.4 Quality versus Costs

The design of a filter can be visualized, on a high level of abstraction, as depicted in figure 1.6. From the specification the filter function can be derived which the

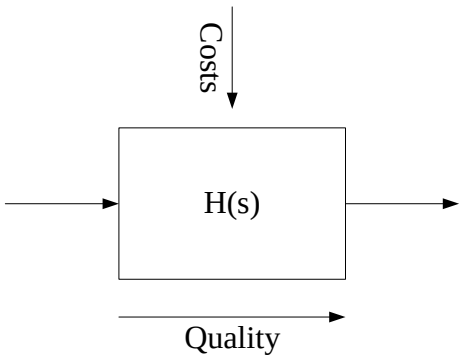

Figure 1.6: The design of a filter on a high level of abstraction.

circuit should implement in the end. At specification level the filter function is still a mathematical expression, i.e. a differential equation. Of course, this mathematical expression is ideal: it does not consume power, it has not a limited

bandwidth or noise production. Of course, the final design which is a real physical implementation will consume power and generate noise and will have a limited bandwidth. Thus as soon as a physical implementation is the goal one should consider how well the circuit implements the differential equation, i.e. the quality.

To reach a certain quality, a certain price has to be paid, i.e. power consumption, chip area et cetera. For a designer a key issue is to have insight in the relation between cost and quality. When this relation is known fast evaluation can be done what the cost will be for a certain quality or the other way around when for instance the power budget is limited, what quality can be reached for that power budget.

Technical merits (quality) of an electronic circuit are valued from the way it performs its function. Of course there are more factors involved than just the technical ones, like the costs to produce the circuit and so on. When only the pure technical merits are evaluated, it is found what can be ultimately achieved; thus when all freedom is given to the designer. When due to other constraints some options are not allowed, it can easily be found how much the decrease of performance (quality) will be. It depends on the circumstances whether a manufacturer is willing to pay the price for extra performance or not, but it is at least known how much improvement can be expected and what price has to be paid to get it. In some cases it can save the manufacturer from paying a price for finding an improvement that is fundamentally not feasible. For example migrating to a more complex and extended technology for getting a higher performance is only useful when the extra options play a role in the optimization of the particular circuit.

1.4.1 Fundamental specifications

There are many ways to specify the performance of electronic circuits, but there are only three fundamental aspects of the performance:

- Noise (N)
- Signal Power (S)
- Bandwidth (B)

related to each other via Shannons equation:

$$C = B\ ^2 \log \frac{S+N}{N} \tag{1.2}$$

C is the signal-handling capacity of the circuit. It is a measure for the information the circuit can process per second. The ultimate goal of a designer is to maximize the signal-handling capacity within the constraints given by the environment. It

can be seen from equation (1.2) that the bandwidth is linear in the expression and the signal to noise ratio is in the logarithm, so an increase of the bandwidth yields more improvement than a comparable increase of the signal to noise ratio. Circuits that "are good at bandwidth" together with signals that are properly coded—so they have little dynamic range and much bandwidth—are therefore likely to be favored when a large signal-handling capacity is required. Digital circuits and signals perfectly match to these requirements and they are indeed by far the "dominant species" in the signal processing world.

However, in areas where bandwidth is a problem, like it is under low-power conditions or at very high frequencies, or when there is no freedom to code the information properly the signal-handling capacity has to be optimized via the signal to noise ratio. In this area high performance analog circuits have to do the job.

1.4.2 Additional Specifications

The three fundamental aspects described above are *sufficient* to specify the performance of a circuit, but they are not the only specifications that are given in practice. Temperature operating range, EMC specifications and so on are of course also of great importance. These specifications can however be seen as a different appearance of the three fundamental criteria or as additional physical criteria. For instance, the disturbances introduced by EM interference (EMI) are comprised within the Noise (N) of equation (1.2), whereas supply voltage is a physical criterion and is not a quality criterion in the sense of Shannon.

These additional physical criteria limit the freedom in the design and thus limit the performance. In practice it is important to know the effect of a physical limitation on the performance. Then it can be evaluated if it is desired or justifiable to, for example, change a supply voltage or increase a supply current in order to improve the performance of a circuit. It can become clear what the influence of the supply voltage on the performance of a circuit is and what voltage becomes "critical", and marks the limit beyond which degradation becomes dramatic. This approach gives much more insight in the behavior of circuits than for example the blink design of 1V circuits because it is a fashion today.

1.4.3 Costs

To realize the performance indicated by the complete set of specifications, a certain price needs to be paid, i.e. the costs. Examples are: power consumption, chip area et cetera. It is usually not too complicated to show for example the relation between power consumption and noise behavior. From this it can be found what the power costs are for optimum noise behavior or what the noise behavior will be

at a given power consumption.

1.5 Classification of filter implementations

Classification is a powerful method for ordering information. It can be applied to all kinds of engineering problems. It can help the designer to find suitable solutions for his/her design problem in a fast way. A classification presents the existing knowledge in a structured way. It differs from an encyclopedia in the way the relation between the several classes is presented. An encyclopedia presents the solutions sequential, whereas a classification also gives hierarchy.

Classifying existing knowledge implicitly shows the location where a lack of knowledge is, i.e. empty classes. For instance, we need to design a signal processing function: $y = f(x)$. Assume that for coding the signals two methods exist, M_1 and M_2. Further, assume that also for the implementation step two different possibilities exist, P_1 and P_2. Then four different solutions can be found, see figure 1.7. When the classification is obtained, then, for instance, existing solutions from

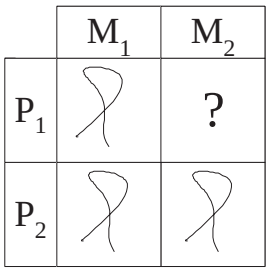

Figure 1.7: An example of finding new solutions via classification

literature can be placed in the classification. The "solutions" which remain empty afterwards, (P_1M_2, in figure 1.7) indicate a *new* solution. In contrast, if an existing solution does not fit in the classification, then the classification in not complete.

In this section classification is applied to filter implementations in order to structure existing filter implementations which may be helpful in selecting what filter implementation to use to obtain for a given application acceptable performance for the lowest price.

Classification starts with a precise definition of what to classify. This definition strongly determines the structure of the classification. Subsequently, the criteria used in the classification need to be chosen.

When classifying filter implementations, criteria can be:

- time behavior (continuous (C) or discrete (D));
- amplitude behavior (continuous (C) or discrete(D));

- type of elements (active (A) or passive (P));
- type of realization (discrete (D) or integrated (Dutch: geintegreerd (G))).

When using these criteria, the classification of figure 1.8 is obtained. From this

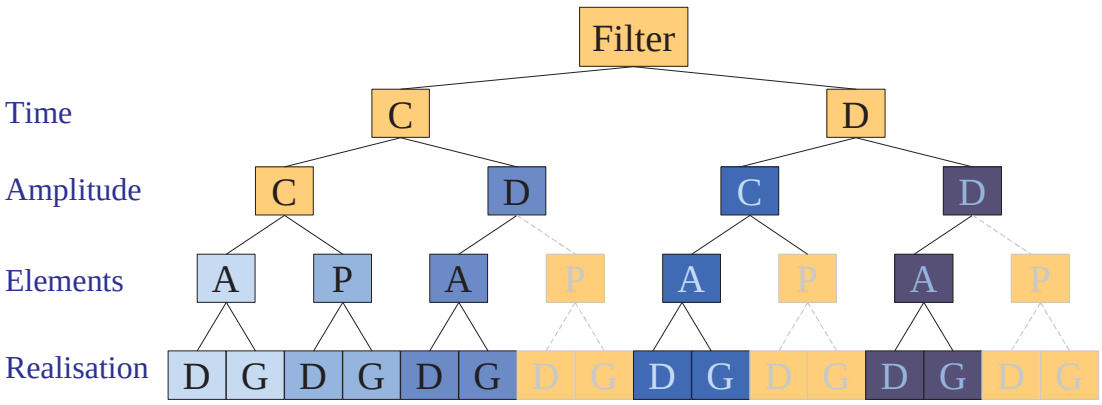

Figure 1.8: Classification of filter implementations

classification three branches can be ignored:

- CDPX
- DCPX
- DDPX

All these options combine passive integration and a time and / or amplitude discrete behavior. As for discrete time or discrete amplitude active devices are required, i.e. switches for discretizing, these options are not possible. From this classification five main types of filter implementations can be distinguished:

- CCAX : analog active time-continuous filters;
- CCPX : passive time-continuous filters;
- CDAX : asynchronous digital filters;
- DCAX : sampled-data filters;
- DDAX : synchronous digital filters.

1.5.1 Analog active time-continuous filters

For these filters all the possible signal values at all time instants are important, i.e. analog. Active elements are used to implement filters without inductors. Passive realization of the same filter transfer would need capacitors and inductors. Analog active filters are often integrated on chip and used in consumer electronics. Because of the active elements these filters can be tuned electronically which is favorable for many applications. Of course, due to the active components in the filter, distortion is introduced, i.e. the output signal of those components is limited by the power supply. On top of that, also additional noise is introduced. Thus compared, for instance, with the passive filters, these filters have less dynamic range.

1.5.2 Passive time-continuous filters

Passive time-continuous filters are realized with passive elements only, i.e. R,L,C, striplines, et cetera. These filters can have an extremely high dynamic range, i.e. signal amplitudes in the filter are not limited by a power supply. Of course, in the end physical limitations set a maximum on the signal amplitudes. For instance, capacitor voltages are limited by the breakdown voltage of the corresponding capacitors. The noise level of these filters can also be low as only the resistive elements introduce noise. Again, because of the absence of active elements no corresponding noise degradation is found.

The bandwidth performance of those filters is limited by the bandwidth of the passive elements only. Therefore, for extremely high frequency filters, e.g. micro-wave filters, passive filters are used.

With respect to analog active time-continuous filters, these filters can be tuned less easily. It requires tuning of capacitor values or inductor values, for instance. Very often this requires mechanical tuning of those elements.

1.5.3 Asynchronous digital filters

When the signal is discretized in the amplitude but not in the time, the class of asynchronous digital filters is found. Asynchronous digital circuits are not governed by a clock signal. Triggering events is based on the information arrived locally so far. For instance, two digital words are added when both have arrived at the adder. In this way no clock-skew problems are found in these circuits. In modern digital systems a big challenge is to minimize the skew in the high speed clock all over the chip.

To the knowledge of the author, no filter implementations are found in this class. An explanation can be that for filtering (in the frequency domain) relative timing information is required. When the mutual timing relation of the digital

words is not determined, it is difficult (may be impossible) to use these digital words to do filtering in the time (frequency) domain. Of course, when the sampling of the (analog) input signal is synchronous, this objection is not valid.

1.5.4 Sampled-date filters

These filters have continuous amplitude behavior and a discrete time behavior. The filter reacts on the signal at fixed points in time only. Examples of these filters are for instance the switched-capacitor filters. In these filters the resistors are replaced by switched capacitors. These filters show the same dynamic range as the analog time-continuous filters for switching frequencies being minimally the double of the maximum signal frequency. This is explained by the fact that the switched capacitor is an exact replica of a resistor for signals with a frequency up to the half of the switching frequency. Of course, due to the switching action additional noise is introduced in the filters.

Thanks to the time discrete behavior these filters are able to implement transfer functions which are not possible for time-continuous filters. An example are the FIR filters (Finite Input Response). Also tuning of these filters is easily realized by changing the switching frequency.

1.5.5 Synchronous digital filters

This class of filters exhibit a time discrete and amplitude discrete behavior. Filtering is obtained by doing numerical processing on the data, for instance by a computer. The subsequent samples have a predefined time relation. With this class reconfigurable filters can easily be implemented, i.e. filter curve is changed when an other algorithm is used on the data stream.

1.6 Filter transfer and topology

The most convenient way of designing continuous time filters is to use the known filter theory of passive filters. The design path followed uses standard filter tables to determine component values in order to attain the desired transfer function. The dynamic range –defined as the maximal signal with respect to the noise level that the filter is able to handle at the same time– is in the case of passive filters (only L and C used) not limited by noise, because reactive components do not introduce noise. The maximally possible signal levels at the capacitors and inductors are (almost) infinite; they are limited by the dissipation in the resistors and the breakdown in the capacitors and saturation of the inductors. Noise of the (also parasitic) resistors puts a lower limit on the smallest signals to be handled.

In active integrated filters, the poles and zeros are determined by resistors and capacitors, as well as by the active components. The use of resistors introduces noise, as does the use of active components. The active components are also assumed to operate within a certain supply voltage, thus limiting the maximal output signal. Hence, a translation is necessary to use the conventional filter theory for constructing active continuous time filters.

1.6.1 The state space description

The state space description is used to combine the transfer function and the topology of a filter in one description. As the filter can be viewed as a linear differential equation, the state space must be able to represent this. The variable s is the Laplace variable, which determines the poles and zeros of the filter. Suppose a transfer function of two polynomials (the order of the denominator is n, which is greater than or equal to the order of the numerator). This transfer function can be represented in a signal flow graph as a connection of n integrators or differentiators. From this point, it is assumed that only integrators are used, as differentiators appear to be difficult to implement because of the critical stability considerations. This does not imply restrictions on the quality aspects of the filter or design freedom. The connection of the integrators can be described in the following equations:

$$s\mathbf{X} = \mathbf{AX} + \mathbf{BU} \tag{1.3}$$

$$\mathbf{Y} = \mathbf{CX} + \mathbf{DU} \tag{1.4}$$

Thus, the new inputs of the integrators ($s\mathbf{X}$) are a function of the old output signals of the integrators ($\mathbf{X}$) and of the input signals ($\mathbf{U}$). The output of the filter ($\mathbf{Y}$) is a combination of the output of the integrators and some fraction D of the input signal. The factor $\mathbf{D}$ can always be neglected in DR calculations, as this signal through this branch does not interfere with the internal structure of the filter.

$\mathbf{A}$,$\mathbf{B}$,$\mathbf{C}$ and $\mathbf{D}$ are matrices, which look like:

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} \tag{1.5}$$

$$\mathbf{B} = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix} \tag{1.6}$$

$$\mathbf{C} = \begin{pmatrix} c_1 & \cdots & c_n \end{pmatrix} \tag{1.7}$$

$$\mathbf{D} = (d) \tag{1.8}$$

The resulting (scalar) transfer function is:

$$H(s) = \frac{Y}{U} = \mathbf{C}(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{B} + \mathbf{D} \tag{1.9}$$

This is rendered schematically in figure 1.9.

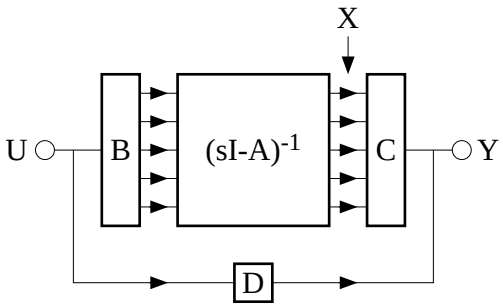

Figure 1.9: Scheme of state-space description

1.6.2 Signal-flow diagram

Signal-flow diagrams are widely used to represent the filter topology and the transfer functions of the comprising elements in a graphical way. In figure 1.10 the basic element is depicted. This diagram should be read as follows. Signal U is trans-

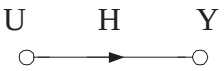

Figure 1.10: Basic element of a signal-flow diagram.

ferred via transfer H into signal Y. The signal U and Y can be either voltage, or current, or charge, etc. So, the signal-flow diagram makes visible what the relation is between two signals. A complete signal flow diagram can comprise several of these branches but also nodes where several signals are added or subtracted.

In figure 1.11 an example is given of the use of a signal-flow diagram of a first-order low-pass filter. Figure 1.11A depicts the circuit diagram of a first-order low-pass filter. The input voltage is U_i and the output voltage is U_o. In order to end up with a signal-flow diagram, the signals and transfers have to be made

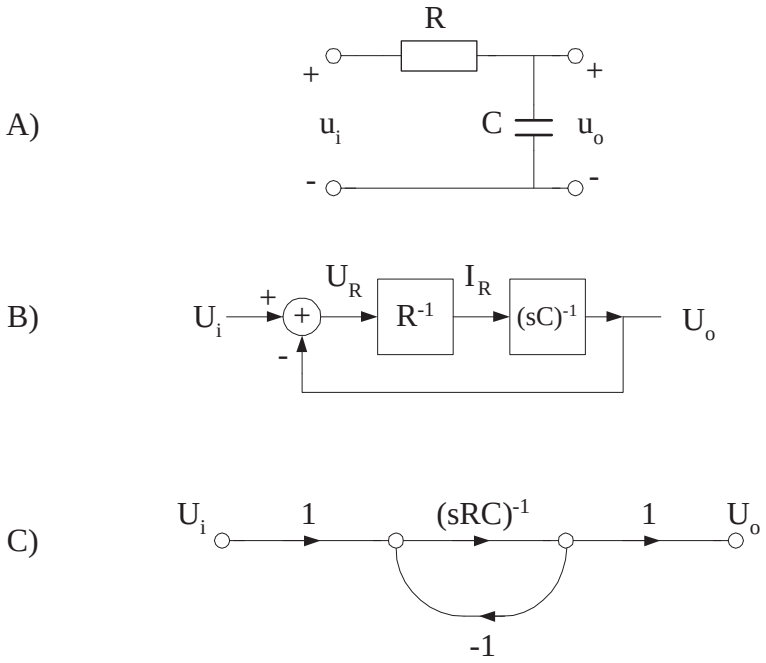

Figure 1.11: A signal-flow diagram used to describe a first-order low-pass filter. A) Circuit diagram B) An intermediate block diagram C) Signal-flow diagram.

explicit. As intermediate step, for the sake of clarity, a block diagram is derived. In the filter of figure 1.11 the following signals and transfers can be found:

$$\begin{aligned} U_R &= U_i - U_o \\ I_R &= U_R/R \\ U_o &= I_R/(sC) \end{aligned} \tag{1.10}$$

This equations can be found by starting at the input signal (U_i) and via the circuit follow the signal to the output signal (U_o). These three expressions are represented by the block diagram of figure 1.11B. The first equation is giving by the subtractor, i.e. $U_R = U_i - U_o$ whereas the other two equations are represented by the two blocks. This diagram is already close to the signal-flow diagram which is illustrated in figure 1.11C. When calculating from this signal-flow diagram the following transfer is found:

$$H(s) = \frac{U_o}{U_i} = \frac{1}{1 + sRC} \tag{1.11}$$

1.6.3 Filter realizations

There are some known methods to determine a topology and to realize a transfer function electronically. Here three ways are shown.

- Ladder realization from passive ladder filter
- Cascade realization
- Direct realization

1.6.4 Ladder realization from passive filter

One method is to derive the state-space description from a passive filter, after which an active implementations is constructed. For a given transfer function the passive filter circuit can be found easily from standard tables. Subsequently, the active implementation can be derived as depicted in figure 1.12. Starting point is a passive realization of a filter. This passive realization is subsequently depicted as a block diagram (like figure 1.11B). Then easily a signal-flow diagram is found; the 3rd step. Subsequently an implementation is depicted by using active a_{ii} (transconductances) and passive intergrators. The fifth diagram in figure 1.12 depicts the same filter function but now with passive a_{ii} and active integrators.

1.6.5 Cascade Realization

The cascade realization can be directly found by splitting the denominator of the transfer function into first- and second-order transfer functions. Applying this method, a third-order Butterworth filter (normalized) is written as:

$$H(s) = \frac{\frac{1}{2}}{(s+1)(s^2+s+1)} \tag{1.12}$$

The corresponding signal flow graph shown in figure 1.13. How to derive a signal flow diagram for a filter when the transfer function is given is described further on. By implementing the $1/s$ branches by active integrators, an active filter results and the scale factors by for instance resistors and active implementation of the filter is obtained.

1.6.6 Direct realization

The direct realization extracts the signal flow graph directly from the transfer function without factorization of the denominator. Taking for this example a

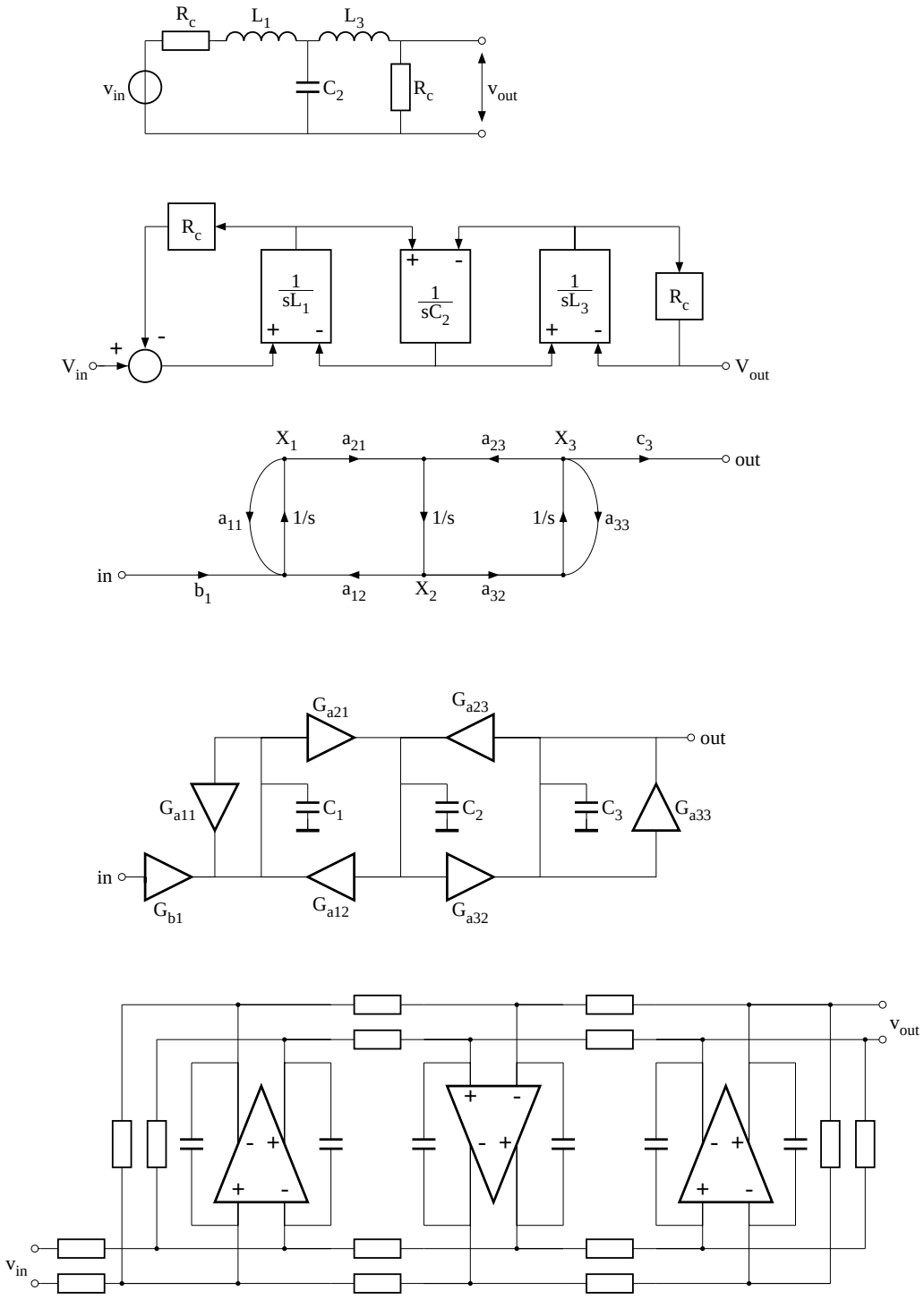

Figure 1.12: Transformation from passive filter to active filter

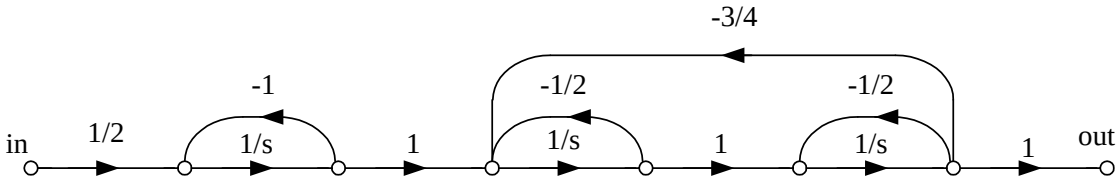

Figure 1.13: Signal flow graph of cascade realization

general expression for a Butterworth characteristic:

$$H(s) = \frac{\frac{\omega_c^3}{2}}{s^3 + 2\omega_c s^2 + 2\omega_c^2 s + \omega_c^3}. \tag{1.13}$$

in which ω_c is the bandwidth of the filter. The corresponding signal flow graph is shown in figure 1.14. Again by implementing the $1/s$-terms by, for instance, active integrators and the scale factor by resistors an active filter is obtained.

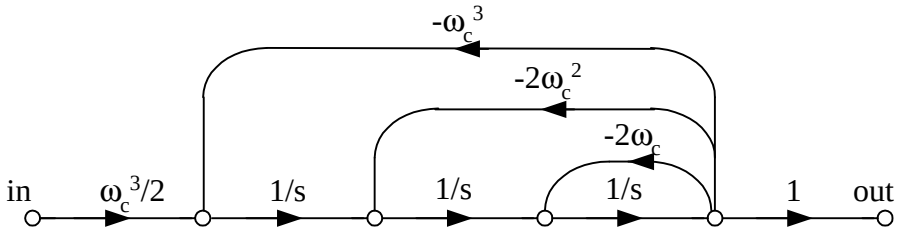

Figure 1.14: Signal flow graph of direct realization

The former examples of the implementation of the transfer functions show that it is possible to use several topologies to realize the same transfer function. This, however, does not imply that the various topologies behave the same. Various topologies, for example, appear to show different dynamic range behavior, but also different sensitivity behavior.

This dependency on the topology necessitate the use of the state space description as described in section 1.6.1. It not only describes the desired transfer function, but also the topology of the filter. Thus the use of the state-space description enables the optimization of filter topology in order to end up with the maximum dynamic range.

1.7 Construction of time-continuous filters

Roughly speaking, the design of filter can be split into two steps:

- determine optimal topology;

- implement topology.

In this section it is shown how a transfer function can be converted, via a state-space description, to a topology with ideal building blocks. In the next chapters guidelines are given for deriving an *optimal* topology, i.e. to attain the highest dynamic range.

Starting point is the required filter transfer function. It might also happen that the specification are in terms of attenuation and so on. Then it is also up to the designer to determine what filter transfer is required to reach those specifications. Here we assume that a filter transfer is already given and we take again the general third-order low-pass Butterworth transfer:

$$H(s) = \frac{Y(s)}{U(s)} \frac{\frac{\omega_c^3}{2}}{s^3 + 2\omega_c s^2 + 2\omega_c^2 s + \omega_c^3}. \tag{1.14}$$

The state-space description is matrix description of a coupled set of n *first-order* differential equations (see equations (1.3) and (1.4)). Therefore, a straight-forward way is to write down the filter transfer as a n^{th}-order differential equation and subsequently convert it to a coupled set of n *first-order* differential equations. Then the state-space description is readily obtained as well as the corresponding topology.

Transfer function equation (1.14) described as differential equation is given by:

$$\omega_c^{-3} \dddot{y}(t) + 2\omega_c^{-2} \ddot{y}(t) + 2\omega_c^{-1} \dot{y}(t) + y(t) = \frac{1}{2} u(t) \tag{1.15}$$

Next step is to write this 3^{rd}-order differential equation as 3 coupled first-order differential equations. This is easily achieved by just defining three states (x_1, x_2 and x_3) which are "coupled" in cascade.

$$\begin{aligned} x_1(t) &= y(t) \\ x_2(t) &= \dot{x_1}(t) = \dot{y}(t) \\ x_3(t) &= \dot{x_2}(t) = \ddot{x_1}(t) = \ddot{y}(t) \end{aligned} \tag{1.16}$$

In order to obtain a complete state-space description, matrices **A**, **B**, **C**, and **D** should be determined. The new states relate to the old states and the input via matrix **A** and **B**:

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}u \tag{1.17}$$

The expression for matrix **A** is easily derived from equation (1.16) as:

$$\begin{pmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -\omega_c^3 & -2\omega_c^2 & -2\omega_c \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ \frac{1}{2}\omega_c^3 \end{pmatrix} u(t) \tag{1.18}$$

The output signal of the filter is related via matrix **C** and **D** to, respectively, the old states and the input signal, according to:

$$y = \mathbf{C}\mathbf{x} + \mathbf{D}u \tag{1.19}$$

In equation (1.16) an explicit expression for $y(t)$ is already given and thus the matrix notation is given by:

$$y(t) = \begin{pmatrix} 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} + \begin{pmatrix} 0 \end{pmatrix} u(t) \tag{1.20}$$

Given this state-space description, a signal-flow diagram is easily obtained. As said, the state-space description is a set of first-order differential equations. Each of these equations should be represented by means of one or more branches and nodes in the signal flow diagram. Drawing the signal-flow diagram related to the example of this section, we'll start with the first line in the matrix equation (1.18), i.e. $\dot{x}_1 = x_2$. In the signal-low diagram derivatives of states should not appear. Thus the expression to be represented should be either (in frequency domain):

$$sX_1 = X_2 \tag{1.21}$$

or

$$X_1 = X_2/s \tag{1.22}$$

in which s is the complex variable. For the first case a differentiator is used, whereas in the second case an integrator is used. For now we choose for using integrators. Further on, it is argued that implementations based on integrators are more easily to implement practically than those based on differentiators. Figure 1.15 shows the signal-flow representation of equation (1.22). Subsequently, the second line of

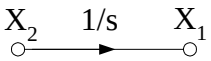

Figure 1.15: Signal-flow representation of $X_1 = X_2/s$

the matrix equation $\dot{x}_2 = x_3$ is rewritten in terms of an integration,

$$X_2 = X_3/s \tag{1.23}$$

This expression is analogous to expression (1.22) represented as a signal flow diagram. The result is added to figure 1.15 and figure 1.16 is the results. The third line of the matrix equation (1.18) yields:

$$\dot{x}_3 = -\omega_c^3 x_1 - 2\omega_c^2 x_2 - 2\omega_c x_3 + \frac{1}{2} u \omega_c^3 \tag{1.24}$$

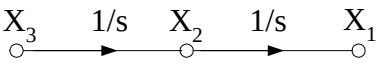

Figure 1.16: Signal-flow representation of $X_2 = X_3/s$ added to figure 1.15

This expression states the the time-derivative of state x_3 is equal to a sum of the stated plus a term proportional to the input signal u. In figure 1.16 the three states x_1, x_2 and x_3 have already been indicated. However, only two integrators are used and thus can never yield a third order system. This third integrator is introduced by equation (1.24). That equations is in terms of $\dot{x}_3$ and in the signal-flow graph of figure 1.16 no $\dot{x}_3$ is found. To make this signal available an additional integrator needs to be added as indicated in figure 1.17. At the input of the most

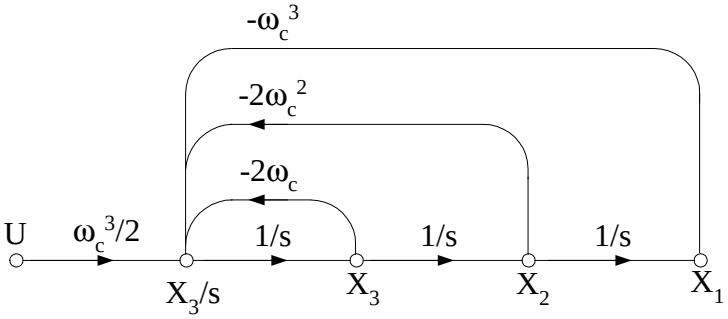

Figure 1.17: Signal-flow representation of matrix equation (1.18)

left integrator signal $\dot{x}_3$ is found as its output signal was defined as x_3. Adding the signals arriving at the node representing $\dot{x}_3$ yields expression (1.24). Finally, to obtain a complete signal-flow representation of the state-space description, also equation (1.20) should be taken into account. That equation determines how output signal y is obtained from the states and the input signal. In this case y equals x_1. This illustrated in figure 1.18.

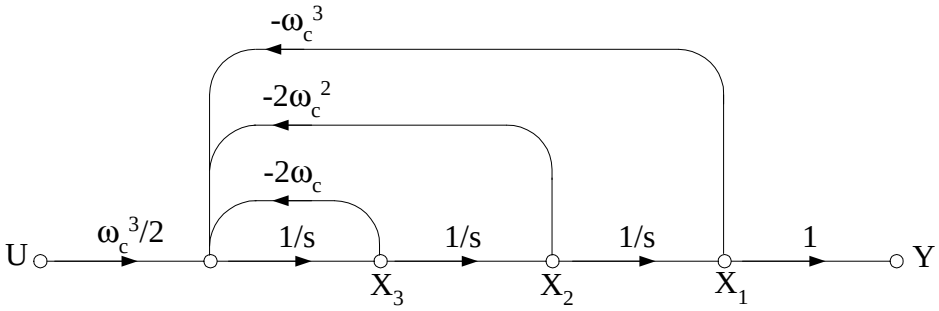

Figure 1.18: Signal-flow representation of state-space description equations (1.18) and (1.20)

1.8 Exercises

1. Give a summary of the history of filter design. What key developments can you indicate.

2. Derive the basic equation for the dynamic translinear filter:

$$CV_T \frac{di_c}{dt} = i_c i_{cap} \tag{1.25}$$

3. What is the main difference between a filter used for selection and a filter used for shaping?

4.

(a) What *quality* aspects do you know in the context of integrated circuit design?

(b) What *cost* aspects do you know in the context of integrated circuit design?

(c) How would you define *quality* and *cost*?

5. To which classes belong the following filters, explain your choice:

(a) A filter realized by combining several taps of a charge-coupled device (CCD)?

(b) A wave guide on a chip which has transmission zeros for certain frequencies?

(c) A filter to do picture enhancement in a digital camera?

6. Given the following applications. From which class of filter should you select a solution, motivate your choice?

(a) Separation of Low and High-frequencies from an audio signal to drive a woofer and tweeter of a loud speaker, respectively.

(b) To implement an RIAA correction in a tape-deck?

(c) A channel selection filter in a GSM telephone?

7. Given the filter of figure 1.19.

(a) To what class belongs this filter?

(b) Explain whether the filter has low-pass, band-pass or high-pass behavior.

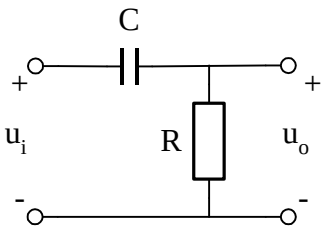

Figure 1.19: Circuit diagram of a RC filter

(c) Derive the signal-flow diagram for this filter.

8. Given the following transfer function:

$$H(s) = \frac{s^2 - 2s + 2}{s^2 + 2s + 2} \tag{1.26}$$

(a) Draw the bode plots of this filter.

(b) For what application can this filter be used?

(c) Design this filter using integrators and gain blocks.

9. Given the filter transfer function according to:

$$H(s) = \frac{s^2}{s^2 + 1.4 \cdot 10^7 s + 10^7} \tag{1.27}$$

This filter function is to be implemented by means of ideal integrators and scalers.

(a) Give the derivation of the state-space description.

(b) Draw the block diagram with ideal integrators and scalers. Do not forget to indicate the input and output of the filter.

10. Given the filter transfer function:

$$H(s) = \frac{(s^2 - s + 1)(s - 1)}{(s^2 + s + 1)(s + 1)} \tag{1.28}$$

(a) What is the type of filter transfer (All pass, Band pass, High pass, Low pass), motivate your choice?

(b) Derive the state-space description.

(c) Draw the block diagram of this filter.

11. Given the following high-pass filter with the poles in a Bessel-Thomson configuration:

$$H(s) = \frac{s^3}{s^3 + 6s^2 + 15s + 15} \tag{1.29}$$

(a) Derive the state-space description

(b) Draw a block diagram with ideal integrators and scalers.

Chapter 2

DR optimization of filters

2.1 Introduction

The use of coils and large capacitances has the disadvantage of not being integratable. The advantage of using coils and capacitors for realizing filters is that coils and capacitors have considerable signal-handling capabilities. The dynamic range of passive filters can in theory be infinite, because only coils and capacitors determine the location of the poles and zeros. The terminating resistors at the input and the output of the filter introduce some thermal noise, but this is very little.

The dynamic range of active filters is not infinite. Limitations arise at the two sides of the dynamic range. The output capability is limited by the supply voltage. No signals appear that go beyond the supply voltage. The noise level is introduced by the use of resistors to determine the filter transfer. The amount of noise becomes worse when active devices are used; this gives rise to a noise factor.

Although not evident, the dynamic range appears to be dependent on the topology of the filter. It can be proven that to every filter transfer there is some corresponding maximal dynamic range together with a given supply voltage, total capacitance and noise factor of the active devices. Only for one topology can it be proven that this maximum dynamic range can be reached. No attention is paid here to the exact calculations, we address only some clarification of the principles.

In this chapter first, in section 2.2, the dynamic range of a filter is introduced. To maximize the dynamic range, the noise level should be minimized and the maximum possible output signal without distortion, ideally, should be maximized. Section 2.3 treats some basic concepts of noise, i.e ,modelling and how to determine an equivalent noise source for a simple circuit. Subsequently, section 2.4 describes what types of distortion can be distinguished and how they are caused. Finally, section 2.5 treats the maximization of the dynamic range of a filter.

2.2 Dynamic Range

In chapter 1 the quality of a filter (or in general of an electronic circuit) was related to the signal-handling capacity C as defined by Shannon:

$$C = B\ ^2 \log\left(1 + \frac{S}{N}\right) \tag{2.1}$$

Besides a linear dependence on the bandwidth, the signal-handling capacity also depends via a logarithm on the ratio of the signal power (S) and noise (N). Clearly, C is the largest (for a given bandwidth) when the ration S/N is the largest. The dynamic range is defined as:

> The dynamic range (DR) is defined as the ratio of the largest and smallest signal that can be processed at the *same* time:

$$DR = \frac{\text{Maximum signal}}{\text{Minimum signal}}\Big|_{\text{at the same time}} \tag{2.2}$$

This definition implicitly states that on the one hand the small signal is not contaminated by noise such that the information on that signal is lost and on the other hand the the information that is on the large signal is not destroyed by distortion. This is illustrated in figure 2.1 The adjectives "too large" and "too small" need some explanations. A signal is called "too small" when it gets lost in the noise. Of course, when lowering a signal level it will gradually be contaminated by noise and there will not be a very distinct (sharp) moment at which you can say "*Now* the signal is lost in the noise". To have an objective criterion, often for the minimum signal level, the signal is used that has the same power content as the power of the relevant noise. Section 2.3 will go into more detail about the power of the relevant noise. For defining a signal to be too large, a somewhat different approach is used. When a signal becomes large, an electronic circuit will produce distortion as a result of its non-linear behavior. Due those this distortion higher harmonics will arise. Often the power content of these higher harmonics is used to a criterion to say that a signal is too large. A criterion could be that the power content of the higher harmonics is equal to the noise power present in the system. Section 2.4 describes in more detail the distortion behavior of electronic circuits.

2.3 Noise

Noise is one of the phenomena that is ubiquitously present throughout physical mechanisms and physical systems. Such systems obey the well-predictable, deterministic rules only to a certain extend; the noise introduces slight non-predictable

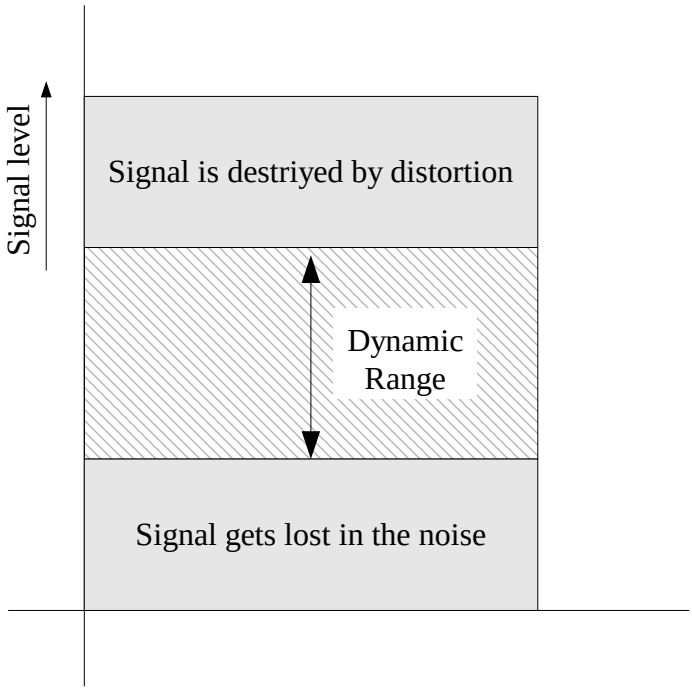

Figure 2.1: The dynamic range depicted as the window between "too large" signals and "too small" signals

perturbations from it. The hissing sounds produced e.g. by radios, televisions, and telephones between messages is perhaps the most widely known example of noise. In engineering, however, the term noise is used to refer to a much wider class of very diverse random phenomena. As you may have expected already, electronic components are subjected to noise as well; their behaviour is also impaired by slight non-predictable perturbations.

In noise analysis of electronic circuits the noise sources have to be obtained from the noise models of the various circuit components. In addition, these models reveal the origin of the noise processes, and their mutual correlations. In this section the noise model for resistors is discussed.

2.3.1 General Characteristics of Electronic Circuit Noise

Although several different types of electronic circuit noise, with slightly different causes can be identified, all of them are fundamentally due to the quantized nature of electric charge on a microscopic scale. As a result of this quantized nature, electric current is not a continuous flow of charge, as it is modeled in circuit theory, but a stream of charged particles moving on average in one direction. Due to several (microscopic) mechanisms, their actual path of motion can be very

irregular and very different from this average direction. As a consequence, the instantaneous current flowing through electronic components will slightly fluctuate in a random manner: it contains noise.

Due to this common cause, the statistical characteristics of electronic noise processes are very similar. All these processes are due to random movements of large numbers of equally charged particles. From stochastics is known that the probability density function of such processes approaches a normal/Gaussian density function. For this reason, all electrical noise processes considered in this book possess an (approximately) Gaussian probability function $p(e_n)$:

$$p(e_n) = \frac{\exp\left[-\frac{(e_n-\mu_n)^2}{2\sigma_n^2}\right]}{\sigma_n\sqrt{2\pi}}. \tag{2.3}$$

The parameters μ_n and σ_n in this expression represent the average (expected) value and the standard deviation (the square root of the variance) of the process respectively. Since electrical noise is generally defined as the fluctuations of currents and voltages with respect to their average value, the average value of electrical noise processes equals zero:

$$\mu_n \equiv 0. \tag{2.4}$$

The value of the variance differs among the types of noise processes, and is determined by circuit and component parameters. In the sequel, the variance for the noise process in a resistor will be specified in terms of the mean square value associated to a frequency band Δf.

2.3.2 Resistor Noise Model

Resistors produce a type of circuit noise that is called *thermal noise*. It is caused by (thermal) kinetic energy gained by free charge carriers when the temperature rises (thermal agitation). If the temperature is the same everywhere in the resistance material, this thermal motion of the charge carriers has no preference for any direction. As a result, thermal energy causes carriers to move randomly, uniformly distributed among all directions, which generates a random current/voltage.

In addition to the thermal agitation, an external applied source gives the motion of the charge carriers a drift component, in the same (or opposite) direction of the external current. The total external noticeable current/voltage equals the superposition of both effects, and therefore contains a random component: thermal noise.

Nyquist showed that the voltage fluctuations (voltage noise) observed at the terminals of a resistor with value R due to thermal agitation possesses a mean-square value associated to a frequency band Δf equal to:

$$\overline{v_n^2}(f, \Delta f) = 4kTR\Delta f, f \geq 0, \tag{2.5}$$

where $k = 1.38 \cdot 10^{-23}$ (J/K) denotes Boltzmann's constant, and T the absolute temperature in Kelvin. Thus, resistors produce white noise, i.e. noise that is frequency independent[1]. Figure 2.2 depicts the Thevenin and Norton equivalents of the resistor with thermal noise voltage source v_n and noise current source i_n, respectively. The resistor depicted in the figure is an ideal one, i.e. the resistor

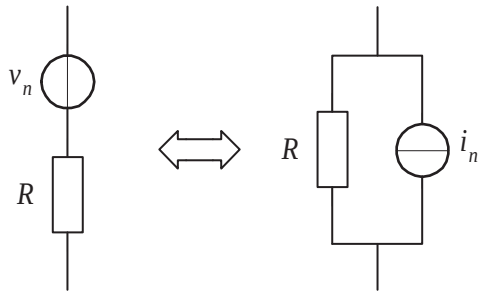

Figure 2.2: Noise model of a resistor.

defined in circuit theory by Ohm's law. So, it is noise free. The noise present in a physical resistor is modelled by the separate noise source.

Using the Thevenin-Norton transform and equation (2.5), we find that the mean square value of the noise current in the Norton equivalent equals:

$$\overline{i_n^2}\,(f, \Delta f) = \frac{4kT\Delta f}{R}, f \geq 0. \tag{2.6}$$

This is explained any further in the next sections.

In resistors made of materials with a very irregular internal structure, another type of noise, so called $1/f$ noise or flicker noise, may yield a non-negligible contribution to the resistor noise production. This is sometimes the case for resistors made of compressed coal powder. The variance of this flicker noise can be written as:

$$\overline{v_n^2}\,(f, \Delta f) = 4kTR\frac{f_l}{f}\Delta f, f \geq 0, \tag{2.7}$$

where f_l denotes the frequency above which the white thermal noise dominates. For frequencies below f_l, the flicker noise dominates. The flicker noise is uncorrelated with the thermal white noise.

2.3.3 Power versus Power-Density Spectrum

From an information theoretical point of view, only the noise that cannot be distinguished from the information signal fundamentally limits the information handling

[1]In reality, the thermal noise of the resistor has an extremely large, but finite bandwidth. Its power spectral density, however, remains white up to frequencies far beyond the range of interest to electronic circuit design.

capacity. In figure 2.3, this is the noise located inside the same frequency band B_{inf} as the information signal. Any noise located outside this bandwidth can in principle be removed by means of frequency filtering. Consequently, the noise power located in the information bandwidth is a lower bound, and the corresponding dynamic range an upper bound on the dynamic range that can exist within the filter.

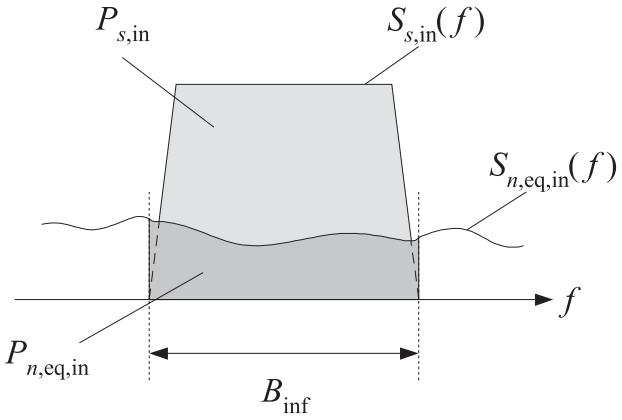

Figure 2.3: Schematic representation of the noise- and signal power spectral densities.

The most important observation, which is generally true, to be made from figure 2.3 is that the noise is spread over a much wider frequency range than the information signal, which is bound to the *information bandwidth* B_{inf} as shown.

For this reason, we define the noise power P_n to be part of the power produced by the noise source that resides in the same frequency range as the information signal. Power equals the area under the power density spectrum, such that:

$$P_n \stackrel{def}{=} \int_{B_{\text{inf}}} S_n(f)\mathrm{d}f. \tag{2.8}$$

2.3.4 Equivalent noise source

Within an electronic circuit mostly various noise sources are present. So determine the dynamic range, the effect of all these noise sources on the signal present in the circuit needs to be analyzed. As one of the first steps in this noise analysis, we have to determine the so called equivalent (output or input) noise source. This source models the noise experienced at the filter output (or input), due to internal circuit noise production. It concentrates the entire circuit noise production into one circuit branch, and thereby provides straight-forward calculation of the experienced noise power.

To determine the contributions of all internal noise sources to the equivalent noise sources, a combination of various transformations, originating from circuit theory, can be used. In contrast, no use of any statistical property of the noise sources is required to obtain the equivalent noise source; these are required only to determine the equivalent noise power.

The equivalent noise source replaces all other noise sources in the filter circuit. Determination of this source may therefore be viewed as 'wiping' all circuit noise to the filter output (or input), resulting in one noise source, and a noise-free filter.

This 'wiping' or transformation of circuit noise to the input is schematically represented by figure 2.4. The circuit noise processes are distributed all over the

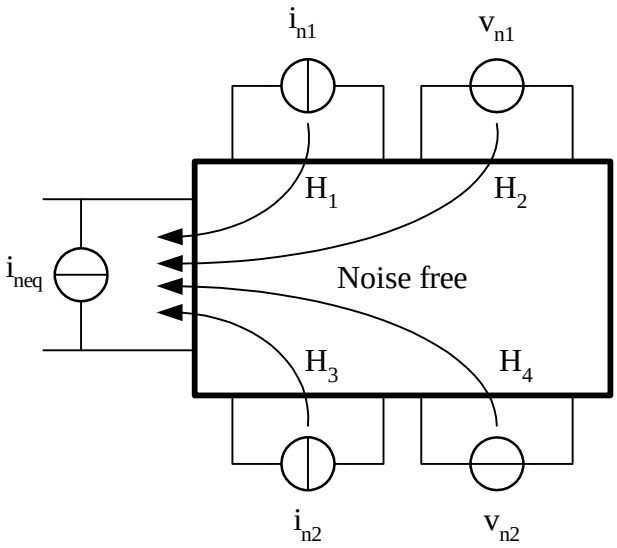

Figure 2.4: Representation of the filter circuit as a noise-free multi-port network, connected to multiple noise sources.

filter circuit, and can each be represented by *independent* current (i_{n1}, i_{n2}) or voltage sources (v_{n1}, v_{n2}). These sources can be considered as external inputs to the otherwise noise-free filter, which is represented as a multi-port network.

The figure shows that in order to obtain the equivalent noise source (i_{neq}), we basically have to determine the transfer functions $H_1(s) - H_4(s)$, with s the Laplace variable. We can do this in a straight-forward way, by evaluating the circuit equations of the filter or by using some network transformation. This is to be discussed further on.

2.3.5 Power of the equivalent source

In the calculation of the power of an equivalent noise source we only have to express the power spectral density of the equivalent noise in terms of the spectral densities

of the various noise processes. From signal theory can be derived that the power spectral density of the equivalent source, $S_{eq}(f)$ is related to the power spectral density $S_i(f)$, in the case they are uncorrelated, through:

$$S_{eq}(f) = \sum_{i=1}^{4} |H_i(s)|^2 S_i(f), \tag{2.9}$$

in which s is the Laplace variable (based on the Wiener-Kintchine theorem). The power spectral densities $S_i(f)$ are often easily to determine. For instance, for a resistor the power spectral density is given by either equation (2.5) or equation (2.6), depending on the type of noise source used, voltage or current source, respectively. The transfer function $H_i(s)$ can be determined also relatively easily.

2.3.6 Network transformations

Determining the equivalent noise source could be done in a straight-forward way, by evaluating the circuit equations of the filter. As will be shown, for the state-space filters this can be a good method to use as all the circuit equation are comprised in the state-space description.

However, this straightforward method is not suited in general. Instead of such a 'global' one-step transform, it is also possible to break the transformation of a noise source to the equivalent input noise into several consecutive steps. In this way, the approach exploits knowledge about the topology of the network, which is very similar for the various circuit types, and in general shows the dominant factors and key parameters determining the circuit noise behavior. Here four different types of transformations are separately discussed.

Transform-I: Voltage Source Shift

The voltage source shift (V-shift) is a transform that enables to move (noise) voltage sources through the network. Generally, it is used to shift these sources to the network input, the output, or a branch where it can be subjected to another type of transform.

The major constraint to be posed on any source transform is that the transform itself does not change the noise current/voltage experienced at the circuit input (or output); it should not be noticeable to an observer measuring the noise voltage/current at these points whether or not noise sources have been transformed/shifted internally in the network. For the V-shift, this means that it must not change the Kirchhoff Voltage Law (KVL) of any mesh in the circuit.

The V-shift obeying this constraint is visualized in figure 2.5. The original (noise) source v_n is shifted out of the branch between the nodes 1,4 into the two other branches connected to node 4; the ones between 2,4 and 3,4. In order

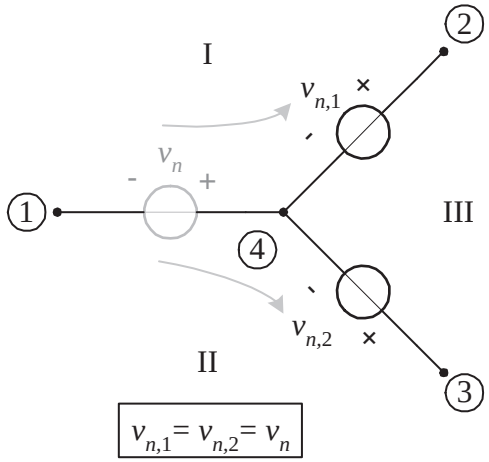

Figure 2.5: V-shift transform.

to guarantee that the KVLs of the meshes I,II,III, associated to node 4 remain unchanged, the new sources v_{n1} and v_{n2} have to be exactly equal to each other and to the original source v_n.

This means that the stochastic processes v_n, v_{n1} and v_{n2} are fully correlated, and possess identical stochastic properties. Observe what happens with the V-shift in the general case, when node 4 is connected to n branches ($n > 2$). Then, shifting v_n through node 4 from its original branch to the $n-1$ other branches yields $n-1$ identical sources, instead of just 2. Further, note that the V-shift allows to move v_n around mesh I and II, but not out of it; this would change the KVL.

Transform-II: Blakesley transform

The dual transform of the V-shift is the Blakesley transform, which allows to move current (noise) sources through the network. It is generally used to move these sources to the network input port, output port, or intermediate nodes that allow another type of transform.

Whereas the V-shift is not allowed to affect the KVL of any circuit mesh, the Blakesley transform is not allowed to change the Kirchhoff Current Law (KCL) of any circuit node, for the same reason. The transform obeying this constraint is depicted in figure 2.6. The original (noise) current source i_n is redirected from the branch between nodes 1,2 through the sources $i_{n,1}$ and $i_{n,2}$ between nodes 1,3 and 2,3. In order to keep the KCLs of the nodes 1,2, and 3 unchanged, $i_{n,1}$ and $i_{n,2}$ have to be exactly equal to each other and to i_n, and also have be be directed as illustrated.

Similar to the V-shift, this means that the original source i_n and the trans-

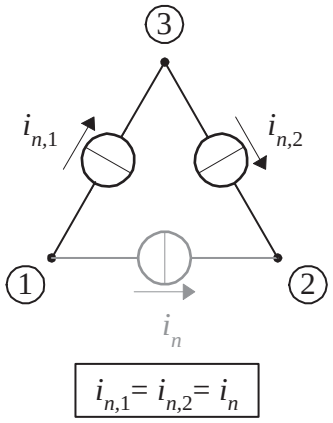

Figure 2.6: The Blakesley transform.

formed sources $i_{n,1}$ and $i_{n,2}$ have to be fully correlated and possess identical statistical properties. We further notice that, similar to the V-shift, the I-shift allows to move a current source around the network, but cannot be used to disconnect it from the original nodes; this would change the KCL.

Transform-III: Norton-Thevenin Transform

The equivalence of the well-known theorems of Norton and Thevenin can be used to transform a (noise) current source into a (noise) voltage source and vice versa. This type of transform does essentially not move sources through the filter network, but is used to switch between the V-shift and Blakesley transform.

The Norton-Thevenin transform, which automatically obeys the constraint that it does not affect the observed output noise (why ?), is depicted in figure 2.7.

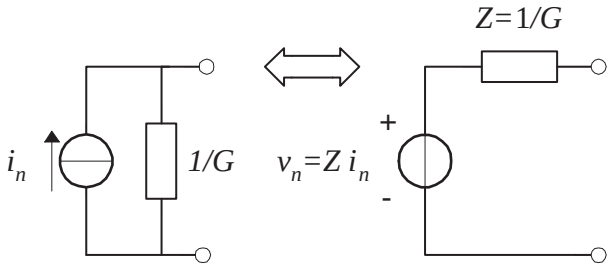

Figure 2.7: Norton-Thevenin transform.

The current source i_n and the voltage source v_n have a one to one correspondence through the impedance Z. The stochastic processes i_n and v_n are therefore fully correlated, and possess similar, though not identical stochastic properties.

Note that this transformation does change the KVLs and KCLs of the circuit; it exchanges a circuit branch for a circuit node, and vice versa. For this reason, the Norton-Thevenin transform, in combination with the V-shift and Blakesley transform can be used to eliminate a voltage source from a mesh, or a current source from a node.

Transform-IV: Shift through twoports

The three transformations we have considered so far are all concerned with two-terminal elements (one-ports) only. If a network consists entirely of such elements these three transforms are all we need to determine the equivalent input noise.

Many circuits, however, also contain elementary twoports, controlled sources (included in transistor models) or a nullor, that cannot be replaced by any combination of one-ports. For such networks, we need an additional transform: the twoport shift. This transformation is illustrated by 2.8. The output voltage v_o and

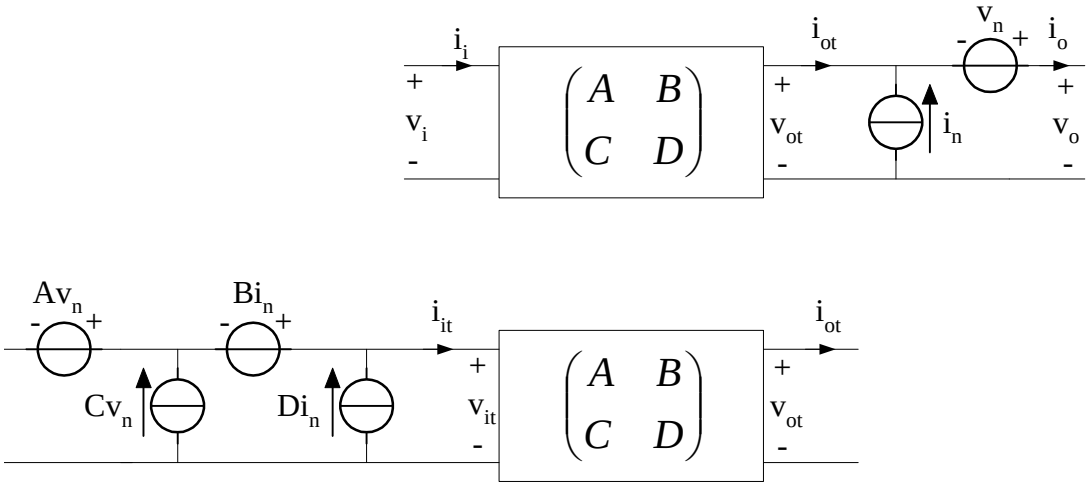

Figure 2.8: Transformation of noise sources through a twoport.

the output current i_o of the twoport are mutilated by a noise voltage v_n and noise current i_n respectively, as depicted in the upper part of the figure. The purpose of the twoport shift transform is to obtain the equivalent input noise sources that yields this output noise. The result depicted in the lower part of the figure is easily obtained from the chain matrix equation for the twoport:

$$\begin{pmatrix} v_{it} \\ i_{it} \end{pmatrix} = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} v_{ot} \\ i_{ot} \end{pmatrix} \tag{2.10}$$

and substitution of v_i, i_i, v_o, i_o, v_n and i_n. Observe that the output noise voltage source v_n is transformed into an input noise voltage source Av_n and a noise current

source Cv_n that are fully correlated. Likewise, the output noise current source i_n transforms into an input noise voltage source Bi_n and an input noise current source Di_n.

2.3.7 Correlated noise sources

Via the transforms discussed in the previous section, an expression is found for the equivalent noise in terms of the comprising noise sources. For circuit design, the important property of the noise source is its spectral power density or its power content. Equation (2.9) shows how the spectral power density of the equivalent sources relates to the spectral power densities of the comprising noise sources. An important constraint for using this equation is that the noise sources are uncorrelated. It is easily demonstrated that by using the transforms fully correlated sources arise. So, care has to be taken when determining the spectral power density or power of the equivalent source.

How to proceed when fully correlated sources are obtained can also be found from equation (2.9). The equation use for each original noise source in the circuit a transfer function describing the transfer from the original noise source to the equivalent noise source. Thus, when fully correlated noise sources are obtained, i.e. originating from the same noise source, the corresponding transfer functions should be added to a new transfer function. By doing this for all the fully correlated noise sources, the remaining sources will be uncorrelated and the expression (2.9) can be used.

Assume that after some transformations, a series connection of three noise-voltage sources is obtained as illustrated in figure 2.9. Then, combining the correlated sources to a new source yields two uncorrelated sources:

$$v_{n1} \text{ and } v_{n2} \tag{2.11}$$

with corresponding transfer functions

$$H_1 = a - b \text{ and } H_2 = c. \tag{2.12}$$

Thus, the equivalent source is described as

$$v_{neq} = H_1 v_{n1} + H_2 v_{n2}. \tag{2.13}$$

As both sources are uncorrelated, equation (2.9) can be used:

$$S_{neq} = |H_1|^2 S_{n1} + |H_2|^2 S_{n2} \tag{2.14}$$

yielding

$$S_{neq} = (a - b)^2 S_{n1} + c^2 S_{n2} \tag{2.15}$$

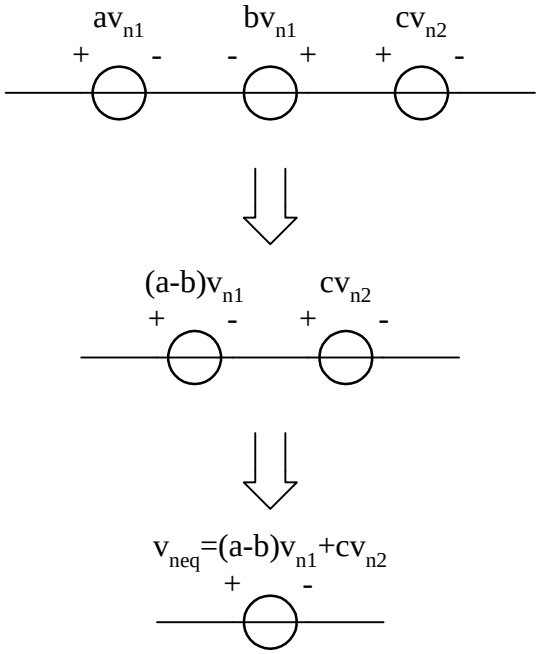

Figure 2.9: Three noise sources that are not completely uncorrelated, that add up to an equivalent noise source

2.4 Distortion

Distortion of a signal occurs, generally speaking, when the actual output signal deviates from the expected output signal, i.e. the signal which would be found at the output of the ideal filter. Depending on the nature of the distortion, information may be lost. Therefore a distinction is made between:

- weak distortion;
- clipping distortion.

Weak distortion occurs when the static transfer of a system deviates from the intended static transfer, see figure 2.10. The actual static transfer still has an inverse function so that a compensating function can be found for preventing the system from losing information. Thus, for instance, intentionally using the exponential transfer of a transistor does not mean that information is lost (remind the dynamic translinear filters); by means of the inverse function, the logarithm, the information can be retrieved again.

The other type of distortion is found when the transfer of a system no longer has an inverse function, i.e. the transfer has become ambiguous; and thus the original information can no longer be retrieved, see also figure 2.10. This occurs, for instance, when signals clip to the supply voltage. Two different types of input

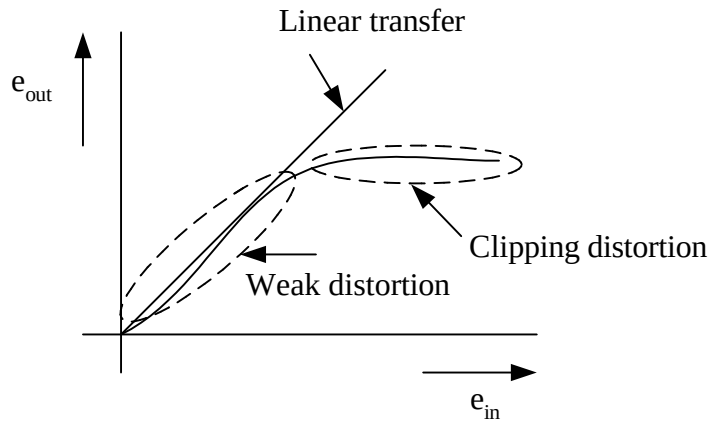

Figure 2.10: Weak and clipping distortion arising from a function not being linear as intended

signals, one causing clipping and the other being close to clipping, may result in the same output signal. A very straightforward example is the use of a limiter to get rid of all the amplitude *information* in a signal.

When circuit transfers are perfectly corresponding with the intended transfer, no weak distortion is obtained. In that case the upper limit of signal amplitudes is given by the clipping distortion. Thus an upper limit for signals is given by supply voltage and supply current.

2.5 DR optimization

Clearly, the dynamic range of a filter depends on the noise and distortion performance of the comprising elements. For instance, the noise of the resistors used in the filter implementation will somehow limit the dynamic range of the filter at the lower bound. The upper bound of the dynamic range is likely to be limited by the power-supply voltage limiting signal swing. Doing careful calculations, it appears that the effect on the dynamic range of all the separate components in the filter are completely taken into account when the dynamic range of the integrators are studied, which is not trivial.

In chapter 1 several topologies where shown having the same transfer function. Noticing that for the different topologies the integrators are connected differently to each other and the input and output of the filter, it is not surprising that the different topologies exhibit different dynamic ranges.

It can be said that for optimizing the dynamic range of a filter two key issues need to be addressed:

- optimization of the dynamic ranges of the integrators in the filter;

- optimization of the interconnection of the integrators, i.e. the topology of the filter.

2.5.1 Dynamic range of integrators

The dynamic range of integrators consisting of a resistor and a capacitor is partly determined by the maximal signal levels the integrator is able to handle. Because integrators are coupled in a filter, all integrators must have approximately equal signal-handling capabilities. When the maximal signal amplitude is V_{max}, the maximal signal power level is by definition $V_{max}^2/2$.

The noise of the integrator of figure 2.11 can be modeled as a noise voltage source at the input, with a single-sided spectrum (f $\geq$ 0) of:

$$S_{ni}(f) = 4kT\xi R \tag{2.16}$$

in which ξ is the noise factor of the integrator with a minimum value of 1. This noise factor accounts for the surplus noise of the active devices. The mean squared

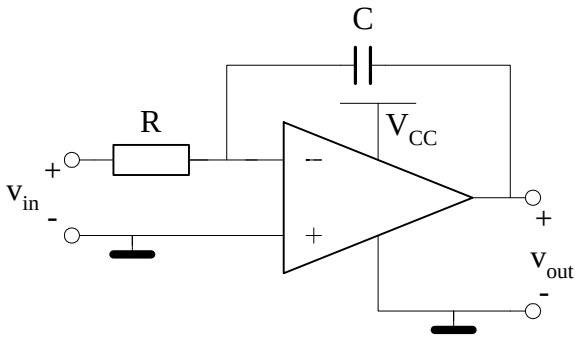

Figure 2.11: An active integrator

noise voltage can be determined by integrating over the noise bandwidth B, which is chosen to be equal to the unity gain frequency of the integrator, $B = 1/(2\pi RC)$:

$$P_{ni} = \int_0^\infty S_{ni}(f)df \approx \int_0^B S_{ni}(f)df = \frac{2kT\xi}{\pi C} \tag{2.17}$$

From this a very remarkable conclusion can be drawn: in the integrator the resistor is the element generating noise, however, looking to the total noise power of the integrator, only the value of the capacitor is what matters! The dynamic range of the integrator becomes:

$$DR = \frac{V_{max}^2/2}{\frac{2kT\xi}{\pi C}} = \frac{\pi V_{max}^2 C}{4kT\xi} \tag{2.18}$$

Note that for a single-sided supply voltage holds: $V_{max} = V_{CC}/2$.

The same calculations can be performed for the completely passive integrator (first-order low-pass filter) as shown in 2.12. The spectral power density of the

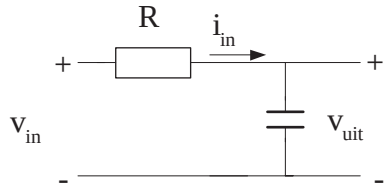

Figure 2.12: A first-order low-pass filter used as integrator

noise voltage source (v_{nin}) at the input equals:

$$S_{nin}(f) = 4kTR \tag{2.19}$$

The spectral power density of the noise voltage at the output of the filter becomes:

$$S_{nout}(f) = |H(s)|^2 S_{nin} = \frac{4kTR}{1 + (2\pi fRC)^2} \tag{2.20}$$

Integrating over the total frequency band, the total output noise power P_{nout} appears to be kT/C exactly. The dynamic range becomes, by assuming again a maximum signal level of $\frac{1}{2}V_{max}^2$:

$$DR = \frac{V_{max}^2 C}{2kT} \tag{2.21}$$

The difference between this expression for the DR and expression (2.18) is a result of the different definition of the band of integration. In the last situation the band of integration is infinite. For the active integrator of figure 2.11 it is not possible to integrate over an infinite band; for frequencies approaching 0 Hz the gain of the integrator becomes infinite. This integrator needs to be placed within the context of the filter network to be able to do calculate the total noise over an infinite band. So, for the active integrator an assumption had to be made. This assumption does not have any influence on the optimization of the dynamic range, only on the absolute value of the dynamic range that results.

Conclusion for the integrators is that the larger the capacitance, the larger the dynamic range is. On top of that, the active part of the integrator should be designed such that ξ approached 1 for optimal performance.

2.5.2 Matrix transforms

As the state-space description describe the filter topology and transfer function, it is suitable for doing dynamic range optimizations. It is used to calculate new or adapted topologies in order to optimize the dynamic range.

There are two methods of optimization. The first one is called *scaling*. Scaling does not change the topology of the filter. Simple "start" topologies are locally optimized. This method does not usually reach the maximally possible dynamic range. *Full optimization* is, however, able to reach this limit. In this case the topology is also changed. All possible connections between integrators may be used to obtain the maximal dynamic range.

The change of topology, without changing the transfer function, can be described by the following transformation with the transformation matrix $\mathbf{T}$:

$$\mathbf{A}' = \mathbf{T}^{-1}\mathbf{A}\mathbf{T} \tag{2.22}$$

$$\mathbf{B}' = \mathbf{T}^{-1}\mathbf{B} \tag{2.23}$$

$$\mathbf{C}' = \mathbf{C}\mathbf{T} \tag{2.24}$$

$$\mathbf{D}' = \mathbf{D} \tag{2.25}$$

By which the states of the filters are transformed according to:

$$\mathbf{X}' = \mathbf{T}^{-1}\mathbf{X} \tag{2.26}$$

By these transforms, the transfer function of the filter remains the same (verify by using equation (1.9) and equations (2.22)-(2.25))

$$H'(s) = H(s) \tag{2.27}$$

Thus the transfer function remains the same, as the topology is changed, in order to attain the maximal dynamic range.

2.5.3 Optimization of the dynamic range of filters

Two items are important when optimizing filters: the maximal signal capability and the noise. The problem can be viewed as that of looking through a window. This is depicted in figure 2.13. All windows have a view which extends from left to right; it can be large or small. The total range is determined by the highest noise level of all integrators, and the lowest output level of all integrators. Although all integrators can have a large dynamic range, not all integrators make optimal *use* of it. Better performance results when all the integrators are scaled to each other.

To show the effect of scaling, an implementation was made of a signal-flow diagram, analogous to those of figures 1.13 1.14, of a third-order filter. The $1/s$

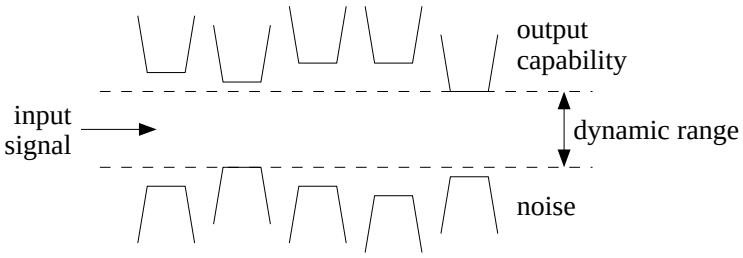

Figure 2.13: Window of several integrators

branches were implemented by means of active integrators like figure 2.11. Simulation results of the output levels of the integrators as a function of frequency are depicted in figure 2.14. It can be seen that the first integrator, magn(f1) in figure 2.14, has to handle larger signals than the second and the third integrator, magn(f2) and magn(f3) in figure 2.14.

By scaling the top levels of the integrator's output levels, the output signals as shown in figure 2.15 results. Now the maximum output level of the integrators are equal and thus all the integrators determine equally the maximum signal that can be handled; all of them are using their signal capability for the large signals completely. This kind of optimization is especially suited for sinusoidal input signals.

Another possibility of scaling is "scaling on the integral of the signal outputs over the total frequency spectrum". This can be viewed as the total signal power level the integrator should be able to handle, which is represented by the total area under the curve of representing the output signal as function of frequency of a single integrator. Such scaling results in the frequency characteristic shown in fig.2.16. This optimization is most suitable for white (noisy) input signals, which is, for example, a good model for a radio input spectrum.

The optimization carried out only considers the output signal handling capabilities of the integrators. This is usually called scaling. Referring to the state-space description, the transfer from the input of the filter to the output of the integrators is equal to, see figure 1.9:

$$\mathbf{F} = \begin{pmatrix} f_1 \\ \vdots \\ f_n \end{pmatrix} = (s\mathbf{I} - \mathbf{A})^{-1}\mathbf{B} \tag{2.28}$$

From this the "controllability matrix", $\mathbf{K}$, can be constructed:

$$\mathbf{K} = \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathbf{F}\mathbf{F}^{*} d\omega \tag{2.29}$$

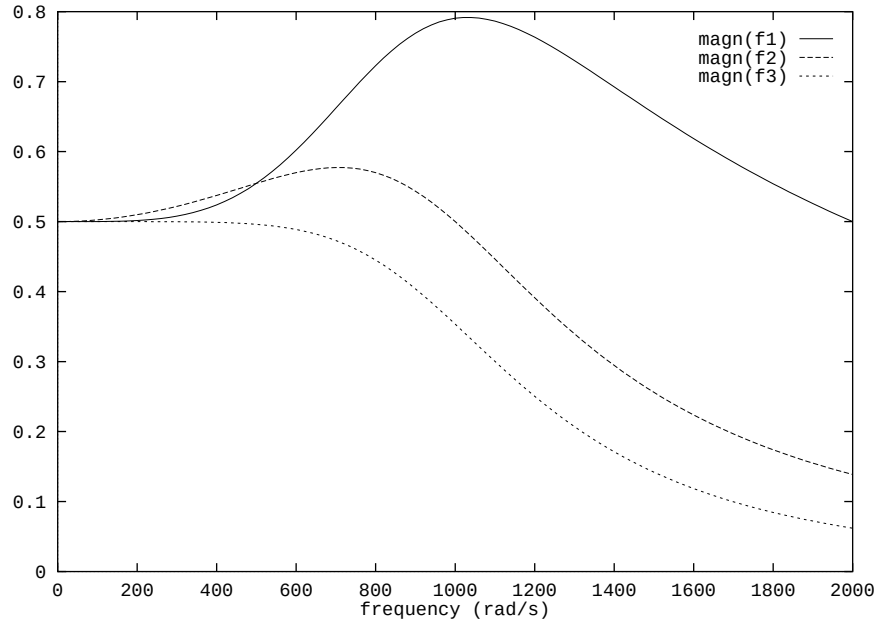

Figure 2.14: Integrator outputs of the unscaled filter

This matrix describes the (power-)transfer from the input of the filter to the output of the integrators.

Scaling does not have any effect on the topology of the filter. This is clear as the scaling action is as shown in figure 2.17. By equalizing the output levels of all integrators, no single integrator will limit the dynamic range at the upper bound of the dynamic range. This is the basic idea behind scaling. It is clear that scaling actually replaces a branch in the filter by another branch with another amplification factor. By means of this amplification factor, the actual integrator is optimally used, with respect to output capability, without changing the topology.

Full optimization –in contrast to scaling– makes use of the noise transfer of the integrators to the output too. Not only is the output capability of the integrators important, but also the noise level. Equalizing the noise levels as well as keeping the output signal capabilities equal makes full optimization feasible. The transfer of the noise sources at the inputs of the integrators to the output of the filter is described as, see again figure 1.9:

$$\mathbf{G} = (g_1 \cdots g_n) = \mathbf{C}(s\mathbf{I} - \mathbf{A})^{-1} \tag{2.30}$$

Now the "observability matrix", $\mathbf{W}$, can be defined as:

$$\mathbf{W} = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathbf{G}\mathbf{G}^* d\omega \tag{2.31}$$

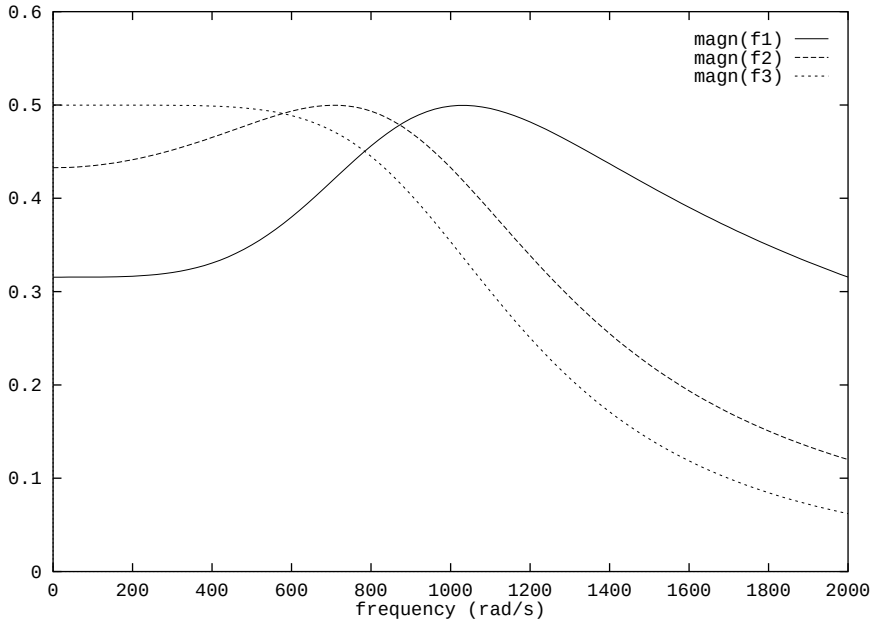

Figure 2.15: Integrator outputs after scaling to the tops

To minimize the total noise of the filter, the main diagonal of the matrix W must be equalized. The main diagonal entries can be viewed as the noise transfers of the inputs of the various integrators to the output of the filter. The difference in the noise of the various integrators also has to be taken into account. This can be done by choosing an ideal capacitance division over the integrators.

Because both the **W** and **K** matrices are used to optimize the filter, the topology changes. The result is usually a filter with non-zero coefficients in every matrix entry. Thus, a fully connected network of integrators evolves. This is most often a structure too large and too difficult to implement on a chip. The dynamic range of this optimal filter is often only used to compare with the actual scaled design, which gives an indication of the quality of the filter structure used, i.e. is there still a lot to be gained or does it not make much sense to optimize my current filter any further. Because a method is known to obtain a topology that actually yields an optimal filter, it is important to know the fundamental limits in advance, so that a filter designer is able to know in advance if it is possible to realize the specifications.

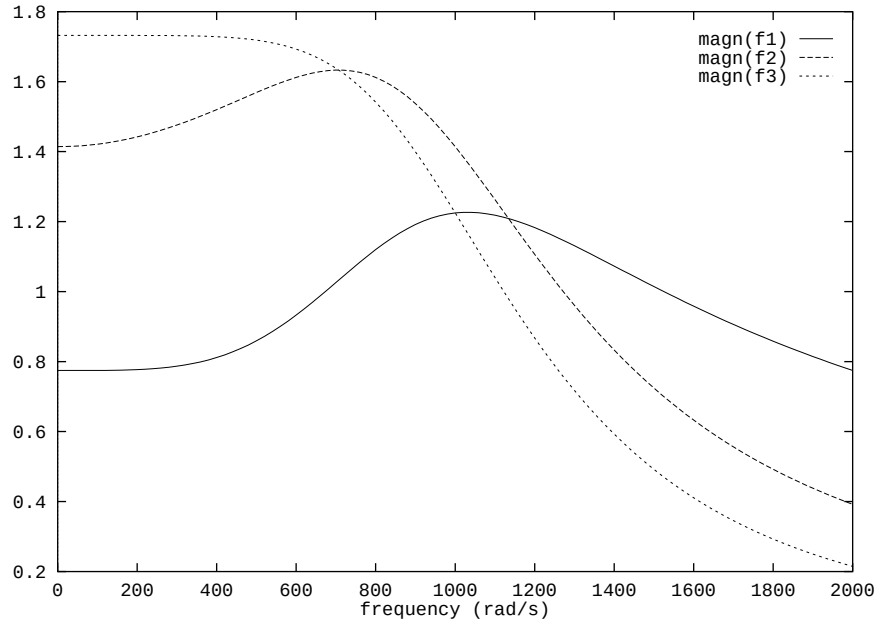

Figure 2.16: Integrator outputs after scaling to the power transfer

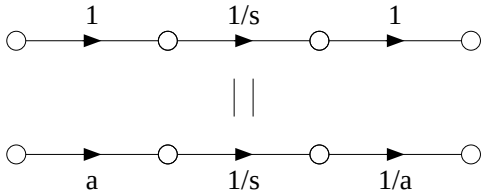

Figure 2.17: Effect of scaling

2.5.4 Fundamental limits

It is possible to derive fundamental limits for the dynamic range of bandpass filters. The dynamic range can be maximally (its derivation is beyond the scope of this book):

$$DR_{opt} = \frac{V_{max}^2 C}{4kT\xi Q} \cdot f(H(j\omega)) \tag{2.32}$$

Q being the quality factor of the factor, approximately equal to the ratio of the bandwidth and the center frequency.

The first part of the expression shows that the output capability and the total capacitance increase the dynamic range when they are enlarged. The noise factor of the active components must be as small as possible. It is also clear that the Q of the filter should be kept as low as possible from dynamic range point of view.

The second part of this equation is only dependent on the transfer function of the filter.

The total minimal power consumption can be derived from this expression. The maximal signal level is equal to V_{max}. Current flows optimally only through the integration capacitors, at a frequency of at most ω_c, the cut-off frequency of the filter. Thus the supply current becomes:

$$I_{sup} = \sum_{i=1}^{n} I_{sup_i} = \sum_{i=1}^{n} \frac{\omega_c C_i V_{max}}{\pi} \tag{2.33}$$

in which I_{sup_1} is the current through integrator i, as the current through the resistors is not taken into account, nor is the surplus current of the biasing for the active circuits.

2.5.5 Example

To show what scaling on tops means for a practical filter design, the filter of figure 1.18 is implemented by means of active integrators. As the transfer of the active integrator of figure 2.11 equals:

$$H(s) = \frac{-1}{sRC} = \frac{-\omega_c}{s} \tag{2.34}$$

the signal-flow graph of figure 1.18 is drawn now with that transfer function for the integrators, i.e. $-\omega_c$ is shifted into the branches representing the integrators, see figure 2.18. Subsequently, this signal-flow diagram is implemented. The bandwidth

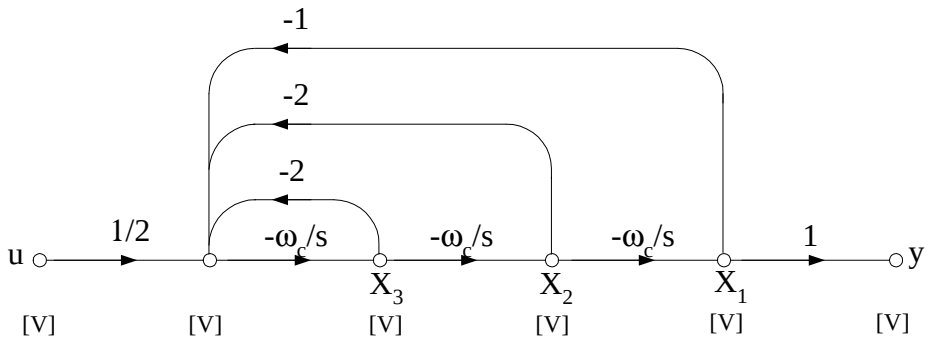

Figure 2.18: Modified signal-flow diagram of figure 1.18, included the dimension of the signals

is chosen to be 1 Hz and for the resistor in the integrator a value of 1 Ω is chosen. For the integrators, active integrators are used as depicted in figure 2.11.

A difficulty arises when we want to implement the scaling blocks. This illustrate in figure 2.18 by the dimension of the signals at the "state nodes", i.e. voltage.

As becomes clear from the picture, the dimension of the scaling blocks is [V/V], i.e. dimensionless. Further, at the input of integrator 3, several voltages need to be added. Adding voltages is relatively complicated when compared with adding currents. By means of some little changes, we can implement current additions. For this, consider a part of the signal-flow diagram of figure 2.18 as depicted in figure 2.19A. First, in the signal-flow diagram the transfer of the active integrator

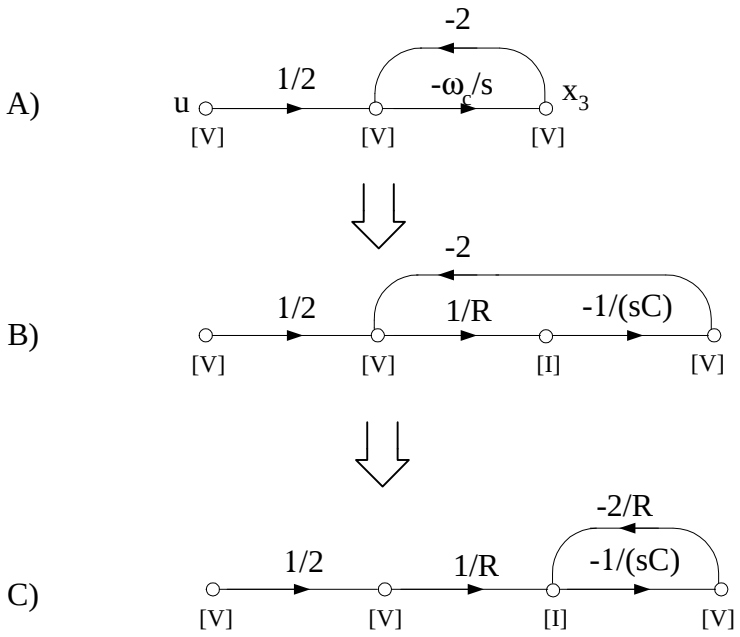

Figure 2.19: Changing from adding voltages to adding currents

is expanded in a transfer of the conductance (resistor) with transfer $1/R$ and of the active capacitor $1/(sC)$, see figure 2.19B. As the conductance $1/R$ is a linear transfer, the addition point can be shifted from the input of this transfer to the output, which has a dimension of a current [A]. Of course, the transfer of "-2" has to be changed into "-2/R" to have again same transfer between the states. Using this adaption, the positive scaling blocks can be implemented by resistors only. For a negative gain block a inverter is required. The resulting implementation is given in figure 2.20. The corresponding output voltages of the integrators are depicted in figure 2.21 for a input voltage of 2 V. From this figure it easily follows that the maxima of X_2 and X_3 are both about 0.73 V whereas the maximum in X_1 is 1 V. Thus X_2 and X_3 have to be scaled by 1.37. This means that the scaling as depicted in figure 2.17 has to be applied to the corresponding branches. This is illustrated in figure 2.22A. In subsequent shifts, the additional scaling factors are shifted to locations where they can easily be implemented, i.e. at the gain blocks or at the inputs of the integrators (just changing a resistors value). The

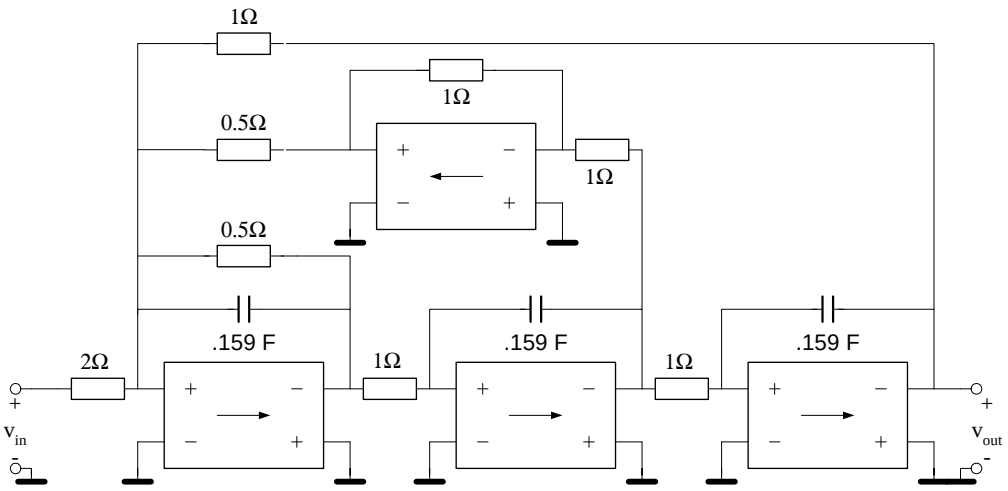

Figure 2.20: Implementation of the filter of figure 2.18, $f_c = 1$ Hz

implementation corresponding with figure 2.22D is depicted in figure 2.23. The corresponding output signals of the three integrators are depicted in figure 2.24. Clearly, for all the integrators, the maximum output signal (for 2 V input signal for the filter) is now 1 V. The only measure that had to be taken was changing some resistor values. As was stated in a previous section, the noise of the integrators is completely determined by the capacitor values used. Therefore, changing resistor values does not change the noise level and thus an increased dynamic range is obtained.

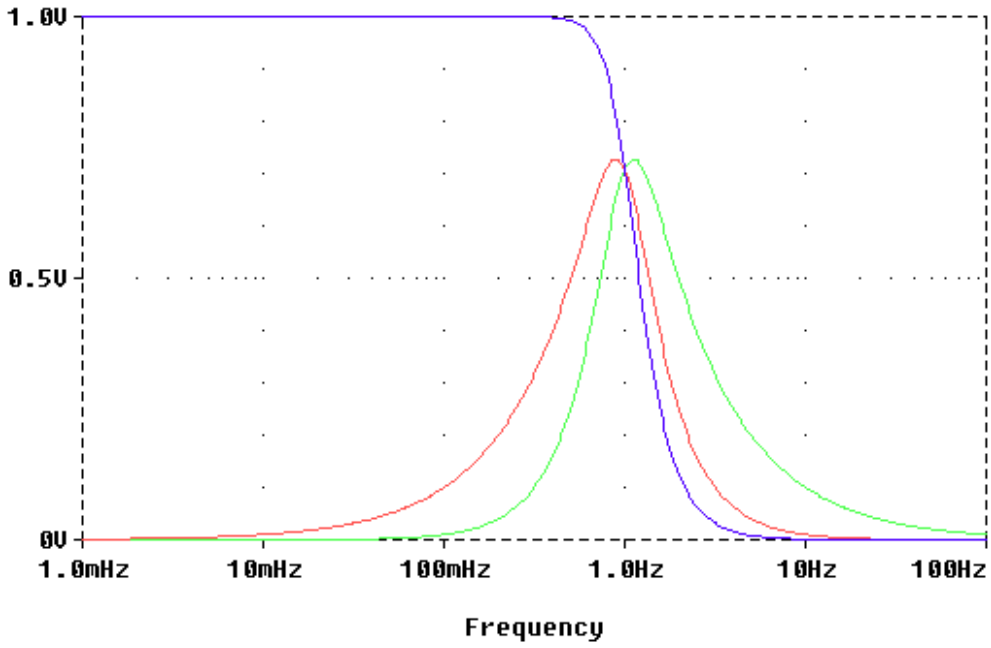

Figure 2.21: Integrator output voltages of the unscaled filter, input voltage is 2 V. From left to right X_1, X_2 and X_3.

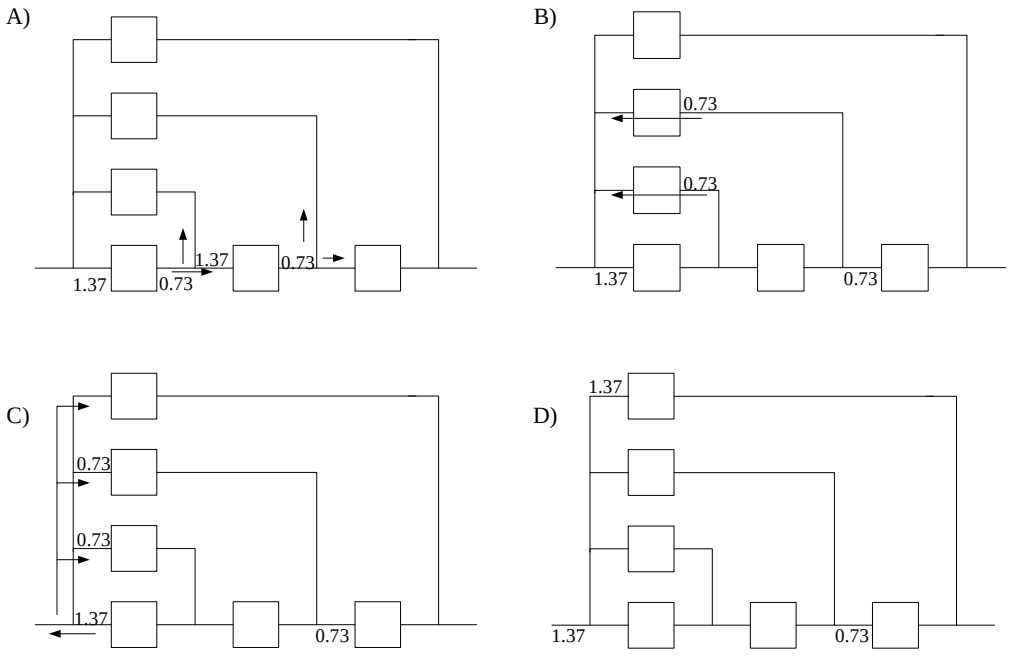

Figure 2.22: Scaling the filter of figure 2.20 to obtain for each of the integrators maximum output levels of 1 V

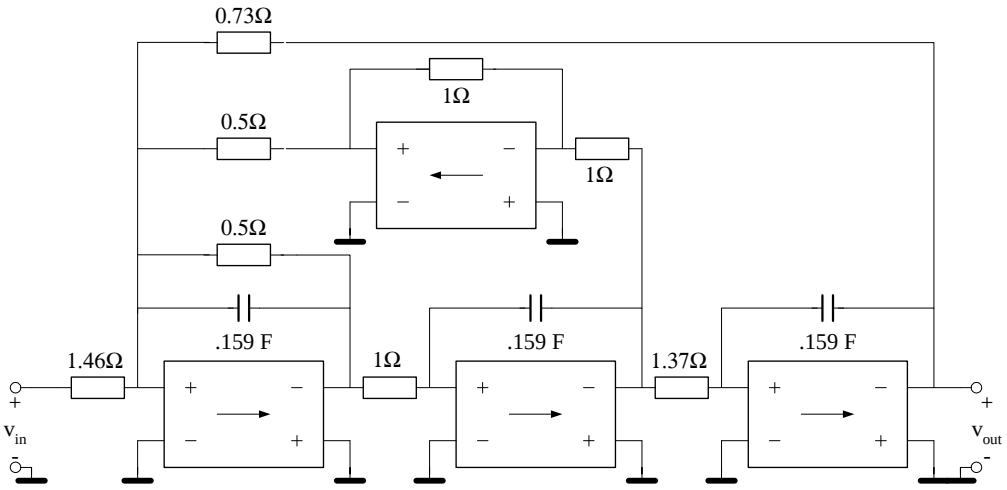

Figure 2.23: Implementation of the scaled filter of figure 2.22D

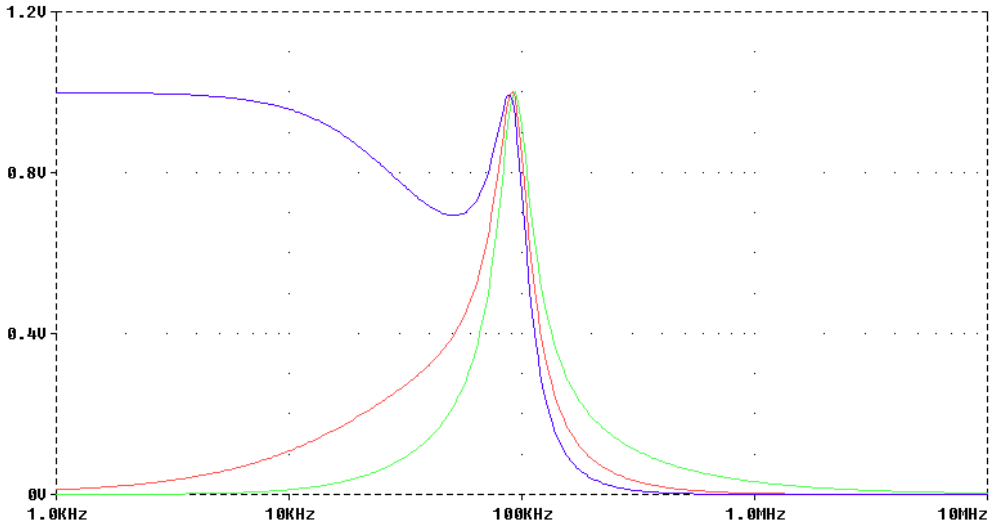

Figure 2.24: Integrator output voltages of the scaled filter, input voltage is 2 V. From left to right X_1, X_2 and X_3.

2.6 Exercises

1. Long Wave (LW) transmission is still very popular in the UK. The frequency range of the LW is from 153 kHz to 279 kHz. For the transmitted signals Amplitude Modulation (AM) is used. Figure 2.25 shows a block diagram of a LW receiver. At the input the signal is received by the antenna. A preselection filter

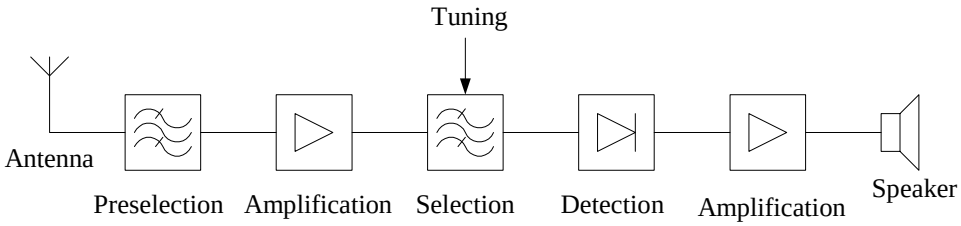

Figure 2.25: A block diagram of an LW receiver.

(bandpass filter) rejects all the signals outside the LW band, for instance due to mobile phones or other radio transmitters. The remaining signals of the LW band are amplified for more robust processing. Subsequently, a selection filter selects, by means of band-pass filtering, a single channel with a bandwidth of 9 kHz. To be able to tune to different stations, this filter is made tunable. After selection of a single channel, the AM signal is demodulated. Finally, the signal is amplified such that it can be made audible via a speaker.

(a) What class of filter is most suitable for the *preselection* filter? Motivate your answer.

(b) What class of filter is most suitable for the *selection* filter? Again, motivate your choice.

(c) Which of the two filter is likely to determine the overall dynamic range of the receiver. Discuss both aspects of dynamic range: noise and distortion.

2. For the LW receiver of figure 2.25, the selection filter was designed in a 0.35 μm CMOS technology. The supply voltage for that technology is 5 V. The filter has a dynamic range of 70 dB and occupies 1 mm^2 of chip area. The filter capacitances are responsible for 75% of this area. The capacitors are made by the gate-capacitances of the MOS transistors.

For the new generation LW receivers a redesign is required in a 0.25 μm CMOS technology. The corresponding supply voltage is 3.3 V. The gate oxide in this process is 30% thinner. The new filter should have a dynamic range of 85 dB.

(a) What will be the size of the redesign in this new technology when one can assume that the active part of the receiver hardly changes in size.

3. Given the cascade of electronic blocks as depicted in figure 2.26. Each block

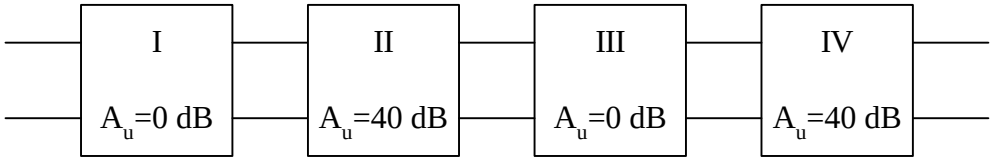

Figure 2.26: A cascade of electronic blocks.

has only got a voltage-to-voltage transfer (i.e. A≠0 and B=C=D=0) and there respective equivalent input noise voltage and maximum output voltage are listed in table 2.6.

Table 2.1: The respective equivalent input noise voltage and the maximum output voltage for the blocks of figure 2.26

Block	$v_{noise,in,eq}$	$v_{out,max}$
I	1 μV	10 V
II	10 nV	10 V
III	10 μV	1 V
IV	10 μV	10 V

(a) Determine the dynamic range of the chain of blocks.

4. Given the two possible representations of the noise of a resistor in figure 2.27.

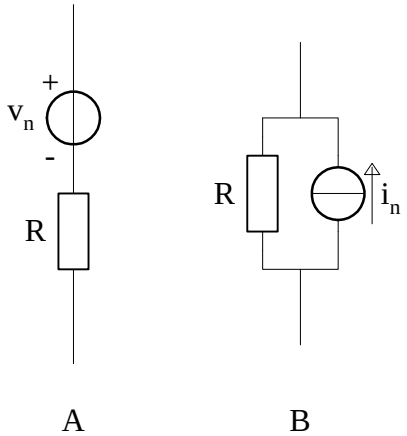

Figure 2.27: Modelling the noise of a resistor. A) with an equivalent noise voltage source and B) with an equivalent noise current source.

(a) What are the power spectral densities of the two noise sources in figure 2.27?

To determine the equivalent noise source of a larger network several transformations can be used. Given a "larger" network as depicted in figure 2.28.

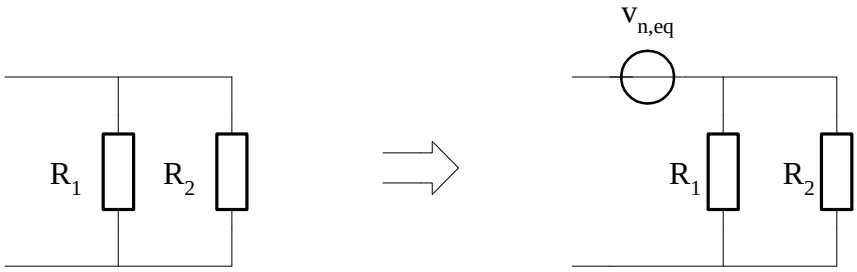

Figure 2.28: two parallel connected resistors with a model for the equivalent noise source.

(b) What four transformations do you know?

(c) Identify in the left part of figure 2.28 the relevant noise sources.

(d) Give an expression for the equivalent noise voltage in terms of the noise sources of R_1 and R_2.

(e) Finally, give the expression for the power spectral density of the equivalent source.

5. Given the network of figure 2.29 which is considered as a one-port . Source

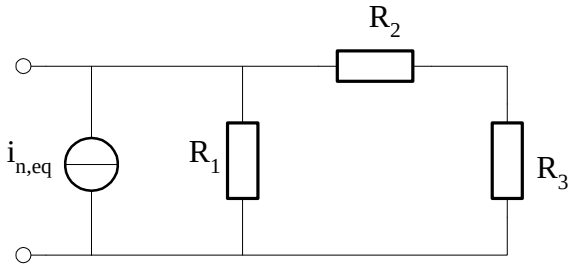

Figure 2.29: A three resistor one-port.

$i_{n,eq}$ is the equivalent noise source of this one-port.

(a) Determine the spectral power density of the equivalent noise source.

Subsequently, the one-port is changed by adding either a parallel capacitance or a series capacitance, see figure 2.30.

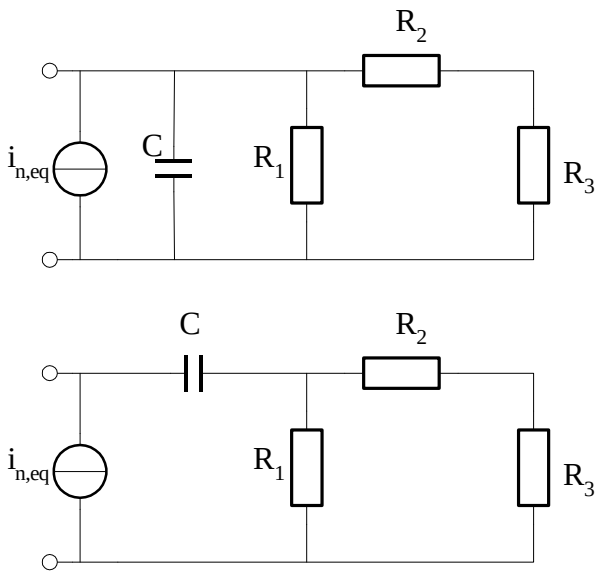

Figure 2.30: Adding either a parallel or a series capacitor to the one-port.

(b) Describe in a qualitative way how the power spectral densities of each of these two situations relate to the power spectral density of question (a).

6. In figure 2.31 a signal model of a bipolar transistor is depicted including the relevant noise sources. The noise source $i_{n,c}$ is due to shot noise on the collector

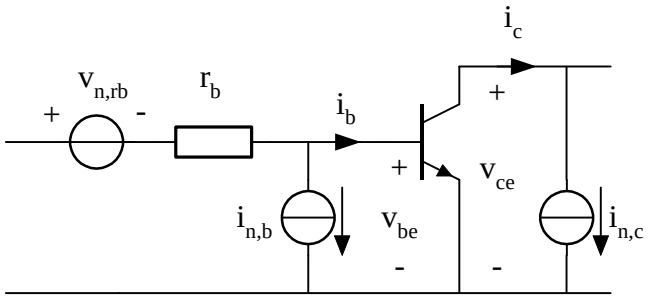

Figure 2.31: A signal model of a bipolar transistor including the relevant noise sources.

current whereas noise source $i_{n,b}$ is due to the shot noise of the base current. The base resistance introduces thermal noise. The corresponding power spectral densities are listed in table 2.2. A simplified chain matrix of the bipolar transistor is given by:

$$\begin{pmatrix} v_{be} \\ i_b \end{pmatrix} = \begin{pmatrix} 0 & \frac{-1}{g_m} \\ 0 & \frac{-1}{\beta} \end{pmatrix} \begin{pmatrix} v_{ce} \\ i_c \end{pmatrix} \tag{2.35}$$

Noise source	S
$i_{n,c}$	$2qI_c$
$i_{n,b}$	$2qI_b$
$v_{n,rb}$	$4kTr_b$

Table 2.2: The power spectral densities of the relevant noise sources of a bipolar transistor.

(a) determine the equivalent noise sources at the input of the transistor, v_n and i_n.

(b) Determine the power spectral density of both sources.

7. Given the two amplifiers in figure 2.32.

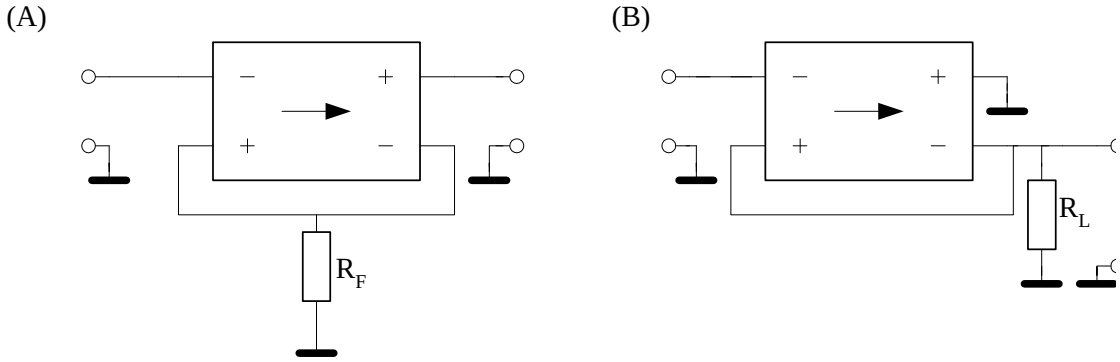

Figure 2.32: Two feedback amplifiers.

(a) Determine for amplifier A) the equivalent input noise source(s) and the corresponding power-density spectrum assuming that only R_F contributes to the noise.

(b) Determine for amplifier B) the equivalent input noise source(s) and the corresponding power-density spectrum assuming that only R_L contributes to the noise.

8. Given the voltage amplifier of figure 2.33.

(a) Identify in the circuit diagram the noise sources. The active part can be assumed to be noise free.

(b) Transform the noise sources to an equivalent noise source at the input.

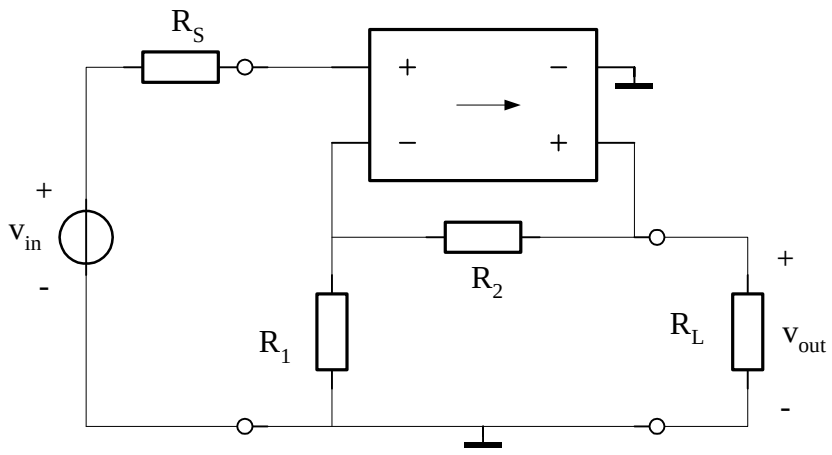

Figure 2.33: A voltage amplifier.

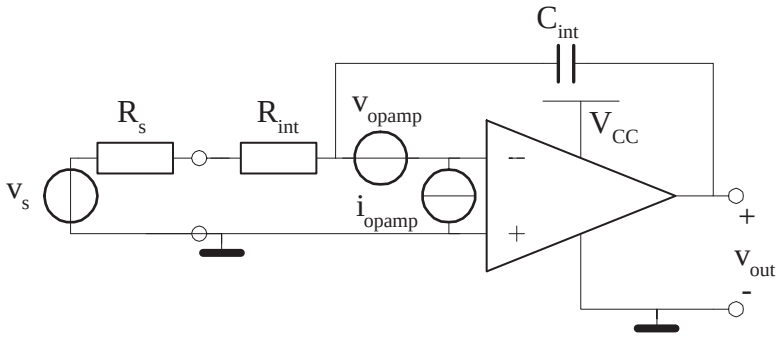

Figure 2.34: An active integrator

(c) Determine the power spectral density of the equivalent source.

9. Given the circuit of an active integrator in figure 2.34. The current source, i_{opamp}, and the voltage source, v_{opamp}, can be used to model the noise contribution of the opamp. Assume for this exercise that $v_{opamp} = 0$.

(a) Determine for the integrator the equivalent input noise voltage, taking into account the noise of the opamp, i_{opamp}, and the noise of the resistors (see figure 2.35.

(b) Determine the power-density spectrum of the equivalent source. Assume that the power-density spectrum of i_{opamp} equals $S_{i_{opamp}}$.

10. In figure 2.36 the block diagram off an all-pass filter is depicted. For an all-

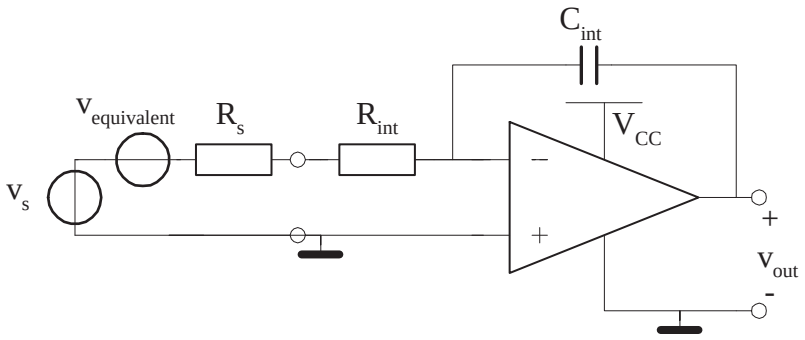

Figure 2.35: The location of the equivalent source

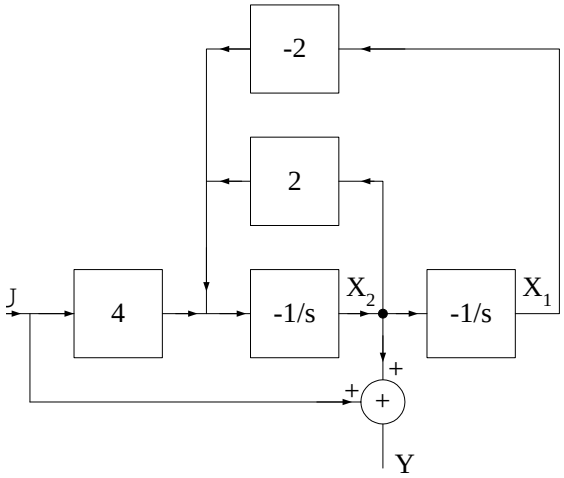

Figure 2.36: Block diagram of an all pass filter

pass filter the magnitude of the transfer is flat whereas the phase of the transfer is not flat (for this filter the phase variation over the frequency is 360 degrees).

(a) Apply power scaling to this filter when it is given that $f_1 f_1^*$ and $f_2 f_2^*$ of the controllability Gramian matrix are 6.25 and 12.575, respectively.

11. Consider the implementation of a third-order Chebyshev filter of figure 2.37. The output voltages of the integrators as a function of the frequency are depicted in figure 2.38.

(a) Apply scaling on this filter such that the maxima of the integrator outputs become equals. Give clearly the different steps.

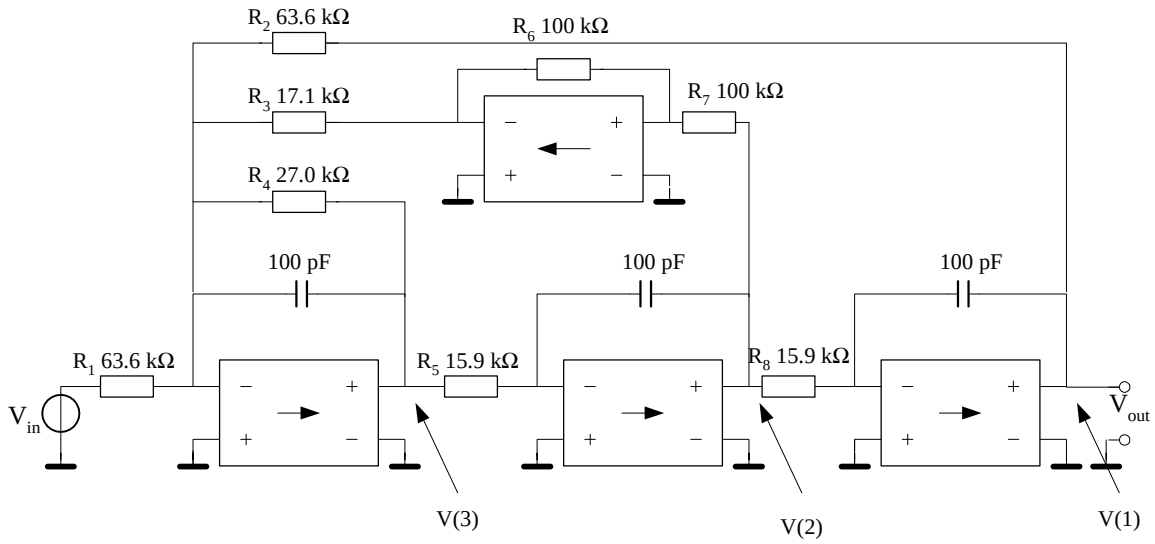

Figure 2.37: Implementation of a third-order Chebyshev filter

12. Given the filter transfer function according to (compare with the corresponding exercise on page 25):

$$H(s) = \frac{s^2}{s^2 + 1.4 \cdot 10^7 s + 10^7} \tag{2.36}$$

This filter function is to be implemented by means of resistors, capacitors and opamps. For the opamp may be assumed that their bandwidth is infinite, the noise is zero and their supply voltage is 15 V. The *integrators* in the filter should minimally have a dynamic range of 120 dB.

(a) Give a circuit diagram of the filter implementation.

(b) Determine the values of the resistors and capacitors using the required filter function and the minimum dynamic range for the integrators.

13. Given the filter transfer function (compare with the corresponding exercise on page 25)

$$H(s) = \frac{(s^2 - s + 1)(s - 1)}{(s^2 + s + 1)(s + 1)} \tag{2.37}$$

(a) Draw an implementation of this filter by using capacitors, resistors and nullors.

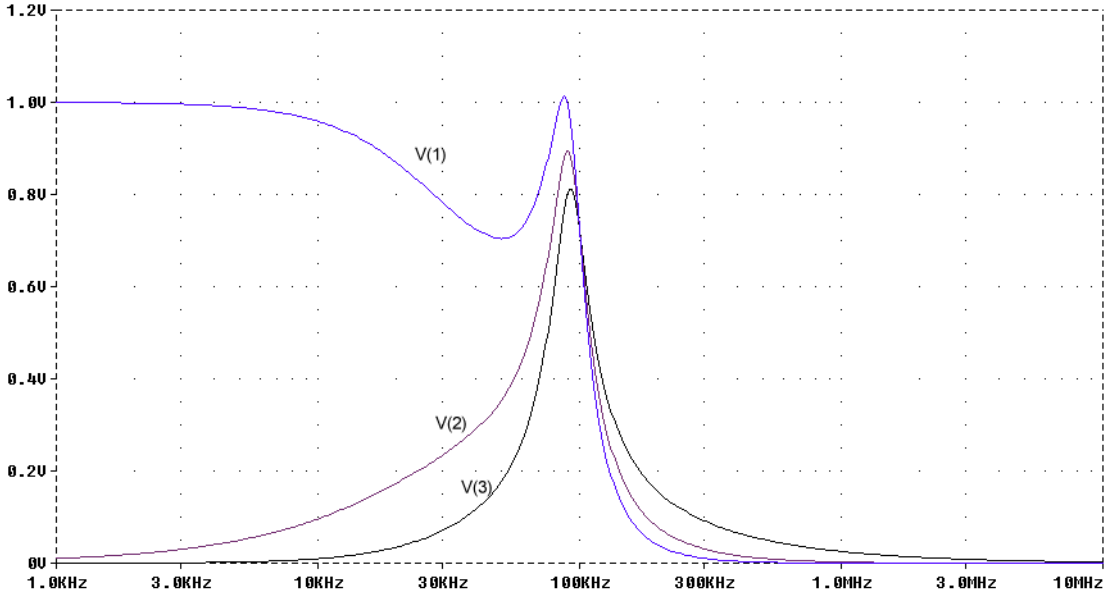

Figure 2.38: The integrator-output voltages as a function of the frequency

Chapter 3

Filters and sensitivity to component spread

3.1 Introduction

In the previous chapters it was shown how to design, starting with a transfer function (differential equation), a filter topology realizing the specified transfer function. Subsequently, implementing the topology by means of resistors, capacitors and nullors, was shown to be a straightforward step. In the example shown, the capacitor values where specified with a precision of four digits. When looking to filter tables specifying for a given topology the required components values, even higher precision are used even up to 10 digits! An example of a passive filter found from a standard topology and a table with component values is given in figure 3.1. "Building" such filters in simulators will work fine; the expected filter

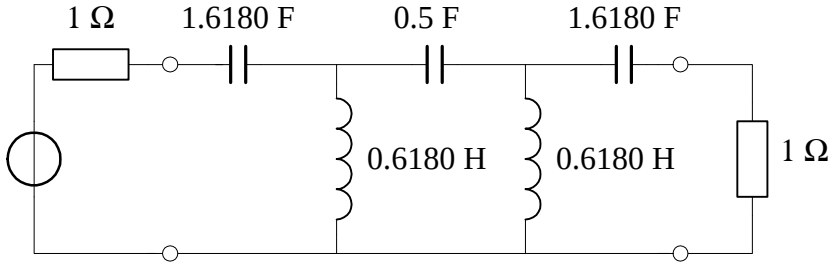

Figure 3.1: A passive filter obtained from a standard topology via tables with component values. Component values are specified by 5 digits.

function is obtained accurately. However, when realizing filters physically, for instance on a chip, one is limited to a certain accuracy. Trying to get an absolute accuracy of 5 digits is not easy to do. Spread on component values can be in the order of 10%. Designing accurate filters with those inaccurate components is

a challenge, especially when you consider that filters comprise *different* types of components, each having their own specific inaccuracy and spread. For instance, for a first-order low-pass filter realized by a capacitor, (C), and a resistor, (R), the corner frequency (f_C) is defined by:

$$f_c = \frac{1}{2\pi RC} \tag{3.1}$$

When for both, the resistor and the capacitor, the spread on their values is 10%, the worst-case error in the corner frequency is 20% !.

Still, designers are able to design very accurate on-chip filters. In this chapter techniques will be shown that enable designers to make accurate filters by using relative inaccurate components. First, some basic reasons for the existence of inaccuracy are treated and how designers have to take that inaccuracy into account during their design. Subsequently, with some more detail the spread in the value of on-chip resistors is discussed including some guidelines to obtain the accuracy that is possible within the used technology. When the required accuracy of a filter is higher than the intrinsic accuracy of the comprising components, system level measures need to be taken: trimming and tuning. In the previous chapters we concluded that different topologies, realizing the same filter function, may have different dynamic ranges. The same goes for the sensitivity of a implemented filter transfer function for component values. It appears that different topologies have different sensitivities for variations in component values. This will be addresses at the end of the chapter and design rules are derived for obtained low-sensitive filter.

3.2 Accuracy and models

For designing accurate circuits, two possible sources of errors should be distinguished:

- too simple models;
- spread on the model parameters.

These are discussed in the next sections.

3.2.1 Model complexity

Models are an abstraction of the real life world. The complex physical effects present in the physical components are described by relative simple relations such

that by using these relative simple expressions the behavior of the physical component can be predicted. With a good model, the relevant physical effects can be predicated relatively easily.

Depending on the context of the model, a good model can be very simple but sometime more details and relations are required in the model. When a model is used that is too simple, given a particular context, for instance when a relevant physical effect is not modelled, the outcome of the prediction will be incomplete and therefore the final design may have a unexpected behavior.

Consider, for instance, the two models depicted in figure 3.2. Figure 3.2A

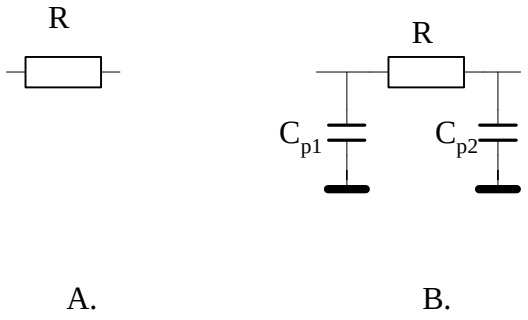

Figure 3.2: Two models for a resistor.

models the basic behavior of the resistor, i.e. Ohms law. For a lot of situations this model is accurate enough (luckily). However, when signal frequencies go up, at a certain point this model will be incomplete. The physical resistor is embedded in some environment, i.e. in the bulk of a chip or on a PCB, and has a capacitive coupling to that environment. This can be modelled, for instance, as indicated by the two additional capacitors in the model of figure 3.2B. When in a design signals are expected requiring the inclusion of these capacitors in the model but the simple model of figure 3.2A is used, unexpected frequency dependency may be obtained. The accuracy of the physical circuit is limited then also by this lack of modelling.

The other way around, the use of a too complex model will cause unnecessary design complexity which may impede the designer from focussing on the relevant and dominant effects determining the accuracy. So care should be taken in determining the simplicity (complexity) of the models to be used in a design trajectory.

3.2.2 Model parameter spread

When the models are accurately modelling the relevant physical effects, then the spread on the model parameters still may cause errors and inaccuracy in a design. Roughly speaking, the spread in the parameters are mainly caused by the limited

Min (Ω)	Nom (Ω)	Max (Ω)
900	1000	1100

Table 3.1: Specification of nominal sheet resistance including the deviations.

accuracy of the production process of the physical realization. For instance, the resistance of a resistor is determined, amongst other things, by its geometrical aspects. On a chip where resistor dimensions can be as low as several μm, it is not difficult to imagine that the final dimensions deviate from the intended ones when taken into account that dimensions are determined via lithographic and diffusion processes.

For reasonable spread on parameters, the consequences on the design are likely to be just quantitative deviations from the expected behavior. This in contrast when too simple models are being used, in which case unexpected phenomena may arise and thus also a qualitative deviation may result.

To be able to design circuits which are capable of surviving parameter spread, the characterization of the statistical information of those parameters is essential. Then, via design measures and simulations the effect of the real life parameter spread can be investigated and, if necessary, countermeasures can be taken.

For instance, in a design manual of a process the sheet resistance of a layer in which a resistor can be realized can be specified as given in table 3.1.

3.3 Resistors

A resistor is a device that defines a linear relation between a current and a voltage. In the ideal case the relation is:

$$u = Ri \tag{3.2}$$

corresponding to the model in figure 3.2A. In equation (3.2) R is a constant. Thus, R does not depend on time, frequency, temperature, and the value of u and i. In practice, R, unfortunately, also depends on all these factors:

$$u = R(t, \omega, T, u, i)i \tag{3.3}$$

In addition, the relation is also subject to statistical variations due to, for example, process variations. Thus, the behavior of a resistor can be predicted with limited accuracy only.

When an electronic circuit is being designed, usually, first the ideal model according to equation (3.2) is used. In such a case, the fundamental limits to the performance of the circuit can be found. This performance is compared to the specifications. Then it becomes possible to estimate the magnitude of the errors

that can be allowed within the demands of the specifications. From this, the design constraints for the devices can be derived when equation (3.3) is used.

In this section focus is on the physical aspect of a on-chip resistor in relation to model complexity and parameter spread.

3.3.1 The value of a resistor

A resistor consists of a "bar" of material with the resistivity ρ and connections to both sides (see figure 3.3). The value of the resistor is determined by ρ, the

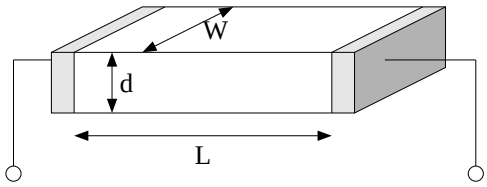

Figure 3.3: A resistor

thickness (d), the length (L) and the width (W).

$$R = \rho \frac{L}{d \cdot W} \tag{3.4}$$

Because, in practice, d cannot be influenced by the designer, only L and W can, ρ and d are usually taken together in a parameter called the *sheet resistance.*

$$R_{\square} = \frac{\rho}{d} \tag{3.5}$$

With the use of this parameter, the layout and value of a resistor is commonly expressed in squares as depicted in figure 3.4. When the number of squares in a

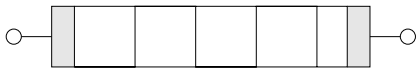

Figure 3.4: A resistor of 4.5 squares

resistor is equal to:

$$N = \frac{L}{W} \tag{3.6}$$

its value can be expressed as:

$$R = NR_{\square} \tag{3.7}$$

From equation (3.7) it can be seen that it is the number of squares that determines the value of the resistor and not the absolute size of the squares. The maximum

current that the resistor can support, its accuracy and its bandwidth, however, do depend on the size of the squares. This is discussed later.
In each technology, there are several layer that are, in principle, suitable for the implementation of a resistor. For each layer, a different number of squares may be necessary to obtain the same resistor value. In table 3.2, for a standard bipolar process, the number of squares necessary to implement a resistor of 1kΩ is indicated.

Name	$R_{\square}$ (Ω/□)	N
epi	2300	0.43
BW	600	1.67
WP	25	40
metal	0.044	22727

Table 3.2: The number of squares necessary to implement a 1kΩ resistor

Suppose the resistor is implemented with the lowly doped P-type layer BW. In figure 3.5, a cross-section of this resistor is shown. The BW layer cannot be

Figure 3.5: Cross-section of a BW-resistor

directly connected from the outside. In order to obtain a good contact between the metal interconnect layer (IC) and the silicon, a highly doped P layer (WP) is used to interface between IC and BW. Therefore, the resistor actually consists of a series connection of several resistors, as is shown in figure 3.6. From the left to the right there are:

- The resistance of the metal (IC) from the connection point to the silicon (WP).
  In table 3.2, it can be seen that the sheet resistance of the metal is much lower than that of the BW layer. Therefore, generally, the resistance of the interconnect does not play a significant role.

- The contact resistance between the metal and the silicon.
  The mask that defines the size of the contact opening between IC and WP is the CO mask. The current through the contact opening is a vertical current,

as opposed to all other currents in the resistor, which are lateral. Therefore, the resistance of the contact (CO) is not expressed in a sheet resistance, but as a resistance per area. In the standard process used as an example in this book, the resistance of a contact with a size of $2 \times 2\mu$ is about 4Ω. The larger the area is, the smaller the contact resistance will be. When the contact area is enlarged, however, care must be taken that the current flows homogeneously through the contact. If not, the contact resistance becomes larger than expected, and the predictability decreases. When this "current crowding" starts, it is of no use to enlarge the contacts any further.

- The resistance of the WP region.
  The sheet resistance of the WP layer is smaller than that of the BW layer, but not small enough to be neglected. Therefore, the number of WP squares in series with the BW resistance should be kept as small as possible.

- The resistance of the BW region.
  This region forms the actual resistor. The number of squares in this layer dominantly determines the resistor vale. Commonly, this part of the resistor occupies most of the area.

- Again the resistance of another WP region, a contact and the interconnect.

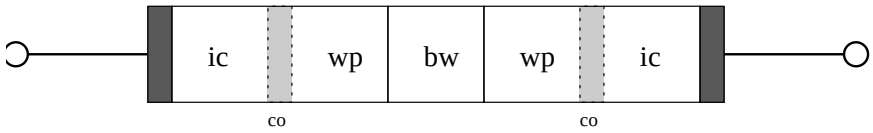

Figure 3.6: The resistor chain of which an integrated resistor consists

From all this, it can be concluded that equation (3.7) is a very simple expression to design an accurate resistor. This equation yields only the resistance of the BW layer. Therefore, the contribution of the other resistors is added via the contact resistance R_c:

$$R = 2R_c + NR_{\square} \tag{3.8}$$

It is difficult to calculate the exact value of R_c. However, the order of magnitude in which its contribution lies can easily be estimated. To obtain accuracy, resistors with various BW layer lengths have to be implemented and measured. The value of R_c can be extracted from this data.
The value of R_c can also be determined numerically by a device simulator, but to still supply this simulator with the correct model parameters, accurate measurements have to be done at least once. The results of a simulator are never more accurate than the measurements that were used to generate the model parameters.

3.3.2 The structure of a resistor

In figure 3.7 a cross-section of a resistor in a standard process is shown. It can be seen that the resistor has *four* connections. Two of them are the usual terminals between the desired resistance exists. Between these terminals, the layers that were shown earlier in figure 3.5 (IC–WP–BW–WP–IC) are found. The contact between the metal (IC) and the silicon (WP) is made via a contact hole. The place and the size of the contact is determined by the (CO) mask.

The resistor is surrounded by the epi well. To isolate this well from the resistor, the WP–epi and the BW–epi junctions have to be reverse biased. Because in this case the resistor body is P type and the eip is N type, the epi connection should be at a voltage that is higher than the highest voltage that can be expected on the resistor body. The positive supply voltage is a safe choice in this respect. For the connection to the epi layer, a standard NPN-collector contact is used. The buried layer (BN) is also present, because also in this case there is the risk of latch-up via the parasitic PNP transistor (BW-epi(+BN)-Subtrate = PNP).

The epi well is isolated from the rest of the chip via a DP (Deep P) ring around it, and the P substrate at the bottom. To keep the junctions involved reverse biased, the substrate and the ring should be at a sufficiently negative voltage; the negative supply voltage or, if not present, ground is the common choice.

There is no need for a separate epi and substrate connection for each resistor. The substrate connection is inevitably a global one, but to have more than one resistor in the epi well is also permissible. They are isolated from each other anyway, because their junctions to the epi layer are reverse biased. This also saves space, because the DP isolation takes much space in comparison to the resistor body itself. Still, for each resistor, the epi voltage should be considered with care.

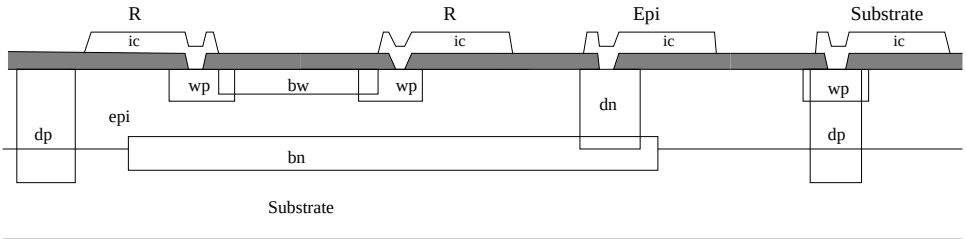

Figure 3.7: Cross-section of a BW resistor

Certainly, in circuits that have more than one supply voltage, it might happen that a resistor is driven at a voltage that is higher than the voltage of its well. Then the junctions between the resistor body and the epi layer become forward biased and the epi layer will "try to follow" the signal voltage at the resistor terminal. Also at that moment, the parasitic PNP starts injecting current into the substrate, since the junction is now biased in forward mode from its base-emitter junction.

In figure 3.8, a more complex model, including the parasitic PNP transistors, for the resistor of figure 3.7 is shown.

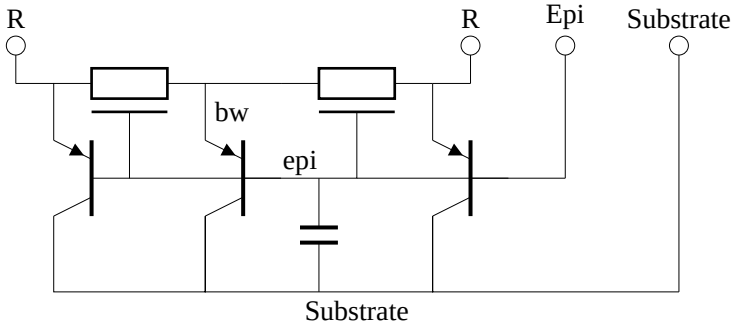

Figure 3.8: A model for a diffused resistor including a the relation to the epi and substrate.

3.3.3 Design aspects

It is common practice in circuit design to start with models including only the basic behavior of the components to get an estimate, without a need of excessive design time, of the performance of the design. Often, when more effects are modelled the corresponding performance lowers. Thus, the estimate done with the simple models can be seen as an upper limit of the performance. This can be very helpful in valuing basic circuit concepts.

When with simple models the performance is good enough two different strategies can be used in dealing with additional model refinements.

- accept the refinements as they are and include them in the design;
- try to find design measures such that the model refinements can be neglected.

The first approach just accepts the refinements and includes them into the design. This results in a more complex design but more important is, also the performance may be reduced by these additional effects. An advantage can be that the final design is more compact.

For the second approach the argumentation is: with the simple models I have a circuit that meets the specification. Thus, now I should try to make those simple models valid (again), i.e. take design measures such that additional effects become negligible. Advantage of this approach is that the design remains relatively simple when the models are considered and as less effects are hampering the performance, higher performances are likely to be obtained. Disadvantage is that due to the additional design measures a less compact design is obtained. However, as

compactness is often not a main issue whereas the quality is, designing via this approach may yield better results. On top of that, as the models remain simple, also better insight is obtained in the design which may open possibilities to complete new designs.

A straightforward example of this second approach is given in figure 3.9. Figure

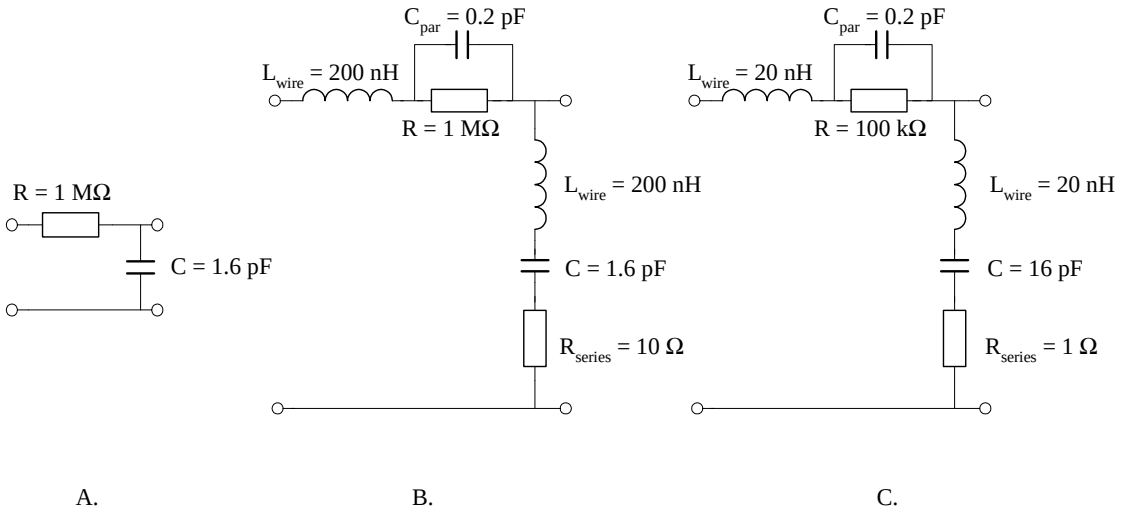

Figure 3.9: A. Simple model for estimation. B. Refined model and relatively high impedance level. C. Refined model and relatively low impedance level.

3.9A depicts a first-order low-pass filter in which the most simple models are used for the comprising elements. The transfer from input voltage to output voltage as a function of frequency is found in figure 3.10, the lowest curve. When the filter is implemented by means of discrete components, wiring inductances, series resistances and parallel capacitances are found. When these parasitic effects are also taken into account, the circuit model of figure 3.9B is obtained. The simulation results of the voltage-voltage transfer of this filter is depicted in 3.10, the upper curve. Clearly, serious deviations arise beyond 1 MHz. The relative effect of the parasitic components should be made less in order to approximate better the ideal transfer. This is realized, for this example, by reducing the wire lengths of the components by a factor ten and lowering the impedance level of the filter resistor and capacitor by a factor 10 also. This is depicted in figure 3.9C. The corresponding simulation result is the middle curve in figure 3.10. Now the transfer starts deviating seriously from the ideal transfer beyond 10 MHz. So, in this case, by simple and trivial measures a decade is gained in the usable frequency range of the filter. Of course, for designing integrated filters, the required measures to be taken for reducing the effect of the parasitic components can be less trivial.

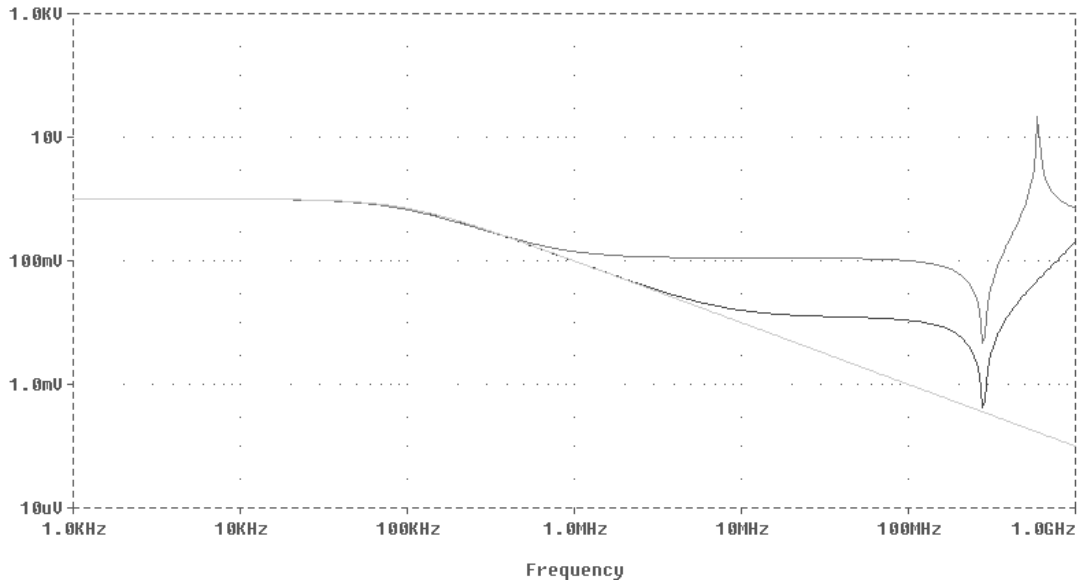

Figure 3.10: Simulation results of different levels of complexity for a first-order filter. From bottom to top: using simplest models; using complex models and low impedance level, using complex models and high impedance level.

3.4 Tuning

The function of filters is to separate signals on the basis of frequencies. The frequencies of interest have to be passed through the filter, as other frequencies are attenuated. The specifications are defined by the system in which the filters are used. As the components in circuits can, for example, have a 10 % tolerance, filters have to be *trimmed* during the fabrication. Often, also filters have to be *tuned* over a certain frequency range, for example, in a radio receiver. The above two comments imply that tuning (trimming) to some reference frequency is desirable. Here focus will be on tuning, in stead of on trimming.

In order to accomplish tuning, a reference frequency has to be available, and the filter must have the possibility of being tuned. This is depicted schematically in figure 3.11. The frequency response, $\tau(x)$, is changed relatively based on absolute frequency information ,(x), from a reference, τ_{ref}.

Filters can be tuned in two ways. Firstly, by varying the capacitance in the integrators. This is possible when using junction capacitances, but this causes problems, because junction capacitances have a polarity, they usually cannot be used floating and they are strongly non-linear with voltage. Secondly, the transfers G_x can be used to tune the filters. In practice, this mostly means that resistors are varied. Examples are MOS transistors in triode, which can be tuned by varying the gate voltage.

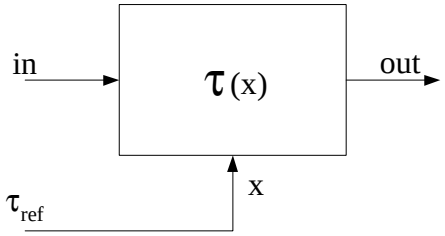

Figure 3.11: Tuning a filter

There are several tuning options, two of which are discussed below. Both methods rely on the matching of components, which determine the time constants of the filter. A Voltage Controlled Oscillator (VCO) or a reference filter can be used for tuning. Both these methods are depicted in figure 3.12.

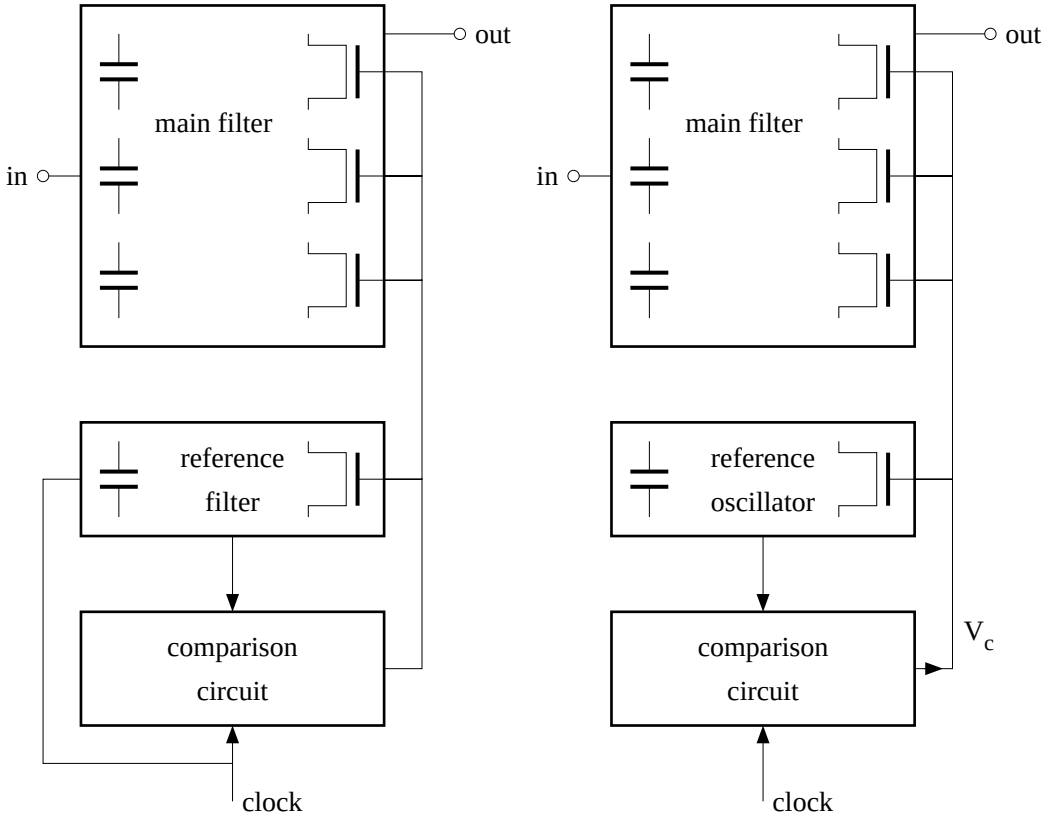

Figure 3.12: Two methods of tuning filters

In both cases, the same components are used in the VCO or the reference filter as in the desired filter. If, for example, an eighth-order bandpass filter has to be tuned, a second-order reference filter can be chosen, with a center frequency equal

to the filter to be tuned. It is also possible to use a VCO (undamped filter) that oscillates at the frequency the filter should be tuned at. By means of a feedback loop, the VCO or reference filter can be tuned. Because of the matching to the VCO or the reference filter, the main filter is tuned, too. The two possible feedback loops are shown in Fig. 3.13.

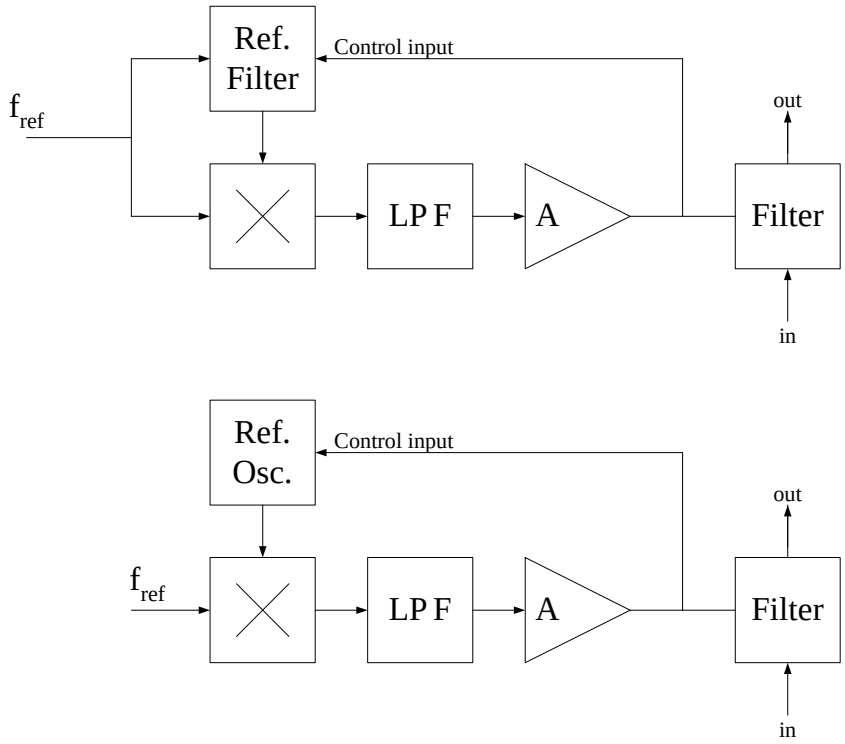

Figure 3.13: Two feedback loops for tuning filters

Of the two methods presented, the VCO tuning has the advantage of not being sensitive to phase errors due to, for example, the phase comparator. VCO non-linearities, however, influence the tuning error. The realized loop actually is a classic phase-locked loop (PLL).

3.5 Correlating errors by using switched capacitors

The drawback of making filters with coils and capacitors is the impossibility of integrating the coils. This resulted in the demand for fully integrated filters. Crystal and ceramic filters (both mechanical) are also commonly used filters. They usually have very high Q, do not need supply voltage, are cheap, but they cannot

be integrated on chip, which is the major drawback. The Sallen and Key filters, which apply active components, can be used to design on-chip filters. They only use resistors, capacitors and transistors as active components.

So far, continuous time filters were treated. They process the signal continuously in time, and use capacitors, resistors, coils and amplifiers to realize the filtering function on the basis of currents and voltages. In contrast, the sampled-data class using switched capacitors or switched currents or switched voltages, are not time continuous. The switched capacitor filters, for example, use switched capacitors to "simulate" resistors. The capacitor is switched between the two connections, where the "resistor" should be. The charge transfer on the system clock signal, causes the capacitor to behave like a resistor. An integrator made in this way is depicted in figure 3.14.

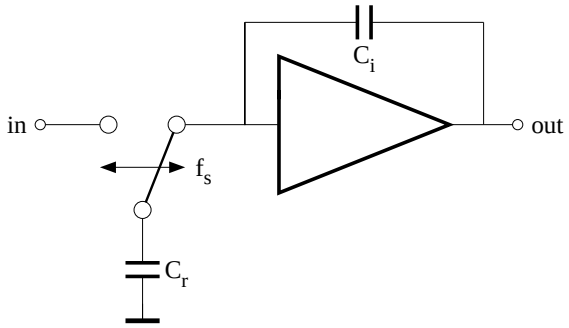

Figure 3.14: Example of switched capacitor (SC) integrator

The continuous time equivalent is depicted in figure 3.15. The equivalent re-

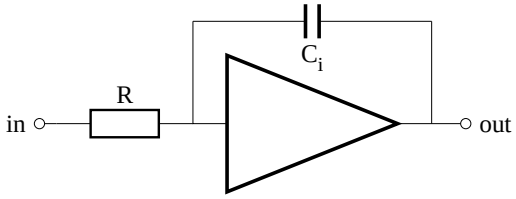

Figure 3.15: Equivalent continuous time (CT) integrator

sistor value yields:

$$R_s = \frac{1}{f_s C_r} \tag{3.9}$$

and thus the resulting transfer function of the circuit yields:

$$\frac{1}{sR_sC_i} = \frac{f_sC_r}{sC_i} \tag{3.10}$$

The big advantage of switched-capacitor filtering is its accuracy. The transfer function is determined not by absolute values of capacitors and resistors, but by a ratio of capacitors. So, when the capacitors match well, thus also have the same relative errors, the ratio of the capacitor values is not affected by the error in the absolute values of the capacitors. Thus the relative-frequency response can be designed very accurately, i.e. based on matching. Whereas the absolute frequency accuracy is governed by the clock signal.

Usually, the requirements of good matching and a accurate clock can be fulfilled. By changing the clock signal, the filter is tuned.

Clearly, a disadvantage of the switched capacitor filter is that a clock signal is necessary, which, via clock feed-through, quite often causes the desired signal to deteriorate. What is more, the use of sampled data systems requires pre-filtering. The sampling causes higher frequency bands to be folded to the base band, i.e. aliasing. Thus, continuous time pre-filtering is required.

The dynamic range of continuous time and switched-capacitor filters appears to be the same for equivalent structures and high sampling rates. This can be explained because of the simulation of the resistor by a capacitor. On every clock pulse, some amount of charge is transferred to the following circuit part, depending on the value of the capacitor. Keeping in mind the aliasing problem, the capacitor seen as a resistor.

3.6 Sensitivity

Different filter implementations, having the same transfer, may have different sensitivity to variations in component values. Key issue for the designer is to find the topology that for the lowest cost implements the required transfer function with the required specification, including capability of dealing with the uncertainty in parameters.

In this section a mathematical description of sensitivity is given. Subsequently, the sensitivity of passive LC ladder filters is considered as it appears to be very low. From that quality of passive LC ladder filter, design rules are formulated to improve the sensitivity performance of active filters.

3.6.1 Definition of sensitivity

The sensitivity, S_x^H, of a transfer, H, to a variation in a parameter, x, is defined as the ratio of the relative variations in H and x:

$$S_x^H = \frac{\partial H/H}{\partial x/x} \tag{3.11}$$

As for a filter transfer H is a complex function depending on frequency, also the sensitivity is in general a complex function depending on the frequency. Ideally, the sensitivity is zero.

Besides sensitivity of a transfer function to a parameter variation, also, for instance, the variation of a -3 dB corner frequency due to a variation in a parameter can be described by a sensitivity function.

Depending on what the key characteristics are of the filter, corresponding sensitivity functions can be derived. As the level of sensitivity is dependent on the characteristic

3.6.2 Sensitivity of passive LC Ladder Filters

From a relative long history of designing passive LC ladder filters, it appeared that this type of filters shows a very low sensitivity to parameter variations. The reason for this can be illustrated relatively easy. Consider the fifth-order low-pass filter of figure 3.16. In the pass band of this filter, the available power of the source is

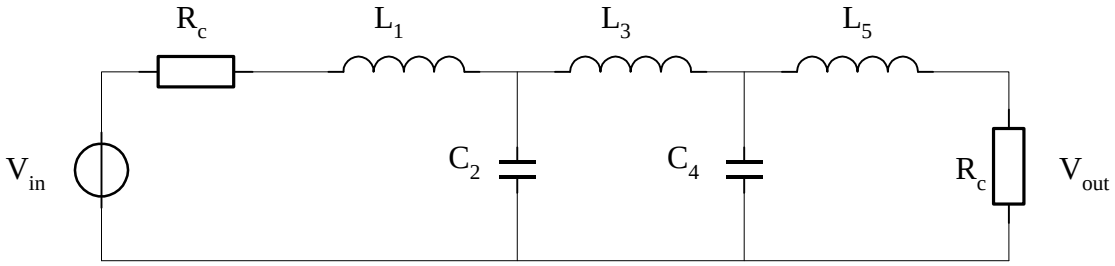

Figure 3.16: A LC ladder filter

completely passed through the filter network to the load, i.e. no power is dissipated in the filter. Consequently, due to variations in one or more of the elements, the delivered power to the load can only be lowered. This is schematically depicted in figure 3.17. Thus the derivative of the transfer with respect to that parameter in that point is zero:

$$S_{L_1}^{H} = \frac{L_1}{H}\frac{\partial H}{\partial L_1} = 0 \tag{3.12}$$

3.6.3 Design rules for active filters

Essential in the low sensitivity of passive filters is the fact that in the pass-band maximum available power of the source is transferred to the input. As the filter is *passive*, due to parameter variations the filter cannot have gain larger than one. Gain only reduces (or in the best case remains constant) when parameters

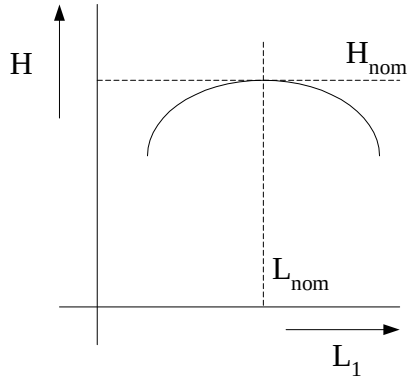

Figure 3.17: The variation in the power transfer as a function of the variation in a component at a specific frequency

change. When designing *active* low-sensitive filters this is a property that needs to be maintained, i.e. the property of a *passive terminal behavior*.

Figure 3.18 shows a capacitor as passive integrator. The transfer function,

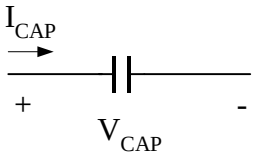

Figure 3.18: A capacitor as passive integrator, I_{CAP} the input current and V_{CAP} the output voltage.

$H(s)$, of the capacitor as integrator is given by:

$$H(s) = \frac{V_{CAP}}{I_{CAP}} = \frac{1}{sC} \tag{3.13}$$

The transfer function will always be passive, independent of variations in C. Only for a negative C this transfer function would become active. Active behavior means that power is increased. Figure 3.19 depicts an active (balanced) integrator. The transfer function of this integrator is given by:

$$\frac{V_{out}}{V_{in}} = H(s) = \frac{-1}{s(R_1 + R_2)C_{eff}} \tag{3.14}$$

in which $C_{eff} = \frac{C_1 C_2}{C_1 + C_2}$. Clearly, due to variations in component values this integrator also exhibits always a passive transfer, i.e. due to variations in the component values no sign change can occur. From straightforward calculations it appears that passive terminal behavior for this integrator is identical to fulfilling the port

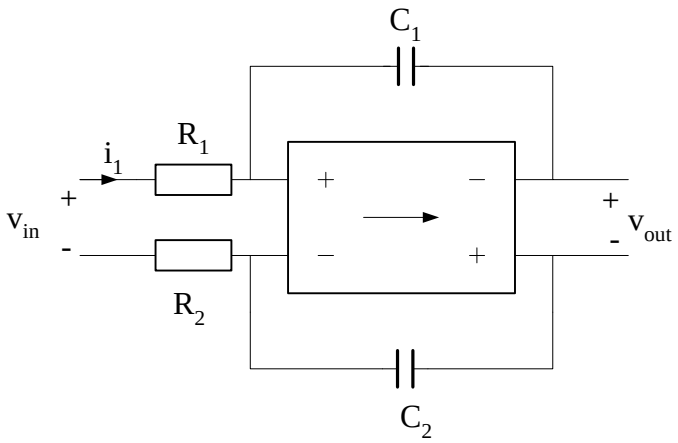

Figure 3.19: An active balanced integrator

conditions. Thus to have low sensitive filters, at least one should use integrators that fulfill the port conditions.

Subsequently, when considering that variations in parameters can also be translated in a kind of disturbances in the filter transfer, i.e. a kind of noise, it is evident that different topologies may exhibit different sensitivity performance. It is beyond the scope of these lecture notes to derive this property.

3.7 Exercises

1. Consider a filter designed with a certain level of model complexity. After the implementation, the filter is measured and it appears that the transfer differs from what is expected. Basically, the difference can have two types of nature.

(a) How would you characterize the nature of the difference in the case that the model complexity was too low?

(b) How would you characterize the nature of the difference when the model complexity is good but a model parameter deviates?

2. Consider the notch filter as depicted in figure 3.20. This type of filters are

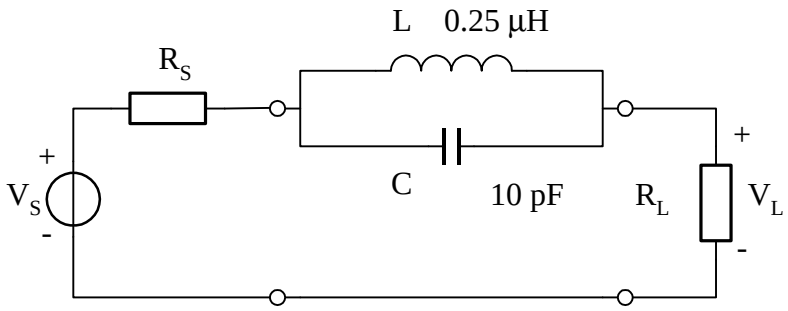

Figure 3.20: A notch filter.

used when a specific, small band, disturbing signal has to be suppressed.

(a) Determine the transfer V_L/V_S of this filter.

(b) At what frequency is the notch?

(c) How large is the suppression of this filter at that notch frequency?

Assume that the value of C_1 and L_1 have a maximum error of 5%.

(d) Calculate the worst-case deviation of the notch frequency.

(e) What is the magnitude of the suppression at the original notch frequency for this worst case situation?

(f) What do you conclude?

3. Simple, but effective voltage references can be made by a diode voltage. A possible implementation is given in figure 3.21. The diode is biased at a current I_D and the resulting diode voltage can be used as a reference voltage.

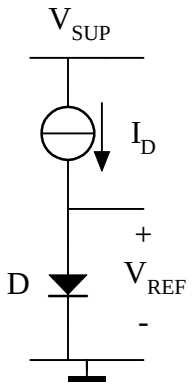

Figure 3.21: A simple diode-voltage reference.

(a) Give an expression for the reference voltage V_{REF}.

(b) What is the error in the reference voltage when the bias current has a relative error of δ_I?

(c) What is the sensitivity of the reference voltage for variations in the diode current?

(d) What is your conclusion?

4. Accurate components are required for to be able to design accurate filters. What should be done to design accurate on-chip resistors? Motivate the effect of the following options.

(a) Large number of squares

(b) Large squares

(c) Low resistive layer

5. In circuits often resistors are required that have a good matching. On chips this matching can be realized relatively accurately. Consider the situation that a resistor is required with value R and a resistor with value of $4R$. Resistor R has got 5 squares and is designed as depicted in figure 3.22.

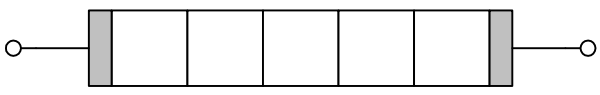

Figure 3.22: Resistor R designed with 5 squares.

(a) How would you design the resistor with value $4R$ for maximum matching? Design it such that a gradient in the doping level (and thus in the sheet resistance) is compensated for. Motivate your answer.

6. Consider the resistor as depicted in figure 3.23.

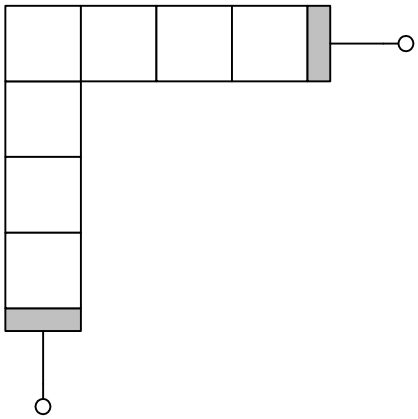

Figure 3.23: A resistor with a corner.

(a) When you can neglect the contact resistances, what do you expect the resistance R is?

- $6R_\square < R < 7R_\square$
- $R = 7R_\square$
- $7R_\square < R < 8R_\square$

Motivate your choice.

(b) What drawbacks can you imagine when you use such a corner in a resistor?

Designers use corners in a resistors just because that without a corner the resistor would be too long.

(c) How would you design this resistor when corners are not allowed (using the same sheet resistance)?

7. Given two resistive layers in a certain technology, see table 3.3. In this table $C_\square$ stands for the parasitic capacitance per $(\mu\text{m})^2$ area between the bottom of the resistor and its environment. C_{edge} is the capacitance per μm length of the sidewalls of the resistor and its environment.

Layer	R [Ω/□]			$C_{\square}$ [fF/(μm)^2]	C_{edge} [fF/μm]	min. □ [μm]	ΔW [μm]		
BW	690	720	750	0.29	0.55	2	0	0.5	1
WP	24.0	24.5	25.0	0.31	0.70	2	0	0.5	1

Table 3.3: Specifications of the BW and WP layer of a DIMES technology.

(a) Design a resistor body (i.e. resistor without contacts) with a value of 10 kΩ and a uncertainty of maximum worst-case 10%. Give the number of squares and the size of the squares and the total size of the resistor you would draw on a lay-out.

(b) Determine the bandwidth of the designed resistors when for a distributed RC-line the bandwidth, when used as one-port resistor, can be described by:

$$B = \frac{\frac{1}{4}\pi^2}{2\pi RC} \tag{3.15}$$

in which R and C are the total resistance and capacitance, respectively.

8. The accuracy of integrated filters is mainly determined by the accuracy of the comprising resistors and capacitors which determine the time constants. In order to obtain high-accuracy filters, tuning can be used. A block schematic of a filter using tuning is depicted in figure 3.24.

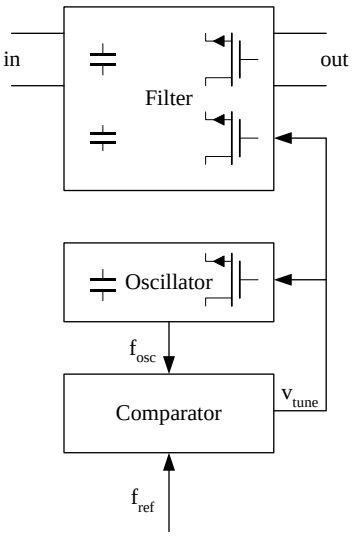

Figure 3.24: Using tuning in order to obtain high accuracy.

(a) Resistor and capacitor values cannot be realized exactly on chip. What reasons do you know for this (3) ?

(b) Explain how the tuning is realized in figure 3.24.

(c) What is the main criterion regarding the relation between the filter and the oscillator?

(d) Explain what happens with the filter when the comparator has got a constant error (offset) at its output.

9. Given the switched-capacitor filter in figure 3.25.

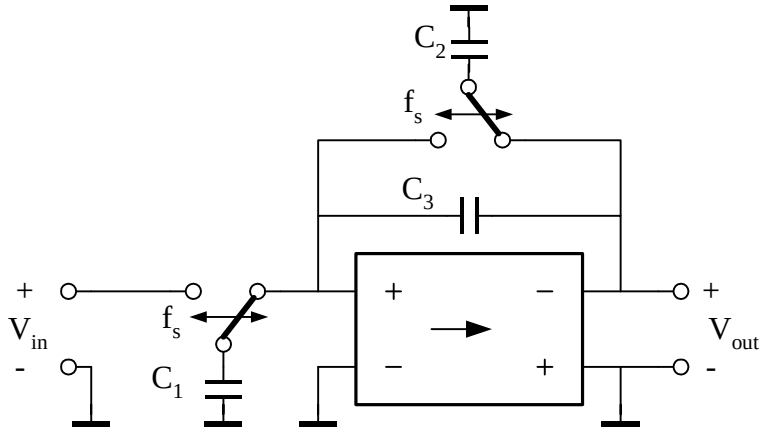

Figure 3.25: A switched capacitor filter.

(a) Draw the "equivalent" circuit diagram in which the switched capacitors are replaced by resistors.

(b) The circuit diagram of the previous question is not completely equivalent with the circuit diagram of the switched capacitor filter. What is the difference in behavior of the two circuit diagrams?

(c) Calculate the transfer of the filter.

(d) What effect has frequency f_s on the transfer of the switched-capacitor filter?

10. Given the circuit of a time-continuous first-order low-pass filter of figure 3.26.

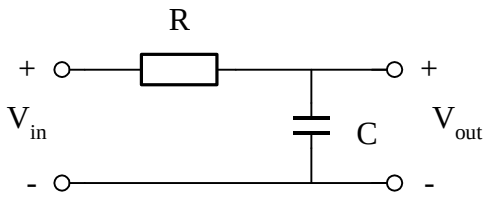

Figure 3.26: A time-continuous first-order low-pass filter.

(a) Show that the total noise power of the equivalent noise voltage at the output of the filter is given by:

$$P = \frac{kT}{C} \tag{3.16}$$

The power of this noise voltage can be interpreted as the uncertainty there is in the DC capacitor voltage when the capacitor would be disconnected from the circuit.

Subsequently, the switched-capacitor version of a resistor is considered, see figure 3.27. The input and output of the switched capacitor can be considered to

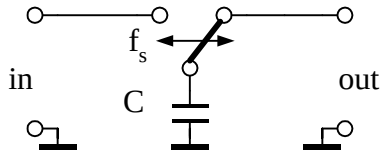

Figure 3.27: A switched-capacitor implementation of a resistance.

be connected to other circuitry.

(b) Show that the power spectral density of the noise voltage in series with either the input or or output is given by:

$$S = 4kTR \tag{3.17}$$

in which R is the equivalent resistance.

(c) Conclusion?

11. Explain why a filter with a high dynamic range is likely to have also a low sensitivity.

Chapter 4

Implementing the integrators

4.1 Integrators

Filters can be considered to be composed of integrators or differentiators. Integrators are used because differentiators are inherently difficult to implement, due to stability problems. In continuous time filters, using the Laplace domain description, the integrating function can be defined as the transfer a_0/s. There are two integrating elements in electronics, i.e. coils and capacitors. As high-quality coils cannot yet be integrated on a chip, capacitors almost always are used to implement the integrator function. The current that flows through a capacitor results in an integrated voltage across the capacitor terminals. If the coil were taken as an integrating element, the voltage across the coil would result in an integrated current through the coil.

Choosing the capacitor as the integrating element implies using a current as input quantity, and getting a voltage as output quantity. This necessitates using voltage-to-current conveyors, or (trans-)conductances in order be able to connect the integrators, see figure 4.1. Between each integrator a conductance, G, is placed to convert the output voltage of a capacitor to a current such that it can be used as an input signal for the next integrator. At each node additional signals can be added and from each signals can be distributed in order to construct a state-space filter. However, adding signals as currents is much easier than adding them as voltages. The analogous holds for the distribution of signals; distributing signals as voltages is much easier than distributing them as currents. As the input voltage and the output current of a conductor have a linear static relation, it does not matter for the filter transfer whether the addition or distribution is done in the voltage or current domain. The only difference is a factor equal to G. Therefore, when adding signals, the easiest way is to do so in the current domain, i.e. at the input of the integrator. Distributing is done the easiest by using the output voltage of the integrator. This means that the gain blocks required to implement a

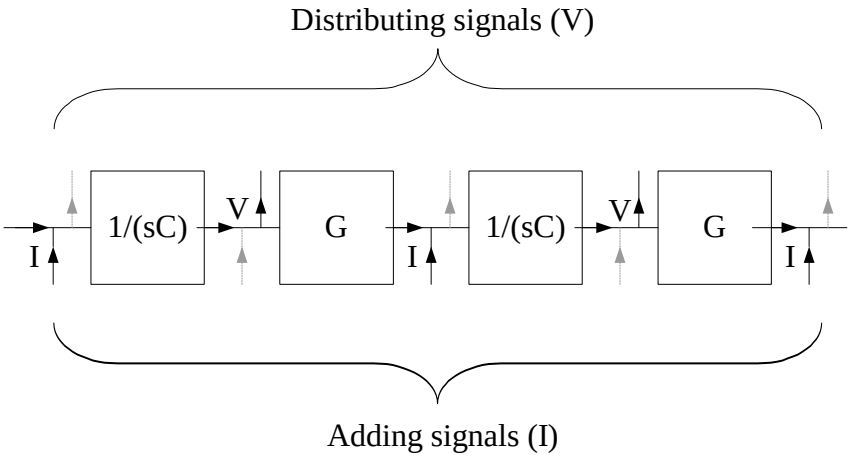

Figure 4.1: A cascade connection of integrators

state-space filter are also conductances (see for instance the example in chapter 2 in which a state-space filter was implemented with opamps, resistors and capacitors.). The simplest solution for conductances is to make use of resistors.

Essentially, it means that to be able to connect the integrators the dimension of the input-output transfer should be dimensionless. Whereas, the transfer of the capacitor has a dimension equal to [Ω]. Therefore, using a cascade of a conductance and a capacitor yields the required dimensionless transfer. It may be clear, that it, consequently, does not differ whether the internal signals of the filter are described in terms of voltages of currents. The location where it matters is at the input and output of the filter. There, the dimensions of the signals should correspond to the requirements.

Thus, an implementation of an integrator (thus with a dimensionless transfer) is composed of a capacitor and a conductor. Taken into account that each of them can be either passive or active, four types of integrators can be distinguished. Figure 4.2 depicts the four types in the case that the input and output signal are considered to be a voltage.

The *admittance-impedance* integrator does not use active components. Both the required conductance and integration are implemented in a passive fashion. When no active elements are used in the other blocks of the filter, it is not possible to make filters with complex poles by using this type of integrator. Therefore, this type of integrator is not used.

The second type of integrator is the *admittance-transimpedance* integrator. In this type of integrator, the realization of the actual integration function is active. The advantage is that the "opamp" used is a well-known electronic function, that can be easily integrated. The opamp can be designed to operate rail-to-rail at the output terminals, so full advantage is taken of the supply voltage. This allows

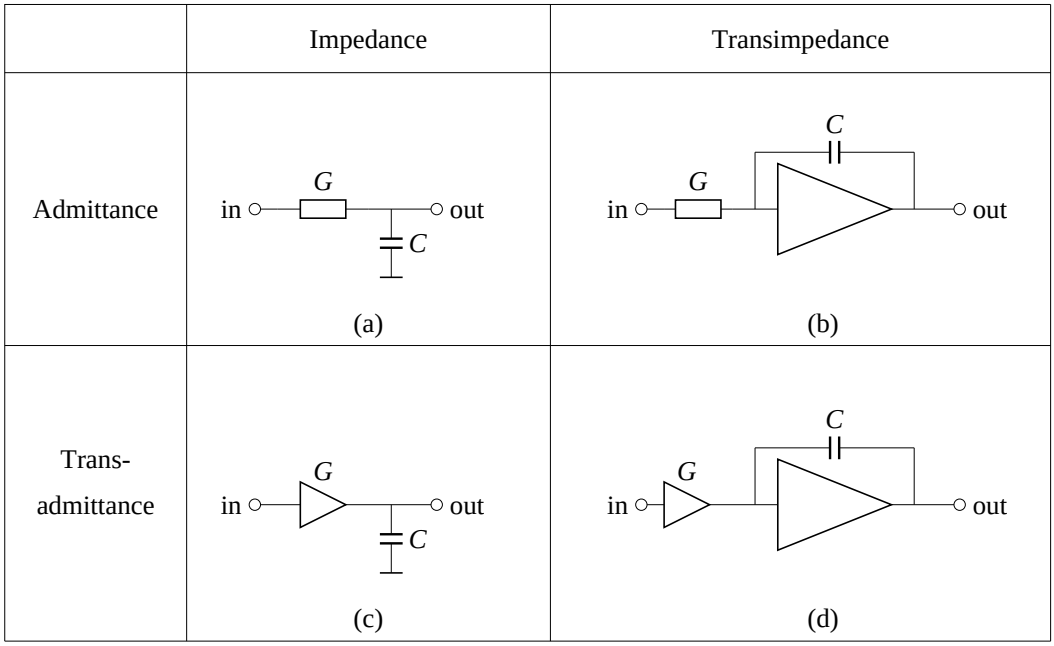

Figure 4.2: Four classes of integrators

for optimal dynamic range performance. The resistor used can be integrated as a diffused resistor, but it could also be implemented as an MOS transistor in the triode region thus yielding the MOSFET-C filters.

The third type, the *transadmittance-impedance* integrator, makes use of active "conductances", or transconductances. The advantage of transconductors is that they are able to operate at high frequencies, because in these integrators the parasitic capacitors of the transconductor are in parallel with the integrator capacitors. Thus, they can be accounted for easily in the dimensioning of the required capacitance. A major drawback, however, is that it seems impossible to implement transconductors with rail-to-rail input capability.

The fourth type of integrator is the *transadmittance-transimpedance* integrator. This integrator has no advantages over the second and third integrators mentioned. The disadvantage is the use of two active parts. Both parts add distortion, as distortion is chiefly formed by active components and, moreover, the power consumption and the noise production increase.

In conclusion, the second and third type of integrators are preferred when designing filters. For both types of integrators an active part is required. In the next section some simple implementations of integrators are discussed. In the next chapters the design of the amplifier (nullor/opamp) active parts is treated.

4.2 Small-signal models for nonlinear devices

For designing high-performance integrators it was found in the previous section that active elements are required. Two types of integrators showed to be preferable.

In the following sections, a start is made with how to implement the amplifier block that is often required in filters. Assuming, that either the active conductance or active integration is realized by means of feedback around a nullor (active part), the two integrators as depicted in 4.3 are obtained.

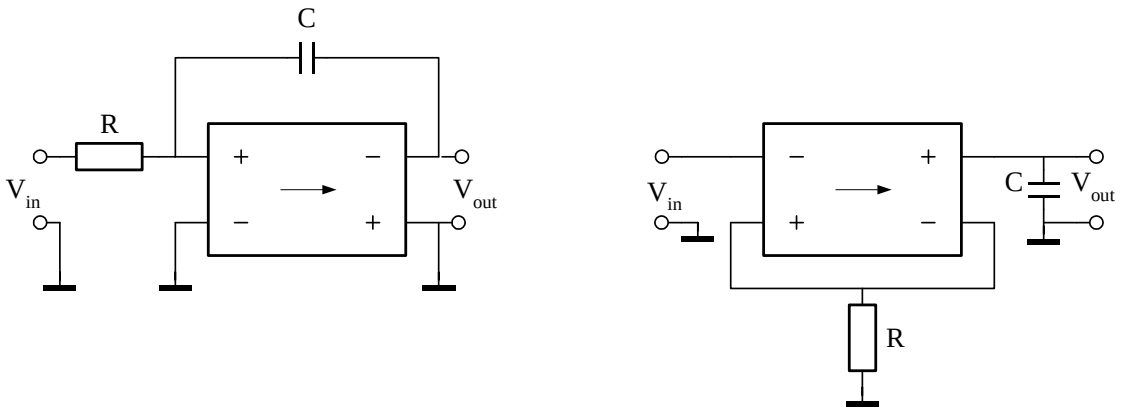

Figure 4.3: Two types of integrators using feedback around a nullor. A) Passive conductance and active impedance B) Active conductance and passive impedance

For implementing these integrators it is required to design nullor implementations. A nullor is an element defined by circuit theory. It can be seen as an ideal gain block, i.e. infinite gain, infinite bandwidth and no noise. In practice the nullor should be approximated by physical elements. The closer the nullor is approximated the better (the more linear and accurate) the input-output transfer of the integrator is. In electronic design the nullor is approximated by combining one or more voltage-controlled current sources. A single voltage-controlled current source is depicted in figure 4.4 The choice for this voltage-controlled source as a basic building block is not arbitrary. As will be seen later on, the commonly used active devices, like bipolar transistor, junction FET (JFET) and the MOS transistor can be described for small variations on a quiescent point as a voltage-controlled current source. This is the small-signal approximation.

In this context electronic design can be seen as first design a circuit by using linear elements, like voltage-controlled current sources, resistors and capacitors. Second, replace the voltage-controlled current sources by active devices (bipolar transistors etc.) with an appropriate quiescent (bias) point.

In the next sections, first the basic concepts of the small-signal analysis are reviewed, and subsequently the small-signal models for the commonly used nonlinear

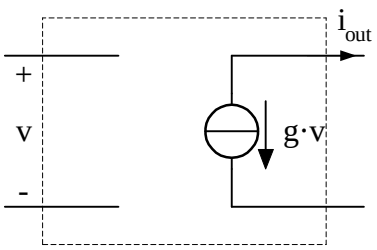

Figure 4.4: A voltage-controlled current source, g, as a first approximation of a nullor.

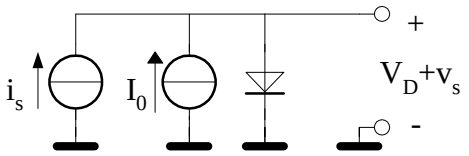

Figure 4.5: A biased diode

electronic elements are derived.

4.3 Small-signal analysis

Small-signal analysis is a technique to describe the behavior of nonlinear elements for small (signal) variations around a quiescent point. The obtained model is the so-called small-signal model. This in contrast to the set of equations describing the nonlinear behavior of the element which is called the large-signal model.

To illustrate this difference consider the biased diode in figure 4.5 (ignore for the moment the signal source i_s). The bias current, I_D, through the diode and the bias voltage, V_D, across the diode are related via the well-known *nonlinear* equation:

$$I_D = \tilde{f}(V_D) = I_S \left[\exp\left(\frac{qV_D}{kT}\right) - 1 \right] \tag{4.1}$$

in which I_S is the saturation current of the diode, q the electron charge, k the Boltzmann constant and T the absolute temperature. This equation relates the total current through the diode to the total voltage across the diode. The combination of I_D and V_D is called the quiescent (bias) point of the diode. To find this point either numerical routines are required, this is used by computer simulators for instance, or graphical representations of equations to find the corresponding intersection point, or straightforwardly solving the set of nonlinear equations. Often, via inspection an accurate approximation of the bias point can be obtained. For the diode circuit of figure 4.5 the quiescent current equals $I_D = I_0$ as the current I_0 can only flow through the diode. From expression (4.1) the corresponding

quiescent voltage V_D can be obtained.

Subsequently, it is assumed that on top of the current I_0 a little variation due to a signal i_s is present. This is modelled in figure 4.5 by the current source i_s. As a result, the diode voltage will have also a small variation, called v_s. Now, this small variation could be calculated via:

$$v_s = \tilde{g}(I_D + i_s) - V_D \tag{4.2}$$

in which $\tilde{g}(.)$ is the inverse of $\tilde{f}(.)$. The variation v_s is found via determining the difference in the diode voltage for the case that signal i_s is present and the case that signal i_s is not present. For this calculation still the nonlinear function $\tilde{f}(.)$, or its inverse $\tilde{g}(.)$ is required.

In contrast, the small-signal analysis uses the fact that for determining the effect of small disturbances a nonlinear function may be considered to be linear. This linear behavior is found from the corresponding Taylor series. The Taylor series for equation (4.1) is given by:

$$I_D + i_s = \tilde{f}(V_D) + \left[\frac{d\tilde{f}(V)}{dV}\right]_{V=V_D} \cdot v_s + \text{higher order derivatives} \tag{4.3}$$

Performing the required calculations yields:

$$I_D + i_s = I_D + g_m \cdot v_s + c_2 v_s^2 + \cdots \tag{4.4}$$

in which g_m is called the small-signal conductance and given by $\frac{qI_D}{kT}$. Assuming that v_s is small such that the higher-order terms with v_s^2, et cetera, can be ignored (the essential assumption for the small-signal analysis), yields:

$$i_s = g_m \cdot v_s \tag{4.5}$$

which is a very simple relation. As g_m can be found very easily, the signal variations can be found via the small-signal analysis also very easily. The graphical interpretation of this small-signal approximation is depicted in figure 4.6. The quiescent point, Q, is found from the large-signal nonlinear relation. Subsequently, the effect of small variations around the quiescent point can be explained via the tangent line in the point Q. The slope of this tangent is found from the first derivative of the nonlinear function yielding; g_m.

From the previous discussion it is clear that for determining the effect of small variations on a nonlinear circuit three steps are required:

- determine the quiescent (bias) point;
- derive the small-signal model;
- perform the analysis on the small-signal model.

For the first two steps corresponding circuit diagrams are used modelling the relevant relations for that specific step.

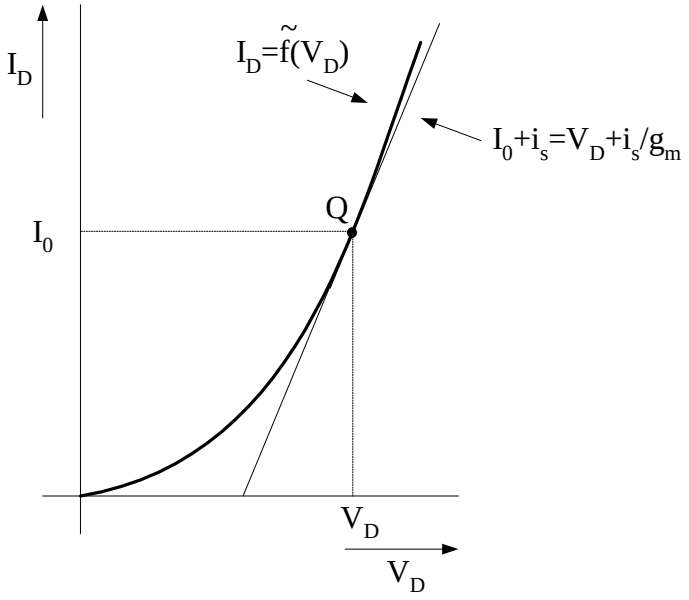

Figure 4.6: Graphical interpretation of the small-signal behavior of a diode

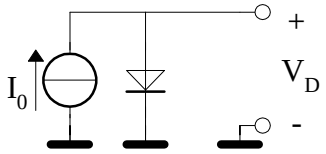

Figure 4.7: Circuit diagram for determining the quiescent point of the circuit of figure 4.5

4.3.1 The circuit for determining the bias point

For finding the quiescent point of a circuit, a simplified circuit diagram can be used. The simplifications are obtained when we consider what a quiescent point is. The quiescent point is the *point* the circuit evolves to in the limit when no signals are applied to the circuit. So, first of all, the signal sources can be set to zero. This means that a signal current source becomes an open and a signal voltage source becomes a short. Further, as we have to do with a point, a static quantity, also the dynamic elements can be set to zero. Thus a capacitor is replaced by an open and an inductor is replaced by a short. In the resulting circuit diagram only *static* elements and quantities are found.

The corresponding circuit for determining the quiescent point for the diode circuit in figure 4.5 is given in figure 4.7. From this circuit the quiescent point can be derived either via numerical methods (computer simulation), via solving the set of expressions or via graphical representation. Often, inspecting the circuit may

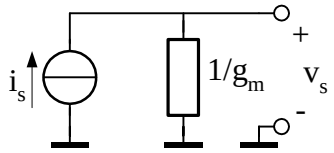

Figure 4.8: The small-signal diagram related to the circuit of figure 4.5

result in a considerable simplification for the calculations.

4.3.2 The small-signal circuit

When the quiescent point is found, then for each element its response on a small variation on that quiescent point can be determined. This results in small-signal models for the elements. The corresponding small-signal diagram for figure 4.5 is depicted in figure 4.8. In a small-signal circuit no static currents and voltage are found which are related to the quiescent point. Further, all the elements in the small-signal circuit are linear. This is because the static voltages and currents in combination with the nonlinear elements are replaced by their linear small-signal models. When DC currents and/or voltages arise in the circuit, this is because the signal source contains signal at DC.

4.3.3 Analysis of the small-signal behavior

Finally, when the small-signal circuit is obtained, analysis can be performed to determine the response of the circuit on a small signal. Analysis can be done using all methods and techniques available for linear circuit analysis, i.e. Kirchhoff laws, Modified Nodal Analysis (MNA), superposition, Laplace transform, etc.

From the circuit of figure 4.8 the variation in diode voltage, v_s, is easily obtained as:

$$v_s = i_s/g_m \tag{4.6}$$

When the results are obtained, one should check whether for the obtained response still the small-signal approximation holds. If the approximation appears not to be valid for the small-signal response, then , consequently, the obtained response is not an accurate prediction of the response of the nonlinear circuit.

This section showed the use of small-signal models to derive the response of a nonlinear circuit to a relatively small signal. Via the small-signal models the calculations can be simplified enormously as soon as the quiescent point is obtained. In the next section it is shown that for the commonly used bipolar and field-effect transistors the relation between the nonlinear model and the quiescent point to the small-signal model is straightforward. This is essential as in the end a designer synthesizes its, in this case, nullor approximation by first using voltage-controlled

current sources. Subsequently, to approximate these voltage-controlled current sources by transistors, the appropriate biasing points should be found such that the small-signal behaviors of the transistors approach the required small-signal behaviors.

4.4 Models for devices

In this section the small-signal models for the commonly used active devices are discussed. These models can be used to synthesize on a small-signal level a nullor implementation.

4.4.1 Bipolar transistor

Several physical models are available for bipolar transistors, like Ebers-Moll model, transport model and the Mextram model. In these lecture notes the transport model will be used. It describes the terminal currents of the bipolar transistor as a function of the terminal voltages, see figures 4.9 In most of the cases the bipolar

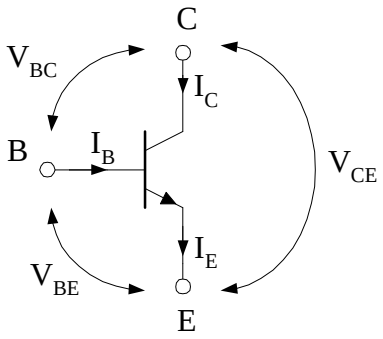

Figure 4.9: The terminal currents and voltages of a bipolar transistor

transistor is used in its active forward mode. For a NPN transistor this means $V_{BE} > 0$ and $V_{BC} < 0$. The corresponding essential part for that mode is depicted in in figure 4.10. For this part of the large-signal model the following relations holds:

$$I_C = I_S \left[\exp\left(\frac{V_{BE}}{V_T}\right) - 1 \right] \tag{4.7}$$

$$I_B = \frac{I_C}{B_F} \tag{4.8}$$

in which I_C and I_B are the collector and base current, respectively, V_{BE} the base-emitter voltage, V_T the thermal voltage and I_S the saturation current and B_F the

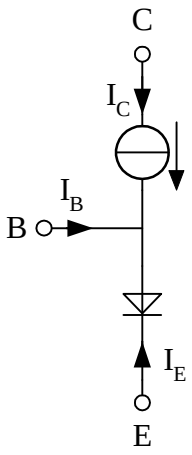

Figure 4.10: The essential part of the transport model for a transistor in its forward active mode.

current-gain factor. The corresponding small-signal diagram is depicted in figure 4.11. The voltage-controlled current source is the small-signal equivalent of the

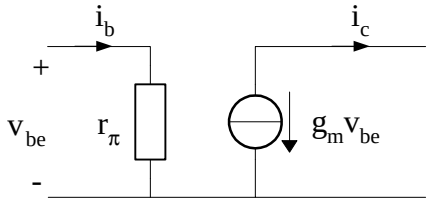

Figure 4.11: Small-signal diagram related to figure 4.10

current source (which has the exponential relation to the V_{BE}) of figure 4.10, i.e. equation 4.8. The transconductance factor, g_m, is given by (compare with the diode):

$$g_m = \frac{I_C}{V_T} = \frac{qI_C}{kT} \tag{4.9}$$

Due to the base current, also a small-signal input resistance (r_π) can be identified. It can be directly derived from determining the derivative of I_B versus V_{BE}. Here we use expression (4.8), as:

$$r_\pi = \frac{v_{be}}{i_c} \cdot \frac{i_c}{i_b} \tag{4.10}$$

The first factor is given by $1/g_m$ whereas the second factor is the small-signal current-gain factor (β_f) that can be derived from expression (4.8):

$$\beta_f = B_F \tag{4.11}$$

which is for silicon transistors on the order of 50 - 100. Thus

$$r_\pi = \frac{\beta_f}{g_m} \tag{4.12}$$

The small-signal model of figure 4.11 models the essential static small-signal behavior of a bipolar transistor. For taking the essential dynamic behavior into account, i.e. charge storage and finite speed of the carriers, a capacitance, c_π in parallel with r_π should be used. This capacitance comprises two physical effects:

- charge storage in the base-emitter depletion layer, c_{je}
- charge storage in the base region, c_d

The c_{je} accounts for the charge variation in the base-emitter depletion area when the base-emitter voltage is changed. Often it is assumed to be constant. The diffusion capacitance, c_d, models the charge present in the base region when a collector current is flowing. The higher the collector current the more charge is present in the base region and the slower the carriers move trough the base region, the more charge is present in the base region. From physical modelling the following relation can be found for the diffusion capacitance:

$$c_d = g_m \tau_f \tag{4.13}$$

in which τ_f is the time it takes for a carrier to cross the base region (on the order of pico seconds). Thus, c_π is given by:

$$c_\pi = c_{je} + c_d = c_{je} + g_m \tau_f \tag{4.14}$$

The corresponding small-signal diagram is depicted in figure 4.12. This model

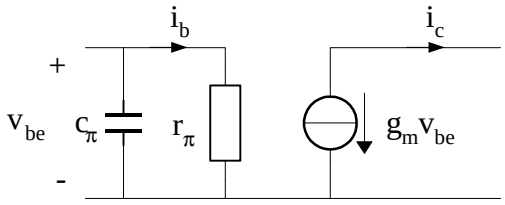

Figure 4.12: Essential dynamic small-signal diagram of a bipolar transistor

is the essential dynamic small-signal model in designing nullor implementations. After designing using this model, often some checks have to be performed in order to validate this simple small-signal model. The following three effects need to be checked:

- r_o: the output resistance

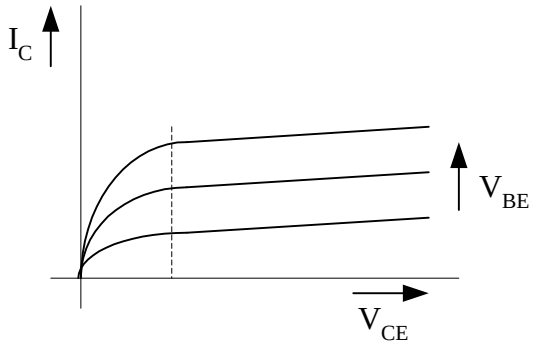

Figure 4.13: Dependency of the collector current on the collector-emitter voltage with the base-emitter voltage as a parameter.

- r_b: base bulk resistance
- c_μ : the base-collector junction capacitance

The output resistance of a transistor is physically explained by the fact that the depletion area of the reverse-biased base-collector junction extends further into the base region for larger reverse voltages. Consequently, the effective base width reduces and the collector current increases. This is found when measuring the collector current as function of the collector-emitter voltage. In figure 4.13 typical behavior is depicted with the base-emitter voltage as a parameter. The effect of the finite output resistance is seen in the slope of the curves beyond the dashed line, where the forward active mode is. The slopes in the curves can be modelled by a resistance equal to:

$$r_o = \frac{V_A}{I_C} \tag{4.15}$$

in which V_A is called the Early voltage. This can be seen by extrapolating the straight part of the curves to the left where they intersect at a single point of the negative x-axis. This point is approximately at V_A. For silicon it is on the order of 50 V to 100 V.

The base bulk resistance, r_b, models the resistance of the silicon between the on-chip connection of the transistor to which the metal can be connected and the internal physical base contact. It is a resistance on the order of 10 Ω to 100 Ω.

Finally, analogous to the base-emitter junction, also the base-collector junction is responsible for a junction capacitance, c_μ. It is located between the base and collector node.

Figure 4.14 shows the model in which r_o, r_b and c_μ are included.

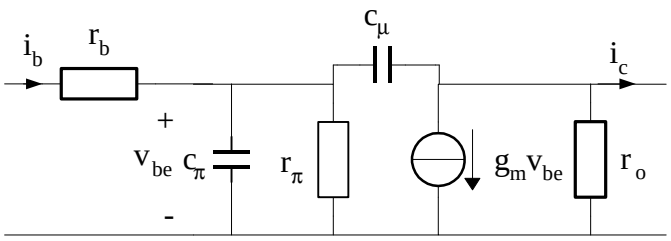

Figure 4.14: A small-signal model for a bipolar transistor including the most relevant second-order effects

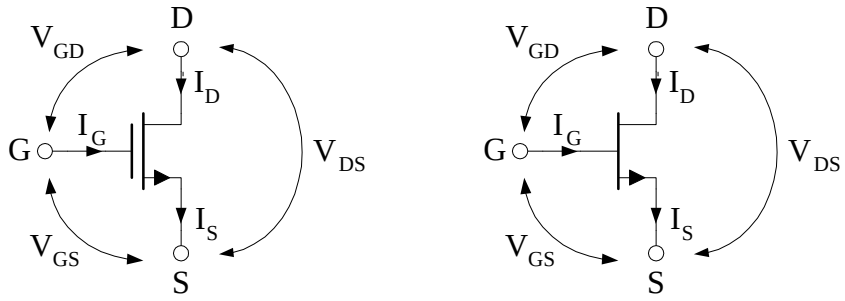

Figure 4.15: Symbols for JFET and MOSFET

4.4.2 Field-effect transistors

The group of field effect transistors (FETs) comprise the commonly used Junction-FETs and MOSFETs. Here a small-signal diagram is derived for these transistors. Basically, the models can be the same, with, however, different physical interpretation of the model components. The component symbols are depicted in figures 4.15.

The usual working mode for field-effect transistors is the forward saturation mode. In this mode the transistors behavior can be modelled as a voltage-controlled current source, like the bipolar transistor in active forward mode.

The basic relations that can be used for the Junction FET in the forward saturation mode are:

$$I_D = I_{DSS}\left(1 - \frac{V_{GS}}{V_{th}}\right) \tag{4.16}$$

$$I_G = 0 \tag{4.17}$$

in which I_D is the drain current, V_{GS} is the gate-source voltage, I_{DSS} the maximum drain current which is obtained at $V_{GS} = 0\text{V}$ (normally on device) and V_{th} is the threshold voltage. The gate current can be assumed to be zero as the gate-source junction is a reverse biased junction.

The basic dynamic behavior is modelled by the gate-source capacitance, c_{gs}, of

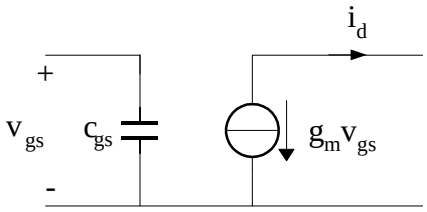

Figure 4.16: The essential dynamic small-signal diagram

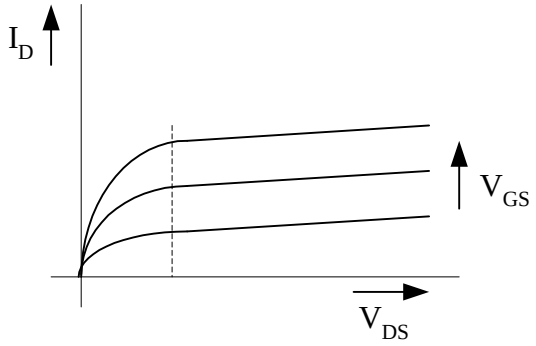

Figure 4.17: Dependency of the drain current on the drain-source voltage with the gate-source voltage as a parameter.

the corresponding reverse-biased diode. Thus the essential dynamic small-signal diagram for a junction FET is as depicted in figure 4.16. As the DC gate current is zero, no analogous small-signal component for r_π is found. The transconductance g_m is given by:

$$g_m = \frac{2}{V_{th}}\sqrt{I_{DSS} I_D} \tag{4.18}$$

The following three additional effects need to be considered for checking the validity of the model:

- output resistance, r_d
- gate-drain capacitance, c_{gd}
- gate resistance, R_G

The small-signal output resistance is found from the output characteristic of the JFET as depicted in figure 4.17. The curves are comparable with the curves for the bipolar transistor. Via

$$r_d = \left(\frac{dI_D}{dV_{DS}}\right)^{-1} \tag{4.19}$$

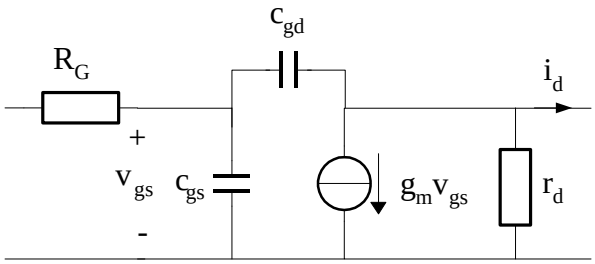

Figure 4.18: The dynamic small-signal diagram in which also r_d, c_{gd} and R_G are included

the value of the output resistance can be obtained. Often this is, analogous to the bipolar transistor described by,

$$r_d = \frac{V_A}{I_D} \tag{4.20}$$

with V_A also called the Early voltage.

Between the gate and drain also a reverse-biased junction is present. Physically the gate-source and gate-drain junctions are just separate parts of one junction. Its capacitance is modelled with c_{gd}.

In the technology in which the JFET is made, often poly-silicon is used to contact terminals, instead of aluminium. Poly-silicon has a relatively high resistance compared with aluminium. Therefore, to account for this resistance a series resistor is added to the gate, R_G.

Thus, the small-signal model for the JFET including these effects is as depicted in figure 4.18.

For the MOSFET also the small-signal model as presented in figures 4.16 and 4.18 can be used for the small-signal analysis. The differences between the JFET and MOSFET are mainly physical. The main difference is that the channel conductivity of a MOSFET is influenced via a field applied via an oxide capacitance instead of a junction capacitance for a JFET. Therefore, the capacitances in the model for the MOSFET relate to different *physical* capacitances. Further, a gate-source junction of a JFET can become forward biased resulting in a gate-current, whereas for a MOSFET the gate current is zero as long as the gate-oxide is thick enough to prevent tunnelling.

Usually, in design MOSFETs are used which are normally-off devices, i.e. no drain current at zero gate-source voltage. In contrast, JFETs are often of the normally on type, i.e. maximum drain current for zero gate-source voltage. Therefore, the relation between the drain current and gate-source voltage is expresses differently, as:

$$I_D = \frac{1}{2}\beta(V_{gs} - V_{th})^2 \tag{4.21}$$

in which β is parameters taken the geometry of the MOSFET into account, like width, length, oxide thickness, et cetera. The corresponding expression for the small-signal transconductance is given by:

$$g_m = \sqrt{2\beta I_D} \tag{4.22}$$

4.5 Simple MOS integrator implementations

MOS transistors in strong inversion are known to operate in two regions, i.e. the saturation and the triode regions. In addition, MOS transistors can be operated actively as well as passively. The difference is that actively operated MOS transistors have the input signal at the gate terminal, and passive transistors have the input signal at the source terminal, i.e. they are used as resistor. By making this division, four different types of MOS integrators are possible.

4.5.1 Passive triode integrator

In the triode region, the MOS transistor behaves according to the following equation:

$$I_D = \beta \left[(V_{GS} - V_{th})V_{DS} - \frac{1}{2}V_{DS}^2 \right] \tag{4.23}$$

in which I_D is the drain current, β a factor depending on geometry and technology, V_{GS} is the gate-source voltage, V_{th} is the threshold voltage and V_{DS} is the drain-source voltage. This can be rewritten as:

$$\frac{I_D}{V_{DS}} = \beta \left[(V_{GS} - V_{th}) - \frac{1}{2}V_{DS} \right] = f(V_{DS}) \tag{4.24}$$

In the triode region, $V_G > V_D + V_{th}$ and $V_G > V_S + V_{th}$ for NMOS transistors. From this equation, it becomes clear that it is possible to use the MOS transistor in the triode region for the voltage-to-current conversion, after which the current can be integrated into a voltage by a capacitor. The transconductance, however, is a function of the drain-source voltage. This implies non-linearity. Most of the even-order non-linearities can be eliminated by using balanced structures. An example can be seen in figure 4.19.

4.5.2 Active triode integrator

The active triode integrator is described by the same equations as the passive integrator. The only difference is the coupling of the signals to the transistor. The transconductance, g_m, of the MOSFET is used to couple the integrators. The distortion can also be reduced by applying balanced structures. An active triode integrator can be seen in fig.4.20.

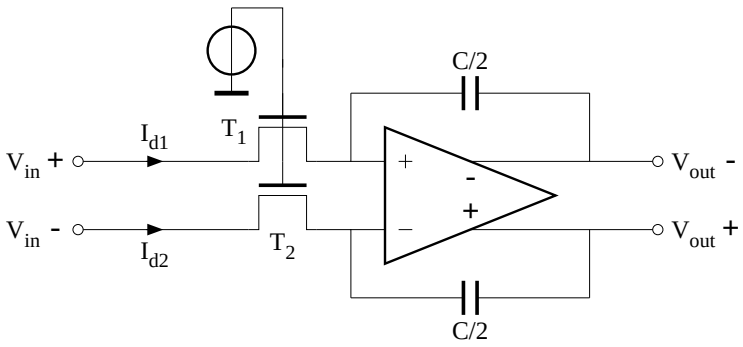

Figure 4.19: Passive triode integrator

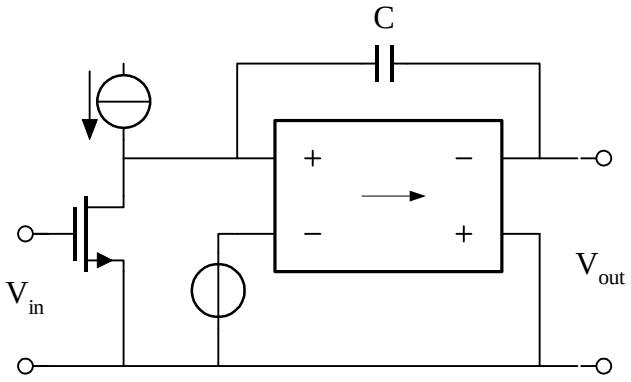

Figure 4.20: Active triode integrator

4.5.3 Active saturation integrator

For MOS transistors in saturation, the following equation describes the relation between the drain current and the terminal voltages:

$$I_d = \frac{1}{2}\beta(V_{GS} - V_{th})^2 \tag{4.25}$$

In the saturation region $V_G < V_D + V_{th}$. Also in this integrator, even-order non-linearities can be canceled by using balanced structures. An example is given in fig.4.21. This integrator has good high-frequency characteristics, because parasitic capacitances are in parallel with the integrator capacitances, such that no parasitic poles occur. Because NMOS and PMOS transistor non-linearities partly cancel out, the distortion reduces but further measures are necessary to decrease the resulting distortion.

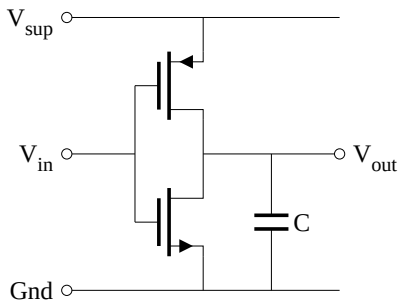

Figure 4.21: Active saturation integrator

4.5.4 Passive saturation integrator

A passive saturation integrator uses the source terminal as the input and is biased in the saturation region. An example is given in fig.4.22. It should be noted that

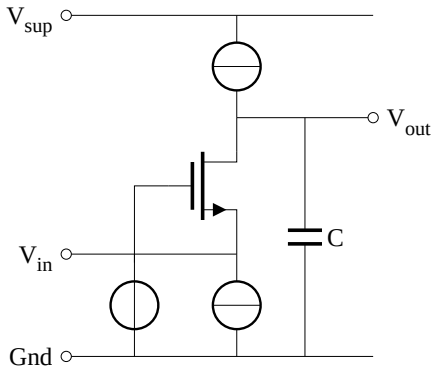

Figure 4.22: Passive saturation integrator

this type of integrator belongs to the first group of integrators as discussed in the previous section. Realizing the integration active yields an integrator from the second type.

4.6 Exercises

1. Given the voltage-to-voltage integrator in figure 4.23.

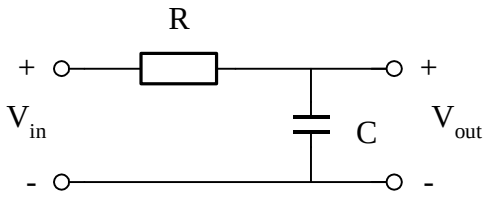

Figure 4.23: A passive voltage-to-voltage integrator.

(a) Determine the transfer function of this integrator.

(b) Draw, equivalent to the depicted integrator, an i-i integrator.

(c) What is the transfer of this integrator?

(d) Give a cascade of three i-i integrators and a cascade of three v-v integrators. You may neglect input and output loading effects.

(e) What is the difference between the two cascades?

2. A first step in approximating a nullor is the use of a voltage-controlled current source.

(a) Derive the chain matrix of the voltage-controlled current source depicted in figure 4.4.

(b) Under what condition is this voltage controlled current source equal to the nullor?

3. Determine for the ideal voltage and current source the corresponding small-signal models.

4. Derive the expression for r_π by directly using:

$$r_\pi = \left(\frac{dI_B}{dV_{BE}} \right)^{-1} \tag{4.26}$$

5. A figure of merit for transistors is the transit frequency. This is the frequency at which the current-gain factor becomes one.

(a) Determine the current-gain factor of the bipolar transistor by using the model of figure 4.12.

(b) Sketch the transit frequency as a function of the collector bias current, I_C.

6. Given the signal schematic of transistor Q_1 in figure 4.24. For the transistor

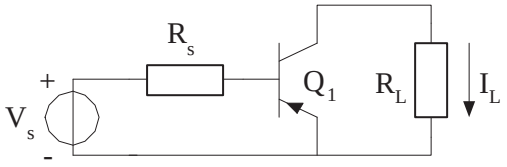

Figure 4.24: A transistor circuit

holds: $I_S = 10^{-14}A, V_A = 40V, kT/q = 25mV$. Q_1 is biased (with circuitry that is not shown) such that $|I_C| = 0.2mA$, $V_{CB} = -10V$ and $\beta_f = 200$. Further, $R_s \approx 25k\Omega$ and $R_L \approx 100k\Omega$.

(a) Calculate the transfer $|I_L/V_s|$.

7. Given the circuit in figure 4.25. For the diode the following values hold:

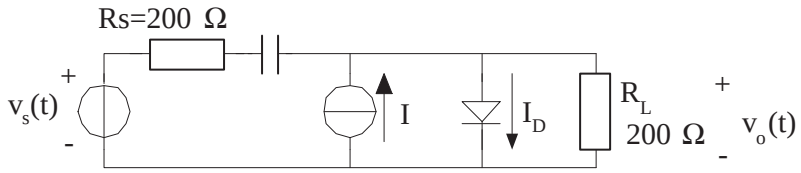

Figure 4.25: A diode circuit

$$I_S = 2.5 \cdot 10^{-12}A \tag{4.27}$$

$$kT/q = 25mV \tag{4.28}$$

The signal source is given by:

$$v_s(t) = 30\sin(\omega t)mV \tag{4.29}$$

For the current source it holds:

$$I = 0.25mA \tag{4.30}$$

The impedance of the capacitor is negligibly small.

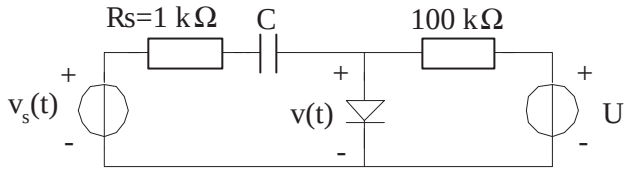

Figure 4.26: A diode circuit

(a) Calculate the peak-peak value of the output voltage $u_o(t)$.

8. Given the circuit in figure 4.26. For the diode the following values hold:

$$I_S = 10^{-14}A \tag{4.31}$$

$$kT/q = 25mV \tag{4.32}$$

Voltage source U is such that the diode bias current is $I_D = 0.1mA$. The signal source is given by:

$$v_s(t) = 10\sin(\omega t)mV \tag{4.33}$$

The impedance of the capacitor is negligibly small.

(a) Calculate the peak value of the diode voltage $v(t)$.

9. Given the amplifier circuit in figure 4.27. For the transistors the following

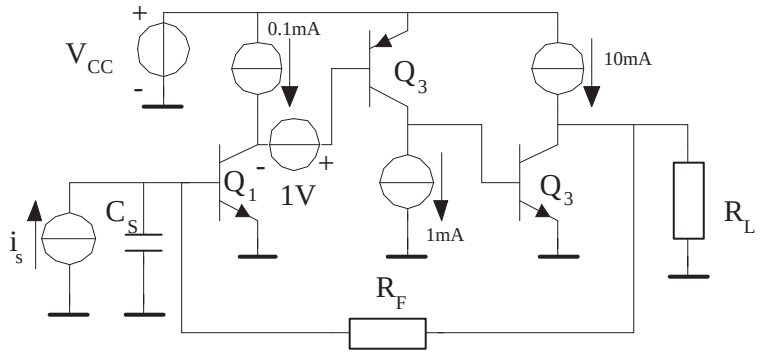

Figure 4.27: A amplifier circuit

values hold:

$$\beta_{1,2,3} = 100 \tag{4.34}$$

$$kT/q = 25mV \tag{4.35}$$

$$C_{\pi_1} = 100fF \tag{4.36}$$

$$C_{\pi_2} = 300fF \tag{4.37}$$

$$C_{\pi_3} = 1pF \tag{4.38}$$

(a) Draw the small-signal diagram and determine the element values for the transistors.

10. Given the differential pair in figure 4.28. Both transistors are equally biased.

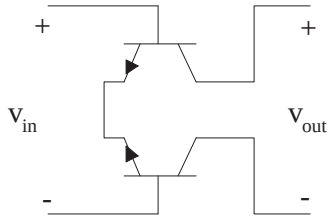

Figure 4.28: A differential pair

(a) Determine the small-signal diagram.

(b) Simplify the diagram of the previous question assuming that the differential pair is driven symmetrically.

11. Given the active saturation integrator of figure 4.29 This integrator imple-

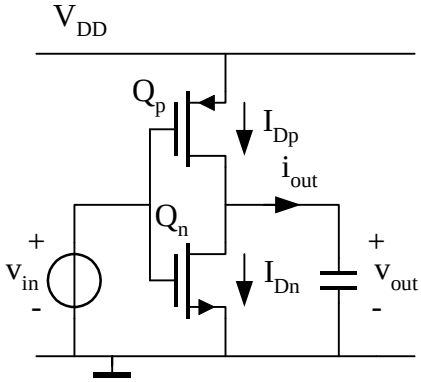

Figure 4.29: An active saturation integrator.

mentation has the advantage that the nonlinearity of the MOS device can cancel upto a large extent. This cancellation is the topic of this exercise.

The input voltage equals:

$$v_{in} = V_{in} + v_s \tag{4.39}$$

in which V_{in} is the bias voltage and v_s is the signal voltage.

(a) What condition such be fulfilled for Q_n and Q_p such that they have equal bias currents.

(b) Determine the combined transconductance of the two transistors.

(c) Determine the second-order distortion in the relation $i_{out} = \tilde{f}(v_{in})$. (Hint: use second term of the Taylor expansion of the nonlinear circuit, see equation (4.4))

(d) What condition must be met to cancel this distortion term?

12. Consider the active *triode* integrator of figure 4.30.

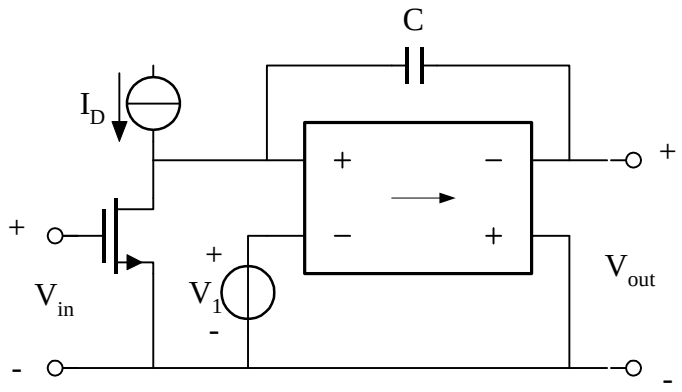

Figure 4.30: An active triode integrator.

(a) Determine for the MOS device in the triode region the small-signal circuit comprising g_m and r_d. Be clear in the sign conventions.

(b) Determine, using the small-signal circuit of the previous question, the transfer of the integrator.

(c) A filter is designed using several of those integrators. How can the filter be tuned making explicitly use of the triode region? Explain!

Chapter 5

Nullor synthesis

5.1 Introduction

The key building block in electronic design is the nullor. In the previous chapters the nullor frequently popped up as a desired building block. The basic function of the nullor is to supply gain, ideally infinite. The symbol of the nullor and its sign conventions are depicted in figure 5.1. Note that the output-current reference

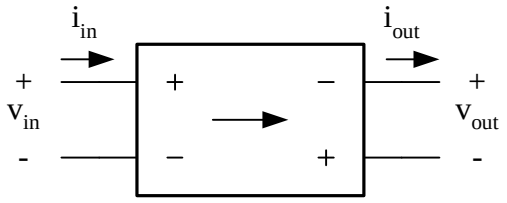

Figure 5.1: The nullor symbol and the sign conventions

direction is defined to be outward. This in contrast to the common definition which defines an inward direction (like for the input port) as the reference direction for the output current, i.e. both the input and output ports are treated equally. The convention of figure 5.1 is more convenient for designers, as in the case of cascading two-ports, no "minus sign" is required.

The nullor is an ideal element. The input voltage, v_{in}, and the input current, i_{in}, are by definition zero. These zero-conditions are not fulfilled automatically in a network. The values of the output voltage, v_{out}, and the output current, i_{out}, of the nullor become such that these input conditions are met. This implies that useful application of a nullor is always in a feedback situation. The input nullor conditions are *requirements* and not *enforcements*, so the nullor input is not a short circuit ($v_{in} = 0$) nor a open ($i_{in} = 0$).

Starting at the definition of the nullor, i.e., the input voltage of the nullor and

the input currents to the nullor equal zero, the chain matrix is readily obtained to be:

$$\begin{pmatrix} v_{in} \\ i_{in} \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} v_{out} \\ i_{out} \end{pmatrix} \tag{5.1}$$

This chain matrix directly follows from the definition. Thus the nullor has by definition a chain matrix filled with zeros.

Chain parameters have the reciprocal value of the transfer parameters. So, the nullor has infinite gain. Thus a nullor implementation is always an active circuit, i.e. comprising a power source. In the subsequent sections the design of a nullor implementation is treated. First, in section 5.2 attention is paid to what basic building blocks are preferably used to obtain the best nullor approximation. The nullor has got infinite gain, whereas single amplifying stages have limited gain. So, a cascade of single stages may be required to obtain sufficient gain. How to compose a cascade of stages is discussed in section 5.3. The nullor is ideal and thus adds no noise to a signal, has got no bandwidth limitation, and does not distort a signal. Practical circuits are not that perfect. Section 5.4 discusses how these quality aspects relate to the nullor implementation and how a design should be organized such that these quality aspects can be optimized separately.

It should be noted that throughout this chapter, when transistors are depicted, the bias circuitry is omitted. Thus, with a symbol of a transistor its small-signal behavior is meant.

5.2 Basic building blocks

Goal in the implementation of a nullor is that the chain matrix of the designed implementation approached the null-matrix as close as required. Thus to select basic building blocks their chain matrices should be evaluated. Firstly, single transistor stages are evaluated and subsequently, balanced configuration are dealt with.

5.2.1 Single-transistor stages

Figure 5.2 shows the possible stages when using a single transistor in the case of bipolar technology and MOS technology. So, a transistor is a three terminal device, whereas the nullor is a two-port with four terminal. When using a single transistor as an amplifying stage, always the input and output will have one terminal in common. Which terminal is the "common" terminal is put in the name of the stage, i.e. CE means common emitter and CD means common drain.

For each of the stages the chain matrix can be derived. The small-signal diagrams used for calculating the chain matrix are depicted in figure 5.3. Using

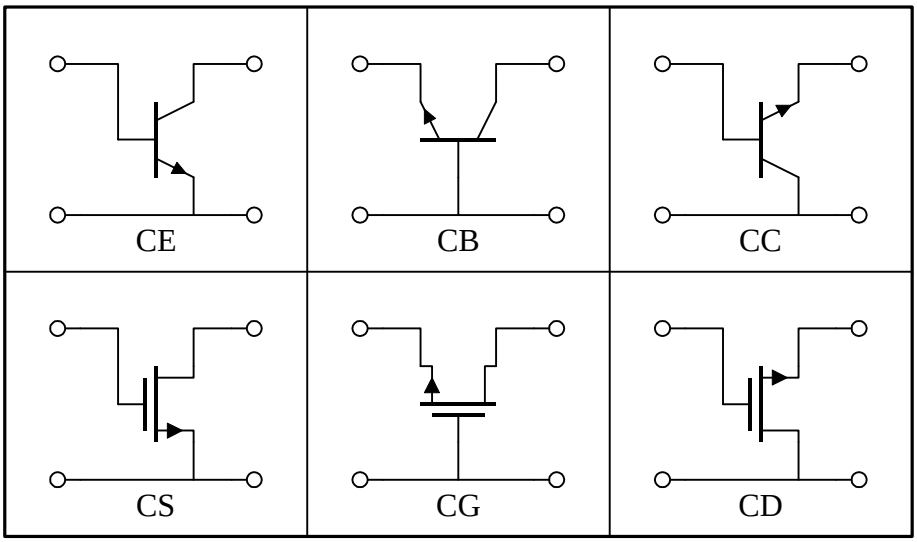

Figure 5.2: Signal diagram of amplifying stages composed of a single transistor, Bipolar and MOS technology.

these small-signal diagrams for the amplifying stages of figure 5.2 yields the chain matrices as listed in figure 5.4. Looking to the matrices as depicted in figure 5.4, several observations can be made.

- First of all, when comparing the chain matrix of the CE (CS) stage with the chain matrix of the CB (CG) and CC (CD) stage, it is clear that they are almost the same. Except a minus sign, only one chain parameter is different. For the CB (CG) stage parameter $D = 1$ and for the CC (CD) stage $A = 1$. The parameters A and D of the CE and CS stage are much smaller than one and thus it can be concluded that, when the smallest chain parameters are the goal, the CE and CS stage approach that the best. Therefore:

  > The CE and CS stage are the basic building blocks or basic amplifying stages comprising one transistor.

  The other stages can be treated as the CE or CS stage but with local non-energetic feedback. This feedback sets one of the chain parameters to one as can be seen in figure 5.4.

- The chain matrix of the MOSFET stages are more simple than those of the bipolar stages. However, this is not generally true. In this section the DC chain matrices are considered and because of the absence of a DC gate current, the chain parameters relating to i_{in} become zero! When AC chain matrices would be considered, the MOSFET chain matrices get the same complexity as the corresponding bipolar chain matrices.

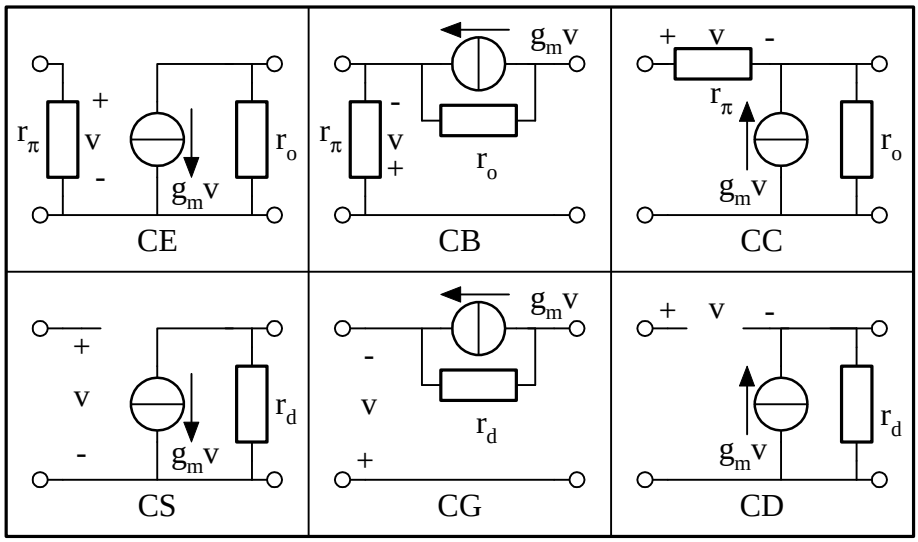

Figure 5.3: DC small-signal diagram for the bipolar transistor and the MOSFET transistor in their possible configurations.

$\begin{pmatrix} -\frac{V_T}{V_A} & -\frac{V_T}{I_C} \\ \frac{-I_C}{\beta_F V_A} & \frac{-1}{\beta_F} \end{pmatrix}$	$\begin{pmatrix} \frac{V_T}{V_A} & \frac{V_T}{I_C} \\ \frac{I_C}{\beta_F V_A} & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & \frac{V_T}{I_C} \\ \frac{I_C}{\beta_F V_A} & \frac{-1}{\beta_F} \end{pmatrix}$
CE	CB	CC
$\begin{pmatrix} -\frac{1}{r_d g_m} & -\frac{1}{g_m} \\ 0 & 0 \end{pmatrix}$	$\begin{pmatrix} \frac{1}{r_d g_m} & \frac{1}{g_m} \\ 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & \frac{1}{g_m} \\ 0 & 0 \end{pmatrix}$
CS	CG	CD

Figure 5.4: Chain matrices for the stages of figures 5.2.

The single-transistor stages have one essential limitation: the terminal that is common for the input and output. This is illustrated by the current amplifier with a single transistor implementation for the nullor, see figure 5.5. Clearly, because of the common-emitter terminal the load resistor, R_L, is shorted!. Placing the transistor up-side down, yields that feedback resistor R_1 is shorted. Conclusion, based on this single transistor it is not possible to implement a current amplifier,

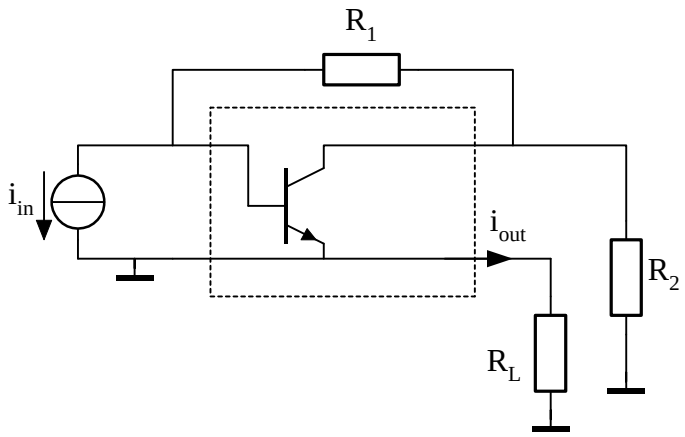

Figure 5.5: A single transistor nullor implementation for a current amplifier.

other amplifying stages need to be found. As all the possible single-transistor stages are evaluated, the need for an evaluation of two-transistor stages arises.

5.2.2 Two-transistor stages

Two-transistor amplifying stages can be found by combining two transistors such that the chain matrix of the new stage is equal or slightly different compared with the single-transistor stage. Visualizing a single transistor as a two-port, combinations can be found by combining two ports. Some commonly used combinations that can be obtained in such a way are depicted in figure 5.6. Assuming that the

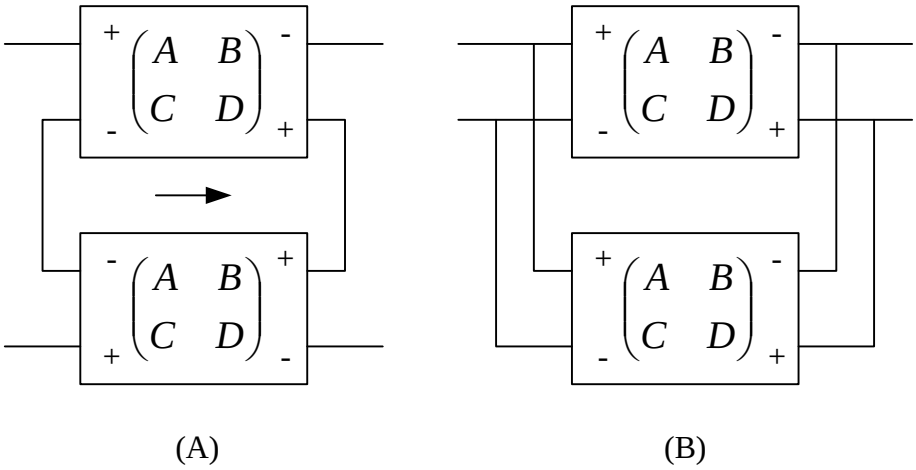

Figure 5.6: Some combinations of two-ports to end up with closely related new two-ports. (A) Anti-series connection (B) Parallel connection

chain matrix of the single two port is given by:

$$K_{single} = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \tag{5.2}$$

then the chain matrix of the anti-series connection, K_{as}, is given by:

$$K_{as} = \begin{pmatrix} A & 2B \\ C/2 & D \end{pmatrix} \tag{5.3}$$

and the chain matrix of the parallel connection, K_p, equals:

$$K_p = \begin{pmatrix} A & B/2 \\ 2C & D \end{pmatrix} \tag{5.4}$$

Clearly, a close relation exists between the chain matrix of the single stage and the chain matrices of these two combinations. The only differences are the factors 2 and 1/2. Thus, when for the two-ports stages are used with the lowest chain parameters, the combination will have chain parameters close to those lowest chain parameters. Replacing the two-port ports by CE stages yield the two stages as depicted in figure 5.7. Transistor combinations like the one depicted in figure 5.7

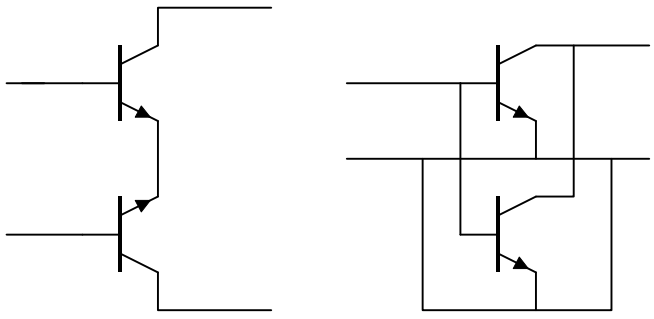

Figure 5.7: Combining CE stages in A) Anti-series connection (B) Parallel connection

are commonly used. The parallel connection of two transistors is used to enlarge the current capability of the amplifying stage. The anti-series stage, or more often called the differential pair, is used when, for instance, the three-terminal CE stage can not be used, as was illustrated in figure 5.5. Because of the anti-series connection of the inputs and outputs, the differential pair has become a four terminal stage. The current amplifier in which the nullor is now implemented with a single differential stage is depicted in figure 5.8.

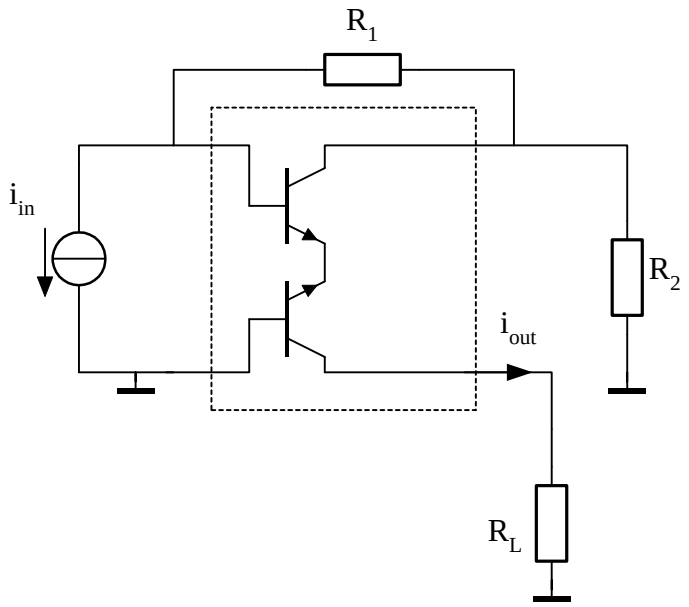

Figure 5.8: A current amplifier with a differential stage implementing the nullor.

The chain matrix of the differential pair is easily found by considering the relation to the chain matrix of the CE stage. It is given by:

$$K_{as} = \begin{pmatrix} -\frac{V_T}{V_A} & -2\frac{V_T}{I_C} \\ -\frac{I_C}{2\beta_F V_A} & -\frac{1}{\beta_F} \end{pmatrix} \tag{5.5}$$

This means, that besides the basic single stages, CE and CS stage, also the differential versions of those stages can be used as basic building block. The basic stages are depicted in figure 5.9. Of course, also the P-type transistors can be used to create basic amplifying stages.

As the chain matrix of the differential pair is closely related to the chain matrix of the CE stage, it is to be expected that also the small-signal diagram of the differential pair is closely related to the small-signal diagram of the CE stage. Indeed, they look very similar. Figure 5.10 depicts the small-signal diagram of the differential pair. Its derivation is left as an exercise at the end of this chapter. The main difference is found in the fact that also the small-signal diagram is a four terminal network. Consequently, when the sign of the nullor-implementation gain is not correct, the output terminals of differential pair can be interchanged to create a sign reversal. Further, in the small-signal diagram it can be seen that $B_{as} = 2B_{ce}$ as the transconductance of the differential pair is half compared with the single transistor.

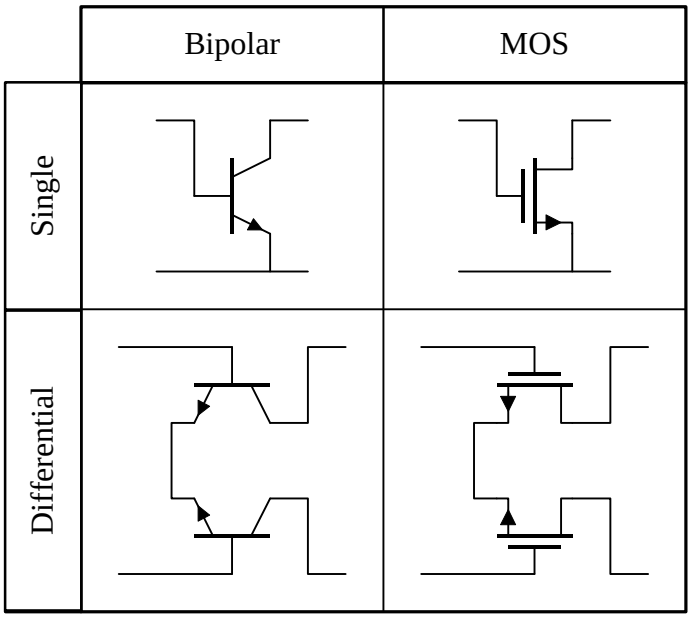

Figure 5.9: The basic stages that can be used for implementing a nullor in a bipolar or MOS technology.

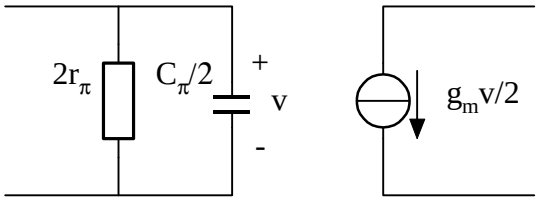

Figure 5.10: The small-signal diagram of the differential pair. Small-signal variables refer to the variables of a single transistor.

5.3 Cascade of stages

The previous section discussed what the optimal basic amplifying stages are for implementing a nullor approximation. Depending on the amplifier one amplifying stage could be insufficient. Insufficient means that the chain parameters are not close enough to zero and thus the gain of the nullor implementation is not high enough. In the next chapter the reason for having a certain gain level for the nullor implementation is discussed in more detail. In this section attention is paid how the chain parameters of the nullor implementation can be made closer to zero by using additional stages.

The straightforward method to increase the gain of a nullor implementation and thus to make the chain parameters closer to zero, is to cascade amplifying stages. This cascading of stages is schematically depicted in figure 5.11. When cascading

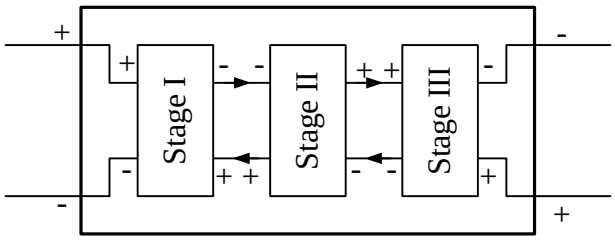

Figure 5.11: A cascade of amplifying stage in order to improve the nullor implementation.

the stages the output port of the first stage is connected to the input port of the second stage and so on. Important is that the output port is *directly* connected to the following input port. No additional elements should be in between. This is because of the highest performance that is obtained in this case. On this level of abstraction it may seem trivial. However, when connecting several transistor stages one may get confused in what to connect to what. Therefore, working with this level of abstraction can help in synthesizing correct nullor implementations.

The overall chain matrix of a multi stage nullor implementation is found by multiplying the corresponding chain matrices. This can be done as the reference direction of an output current is outward whereas the reference direction of the input current is inward and the input and output voltage have same reference direction. As an example, the chain matrix of a two-stage implementation is given:

$$\begin{aligned} K_{cascade} &= K_1 K_2 = \begin{pmatrix} A_1 & B_1 \\ C_1 & D_1 \end{pmatrix} \begin{pmatrix} A_2 & B_2 \\ C_2 & D_2 \end{pmatrix} \\ &= \begin{pmatrix} A_1A_2 + B_1C_2 & A_1B_2 + B_1D_2 \\ C_1A_2 + D_1C_2 & C_1B_2 + D_1D_2 \end{pmatrix} \end{aligned} \tag{5.6}$$

Each chain parameter of the cascade is a sum of two products of chain parameters. As the chain parameters are close to zero, the chain parameters of the cascade become even closer to zero. For instance, the DC chain matrix of the cascade of two CS stages is given by:

$$K_{CS,CS} = \begin{pmatrix} \frac{1}{g_{m1}g_{m2}r_{d1}r_{d2}} & \frac{1}{g_{m1}g_{m2}r_{d1}} \\ 0 & 0 \end{pmatrix} \tag{5.7}$$

The terms $g_m r_d$ are also called the voltage-gain factor of a FET device. This voltage-gain factor can be on the order of 100 to 1000. Clearly, the chain matrix of the cascade approaches the zero matrix closer than the one does of a single stage.

An example of a three-stage implementation is given in figure 5.12. The first stage is a differential PMOS stage, the second stage is a differential bipolar stage (N) and the third stage is a single CE stage (N). Note that each output

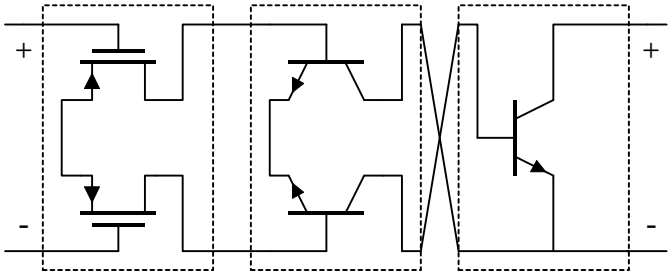

Figure 5.12: A three stage implementation of a nullor.

port is correctly connected to the subsequent input port. The overall polarity of the cascade was chosen such that an interchange of two output terminals of a differential pair was required.

5.4 Design for quality

In the previous section it was discussed how to implement a nullor approximation by using more amplifying stages. Which type of stage (single or differential) and what type of technology (bipolar or FET) should be chosen for optimal results was not treated. This requires a more detailed analysis of the noise, speed and distortion performance of a nullor implementation, which is beyond the scope of these lecture notes. However, in this section attention is paid to the *locations* in the circuit where the different quality aspects (noise, distortion and bandwidth) are determined. This gives the designer insight in how (where) to improve certain quality aspects of the circuit.

A properly designed nullor implementation comprises only amplifying stages: CE, CS and the differential variant. As a result, when the signal level is depicted as a function of the location in the nullor implementation, a diagram like depicted in figure 5.13 is obtained. This diagram is the key to localizing the noise performance, the distortion performance and the bandwidth performance. As a result of the use of amplifying stages only, the signal at the input is the smallest whereas the signal at the output is the largest.

Noise effects the signals dominantly at locations where the signals are the smallest. Consequently, the noise performance is determined by the input stage:

> The noise performance is localized at the input of the nullor implementation.

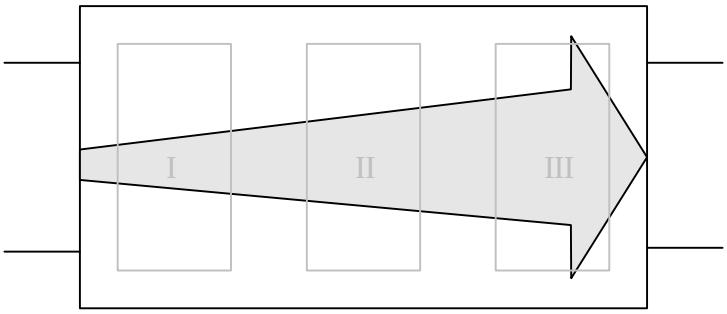

Figure 5.13: The signal level in a cascade of amplifying stages as a function of the location in the cascade.

Analogous, strong distortion, like clipping, is likely to happen at the output, since in a properly designed amplifier, the signals are the largest in the output stage.

> The clipping-distortion performance is localized at the output of the nullor implementation.

The fact that these two properties can be assigned to different parts of the amplifier makes an assumption of orthogonality for these two valid.

The third property, bandwidth, cannot be assigned to any particular part of the amplifier. This is because the nullor implementation is used in a feedback loop and the dynamic behavior of a closed loop is determined by the whole loop.

> The bandwidth performance is determined by the whole nullor implementation

Both the input stage and the output stage contribute to the bandwidth performance of the complete amplifier too. Therefore noise and clipping-distortion optimization always interfere with bandwidth optimization. For this reason noise and clipping-distortion optimization should be performed *before* bandwidth optimization. During these optimizations bandwidth is not taken into account. It is assumed that this can be made correct later outside the first and last stage. The contributions of the first and last stage to the bandwidth are taken for granted during bandwidth optimization. They are taken into account, but in principle not changed.

Since bandwidth calculations are very tedious, before the actual bandwidth of the amplifier is determined, first a prediction of the bandwidth is made by way of a fairly simple calculation. When the circuit passes this test, the actual bandwidth calculations are performed, so bandwidth optimization will consist of two stages:

- bandwidth estimation

- bandwidth optimization or *frequency compensation*

For ease of calculation bandwidth calculations will be performed with a very simple transistor model at first. Later the model is refined gradually after each successful calculation until the full transistor model is applied. Again this is to detect unfeasible solutions before extensive calculations are performed. Also this gradual refinement of models provides a lot of insight for the designer in the exact cause of appearing problems in the amplifier.

5.4.1 Design of the first nullor stage: noise.

The first property that is optimized is the noise performance. In figure 5.14 the nullor configuration is shown that is used for this optimization step. The first

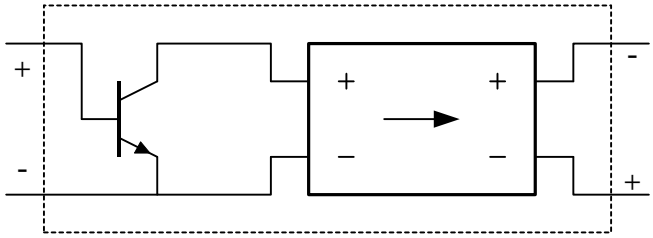

Figure 5.14: The model for the active circuit used for noise optimization.

stage is put in front of the nullor. When the gain of the first stage is sufficient, the noise contribution of the rest of the circuit can be neglected. Therefore as a first stage a CE or a CS stage should be used, or their differential version, because they offer the most gain.

The advantage of starting with the noise optimization is the fact that it is possible to model the circuitry following the first stage by a nullor. In this way, the other two criteria, bandwidth and clipping distortion, remain ideal. The nullor will supply infinite power if necessary, its infinite gain results in infinite bandwidth and zero distortion for the feedback loop. The designer only has to be concerned with noise in this stage. As long as the implementation of the nullor is good enough, the performance of the first stage with respect to bandwidth and distortion is of no importance.

An orthogonal design of the noise performance is therefore possible.

The clipping distortion, though also localized at a specific place in the amplifier, is better not taken as the first aspect to be optimized. This is because it is not possible, like it was for noise, to take the stage concerned and cascade it with a nullor to make the optimization independent of the noise and bandwidth aspects. The nullor would be placed in front of the stage concerned and with its infinite gain, it would reduce the distortion caused by this last stage to zero, irrespective

of the implementation of that stage. So to keep calculations simple as long as possible, by keeping a nullor in the design as long as possible, the design should start with noise optimization.

5.4.2 Design of the last nullor stage: distortion.

When the first stage has been designed for noise, the last stage should be designed for clipping distortion. Bandwidth can be manipulated anywhere in the amplifier, so any restriction on bandwidth caused by the last stage, can be corrected elsewhere in the amplifier. The nullor implementation used, can be as depicted in figure 5.15. The largest signals appear in the last stage. Therefore clipping is most likely to

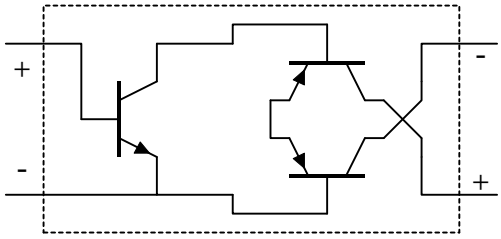

Figure 5.15: The model for the active circuit used for clipping distortion optimization.

happen in this stage and probably most of the power is consumed here. To make the assumption true, the gain of the last stage should be as large as possible. The magnitude of the signals in the preceding stage is reduced by the gain of the last stage. Therefore it is unlikely that that stage will cause distortion problems if the last one does not.

Apart from the clipping distortion, there is also the weak distortion caused by the non-linearity of the devices. The dominant contribution of this type of distortion may not be from the last stage. This type of distortion can be reduced by increasing the loop gain, a measure that can be taken anywhere in the circuit and that is not in conflict with bandwidth or noise optimization.

5.4.3 Design of intermediate nullor stages: bandwidth.

The exact calculation of the bandwidth of an amplifier is very complicated. It is too complicated to waste it for detecting non-feasibility of a solution. Therefore first a prediction will be made on the possible bandwidth of a solution that is not based on complicated calculations. By model simplification, calculations are reduced even further.

The two stages that have resulted from the noise and the distortion optimization form the first "guess" for the complete active circuit. The simple models are

used. A negative-feedback amplifier results that at least meets the noise and the clipping specs.

For this configuration a prediction of the maximum bandwidth can be found. When this upper limit is too low it can be increased by means of, for instance, additional intermediate stages. When the prediction of the upper limit is high enough the real design of the bandwidth can be done and a correct nullor implementation is obtained.

5.5 Exercise

1. Determine the chain matrix of a nullor by directly applying the definition of the chain parameters.

2.

(a) Draw a current follower using a nullor for the active part.

(b) Implement the nullor by one single-transistor stage.

(c) Determine the chain matrix of the configuration of the pervious questions.

(d) What is your conclusion?

(e) Repeat this exercise for a voltage follower.

3. Given the combination of amplifying stages, depicted in figure 5.16. It can be

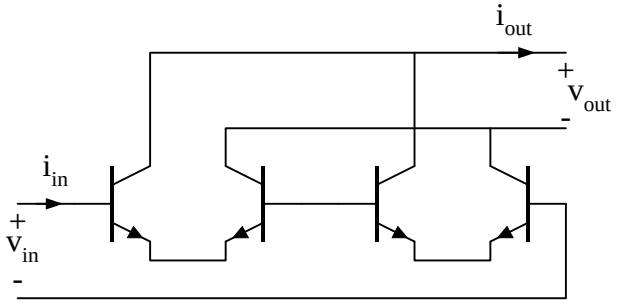

Figure 5.16: A combination of amplifying stages.

assumed that the four transistors are identical and have a chain matrix equal to:

$$K = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \tag{5.8}$$

(a) Determine the chain matrix of the combination of transistors depicted in figure 5.16 (Hint: use hierarchy).

4. The chain matrix of a differential pair is closely related to the chain matrix of the single-transistor stage (CE, CS).

(a) What is the difference between the chain matrix of a differential pair and its single-transistor counterpart (use the DC chain matrix)?

(b) How can those chain matrices be made identical? Discuss this for bipolar as well as for FET technology.

5. Small-signal diagrams are used to analyze the behavior of, for instance, non-linear transistors.

(a) Derive the low-frequency (DC) small-signal diagram of a PNP bipolar transistor.

(b) Compare it to the low-frequency (DC) small-signal diagram of a NPN bipolar transistor.

(c) What is your conclusion? Explain your findings.

6. Figure 5.17 depicted three nullor implementations.

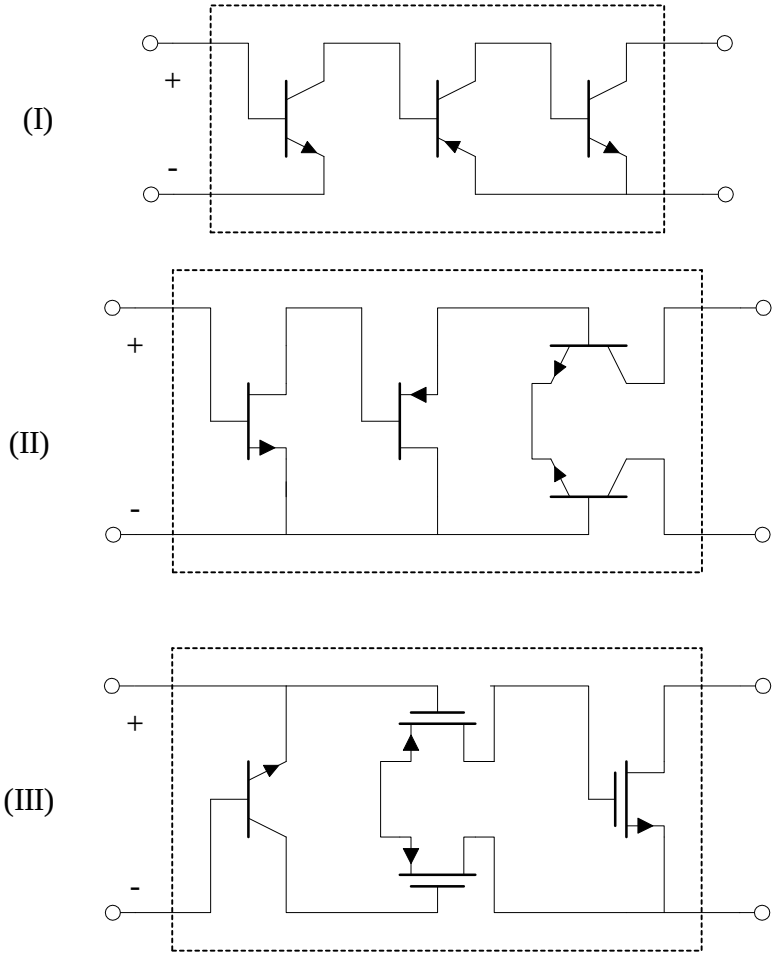

Figure 5.17: Three arbitrary nullor implementations

(a) Which of the implementations of figure 5.17 are correct and which incorrect. Motivate your choices.

(b) Indicate in the figures the intended output polarities. A reference polarity for the input is given in the figure.

7. Figure 5.18 depicts two two-stage nullor implementations yielding a inversion between input and output as indicated.

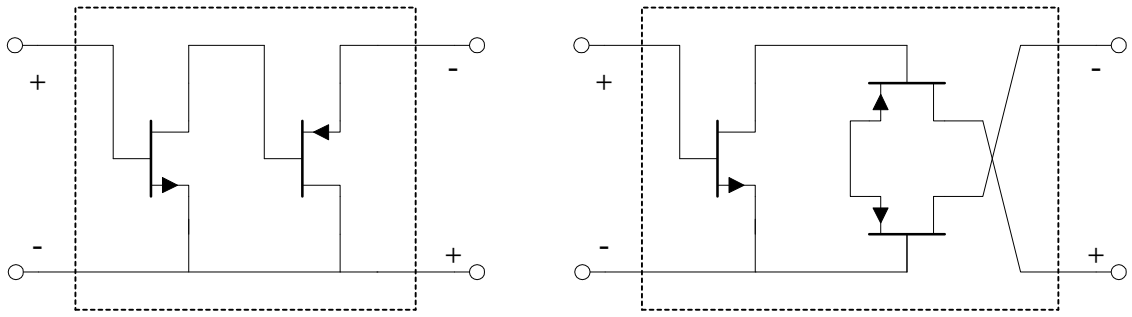

Figure 5.18: Two two-stage nullor implementations yielding inversion between input and output.

(a) Are both implementations correct implementations?

(b) No: describe what is wrong and indicate what the consequence is on the performance of the nullor implementation.

(c) Yes: which one do you prefer? Motivate!

8. Given the chain matrix of a CE-stage:

$$\begin{pmatrix} A_{CE} & B_{CE} \\ C_{CE} & D_{CE} \end{pmatrix} \tag{5.9}$$

(a) Determine the elements of this matrix with the assumption that the frequency dependent elements and the Early effect can be ignored

The stages depicted in figure 5.19 comprise one or more transistors

(b) Determine for the depicted combinations the chain matrices as a function of A_{CE}, B_{CE}, C_{CE} and D_{CE}. It can be assumed that all the transistors are identical.

9.

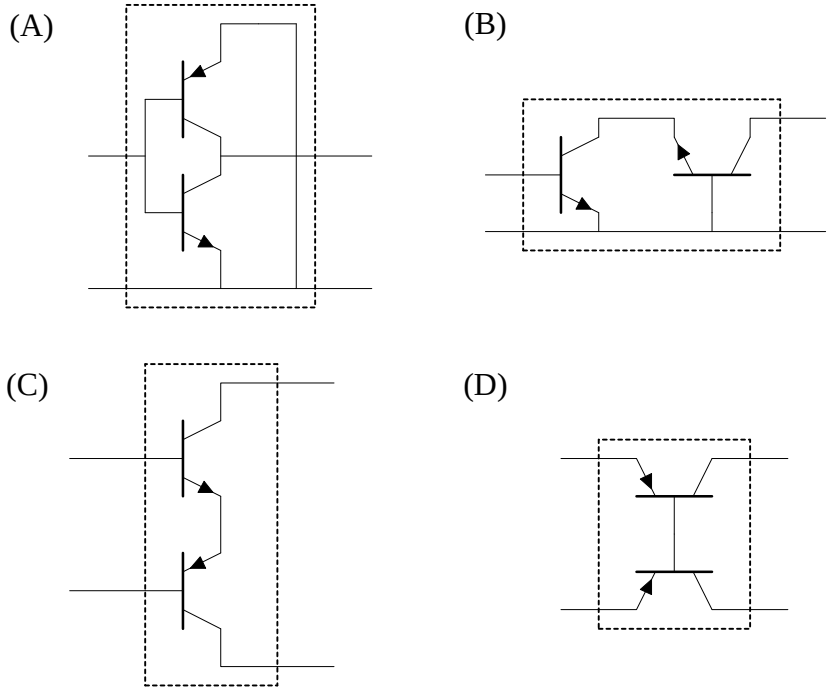

Figure 5.19: Four transistor combinations.

(a) Determine for the stages depicted in figure 5.19 the small-signal equivalents.

10. Given the four amplifiers in figure 5.20.

(a) Determine the transfer of each of these amplifiers by using the nullor constraints.

(b) Give for each of the nullors a two-stage implementation using MOSFETs

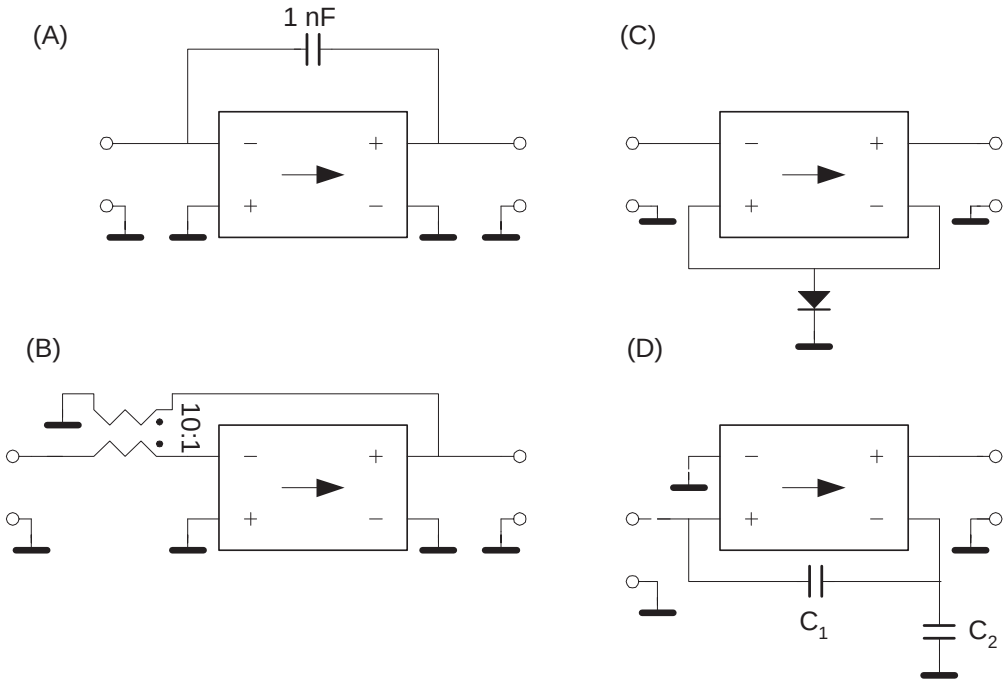

Figure 5.20: four feedback amplifiers

Chapter 6

Accurate Amplification

6.1 Introduction

In the previous chapters several design steps were treated in the design of active filters. Clearly, the chapters dealing, for instance, with the noise performance of circuits and the implementation of the nullor, are more general applicable than to filter design only. The same goes for this chapter. In this chapter a model is presented that fits perfectly to the design of negative-feedback circuits.

Feedback is applied when a highly accurate transfer exhibiting gain or a source or load independency is required using the standard electronic components. The standard electronic components are either active and relatively inaccurate or passive and relatively accurate. By means of feedback, these components are combined such that an accurate transfer is obtained with the characteristics of active devices, i.e. gain, etc. This is depicted in figure 6.1. When the active elements

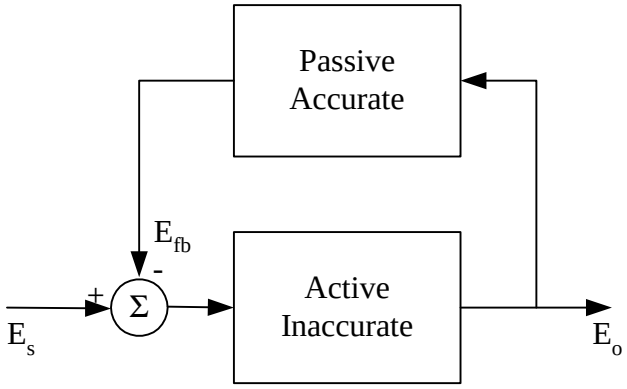

Figure 6.1: Using feedback to combine the quality of passive and active components to obtain highly accurate transfers.

supply sufficient gain, i.e. approaches a nullor, the transfer is determined by the

feedback network. This is because in the situation of high gain, the input signal of the active part approaches zero. This means that the feedback signal, E_{fb}, approximately equals the input signal E_s. As the output signal E_o is related to E_{fb} via the transfer of the feedback network this means that the ratio E_o/E_s is given by $1/\beta$ in which β is the transfer of the feedback network.

6.2 Asymptotic-gain model

The derivation of the asymptotic-gain model starts with the application of the superposition principle to the negative-feedback system. Therefore, two sources need to be distinguished in the negative-feedback system:

- the independent *signal* source, E_s;
- a *dependent* source modelling the active part, E_c

in which the dependent source is assumed to be described by the following relation:

$$E_c = AE_{cc} \tag{6.1}$$

in which E_{cc} is the control input and A the constitutive parameter. Applying superposition to obtain an expression for the output signal, E_o, yields:

$$E_o = \rho E_s + \nu E_c \tag{6.2}$$

in which ρ is the transfer from the signal source to the output when $E_c = 0$ and ν is the transfer from the dependent source to the output under the condition that the input signal is zero. In this equation the value of the dependent source is still an unknown and for getting a fully described system, superposition should also be applied to the dependent source. To make the constitutive parameter explicit in the model, superposition is applied to obtain an expression for the control input of the dependent source, yielding:

$$E_{cc} = \xi E_s + \beta E_c \tag{6.3}$$

in which ξ is the transfer from the signal source to the control input when the dependent source is zero and β is the feedback modelling the transfer from the dependent source to the control input assuming that the signal source is zero. Equations (6.1) - (6.3) can be graphically represented as depicted in figure 6.2. Using (6.1) - (6.3), the transfer of the negative-feedback system, A_t, can be described as:

$$A_t = \frac{E_o}{E_s} = \rho + \frac{\nu\xi A}{1 - A\beta} \tag{6.4}$$

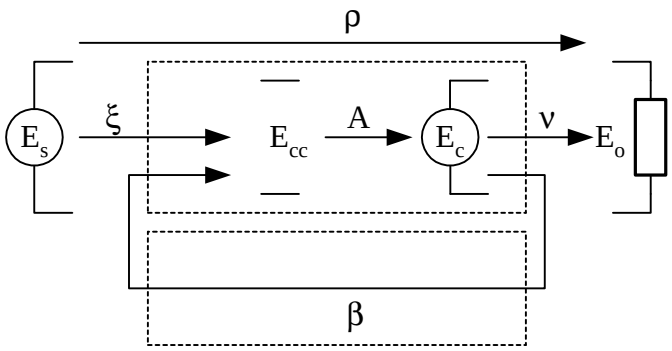

Figure 6.2: Graphical representation of the superposition model of a negative-feedback system.

In a synthesis method it is preferable to start from an ideal situation. In the case of a negative-feedback system, the ideal situation is that the active part is a nullor. When modelling the nullor by means of a dependent source, the constitutive parameters will be infinite. Thus, the gain of the negative-feedback system in the ideal situation, A_∞, is found as a limiting case for $A \to \infty$ as:

$$\begin{aligned} A_{t\infty} &= \lim_{A\to\infty} A_t = \lim_{A\to\infty} \left(\rho + \frac{\nu\xi A}{1 - A\beta}\right) \\ &= \rho - \frac{\nu\xi}{\beta} \end{aligned} \tag{6.5}$$

The gain $A_{t\infty}$ is called the asymptotic gain, i.e. the gain of the amplifier in the ideal (gain of active part asymptotically going to infinite) situation. Rewriting the gain of the negative-feedback system, in which the asymptotic gain appears, is given by:

$$A_t = A_{t\infty}\frac{-A\beta}{1 - A\beta} + \rho\frac{1}{1 - A\beta} \tag{6.6}$$

In this expression the term A is not present individually, but always in a product with β. Looking to the diagram in figure 6.2 it is clear that the product $A\beta$ is the loop gain of the amplifier. Assuming the loop starts at the dependent source, then the loop goes on via the feedback path, β, to end up at the control port of the dependent source and, finally, via the the constitutive parameter, A, the starting point is reached again. Often, this loop gain is written as a single parameter, like L.

Having a closer look at the expression, it appears that the gain, A_t, is a weighted sum of two transfers: $A_{t\infty}$ and ρ. The two weighting factors depend only on the loop gain. In the case of infinite loop gain (the active part is a nullor), the gain equals $A_{t\infty}$ and in the case of zero loop gain (the active part is an open) the

transfer is given by ρ. Therefore, a better name for ρ is A_{t0}, i.e. the gain of the feedback system in the case that the loop gain is zero.

Applying these two changes to expression (6.6), yields:

$$A_t = A_{t\infty}\frac{-L}{1-L} + A_{t0}\frac{1}{1-L} \tag{6.7}$$

Note that the loop gain for proper functioning of the loop should be negative.

Clearly, the asymptotic-gain model fits perfectly on the synthesis of electronic circuits as presented in these lecture notes. Roughly speaking, the presented synthesis procedure comprises two steps:

1. Design of the feedback network, assuming the active part is a nullor.
2. Approximate a nullor by means of a cascade of amplifying stages.

These two steps, translated to the asymptotic-gain model, yields:

1. Design $A_{t\infty}$.
2. Design L such that it is sufficiently high.

6.3 Calculating loop gain

For the asymptotic-gain model two quantities need to be determined:

- $A_{t\infty}$
- L

The asymptotic gain of a negative-feedback system is relatively easy to obtain: assume the active part is a nullor and find for that configuration the transfer.

Calculating the loop gain of the negative-feedback system requires some specific attention. In the asymptotic-gain model the active part is modelled by means of a single dependent source and then the loop gain is A times β. However, most nullor implementations have more than one amplifying stage, resulting in more than one dependent source in the small-signal diagram. The question is: "How to calculate the loop gain in such a situation?"

Assume the simplified small-signal diagram of a negative-feedback amplifier as depicted in figure 6.3. The active part is high-lighted by means of the dotted box. Applying the asymptotic-gain model literally would mean that the active part is modelled as a single dependent source, with input v_{be1} and output $g_{m3}v_{be3}$. Subsequently, β and A should be calculated.

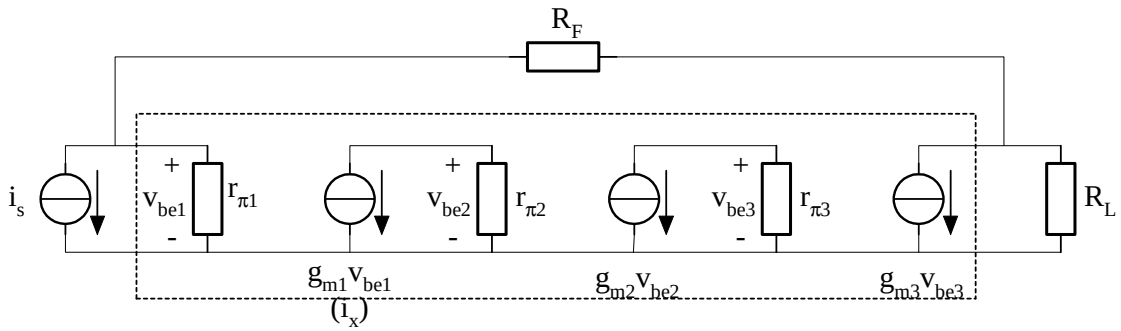

Figure 6.3: Calculating the loop gain of a negative-feedback amplifier.

However, a more direct method is to calculate the loop gain at once, i.e. calculating "L" directly. Therefore, somewhere the loop has to broken and the gain should be calculated between the two open ends assuming that the input signal (i_s) is zero. The loop is broken by making a dependent source in the loop, independent. In this way the topology of the network is not changed and an exact loop gain can be calculated. This in contrast with several methods found in literature. In those methods the loop is broken by disconnecting somewhere in the loop an element. For instance, when the loop would be broken by cutting R_F from the input, the impedance seen at R_L is changed as R_F is floating now.

In the amplifier of figure 6.3, each of the three independent sources can be chosen to break the loop at. Criterion is that when the constitutive parameter, of the dependent source that is made independent, is made infinite, the nullor conditions should be found at the input of the active part. In the case of a correctly broken loop, the constitutive parameter of the respective amplifying stage is in the loop. Thus, when making this parameter infinite, the loop gain becomes infinite, and the active part resembles a nullor. In the example presented in section 6.5 it is shown how an error can be made by choosing the wrong dependent source.

When the loop is broken, the transfer from the new independent source to the original controlling port should be calculated. The loop gain is then found by multiplying that transfer by the constitutive parameter of the original dependent source. The calculation can be done by using the MNA method, or often also by inspection using current and voltage division.

For the amplifier in figure 6.3, the first dependent source is assumed to be independent, i_x. The transfer from i_x to v_{be1} is readily obtained to be:

$$\frac{v_{be1}}{i_x} = r_{\pi 2} \cdot g_{m2} \cdot r_{\pi 3} \cdot g_{m3} \frac{R_L}{R_L + R_F + r_{\pi 1}} \cdot r_{\pi 1} \tag{6.8}$$

and the loop gain is given by:

$$L = -g_{m1} \frac{v_{be1}}{i_x} = -\beta_1 \cdot \beta_2 \cdot \beta_3 \frac{R_L}{R_L + R_F + r_{\pi 1}} \tag{6.9}$$

It is easily shown that when the loop is broken at dependent source 2 or 3 the same results are obtained.

6.4 Black's feedback model

In the previous sections the asymptotic-gain model was presented as a model which should be used for designing feedback circuits. A more commonly used feedback model is Black's model. Question then arises: "Why not using that model, it yields more or less the same kind of expressions?".

Main reason, is the fact that Black's model is defined at signal level using unilateral blocks. Those blocks have a gain that is independent of source and load. These kind of blocks are found, for instance, in control engineering. In electronics, however, circuits have to do with voltages and currents and consequently, loading effects of impedances play an important role.

The asymptotic-gain model has got its roots in the superposition principle. This superposition principle is still valid when considering voltages and currents, which makes the asymptotic-gain model suitable for application in electronic circuit design.

6.5 Example use of asymptotic-gain model

In this section an example is treated that shows how the asymptotic gain and the loop gain can be calculated for the asymptotic-gain model. The amplifier of figure 6.4 is studied. The gain of the amplifier is set by resistor $R_{feedback}$. Input signal

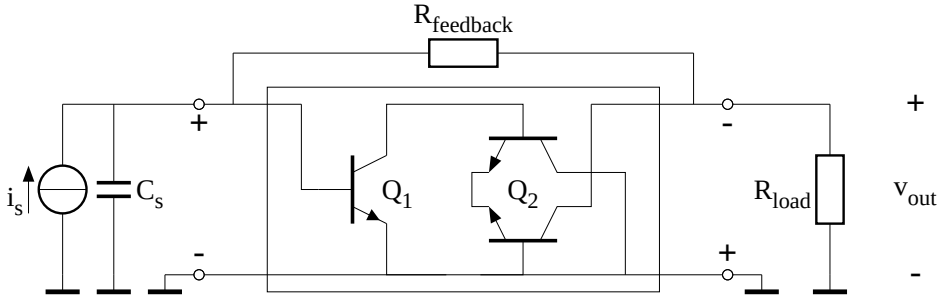

Figure 6.4: A transimpedance amplifier with a two-stage nullor implementation.

is current i_s and output signal is voltage v_{out}. The nullor is implemented by two stages. The first stage is a CE-stage whereas the second stage is a differential pair to guarantee a negative loop gain.

6.5.1 The asymptotic gain

To calculate the asymptotic gain, the active part is replaced by a nullor, see picture 6.5. The input current and input voltage of the nullor are zero and thus all the

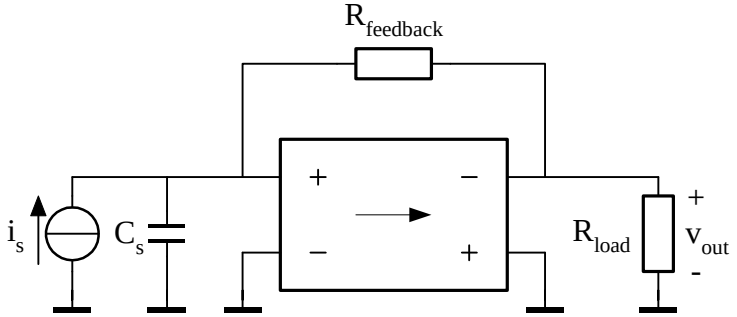

Figure 6.5: Determining the asymptotic gain by replacing the active part by a nullor.

signal current flows through the feedback resistor, yielding and output voltage equal to $-R_{feedback} \cdot i_s$. Consequently, the asymptotic gain is easily found to be:

$$A_{t\infty} = -R_{feedback} \tag{6.10}$$

6.5.2 The loop gain

For calculating the loop gain, the small-signal diagram of the amplifier is considered, see figure 6.6. For the differential pair a simplified small-signal diagram is

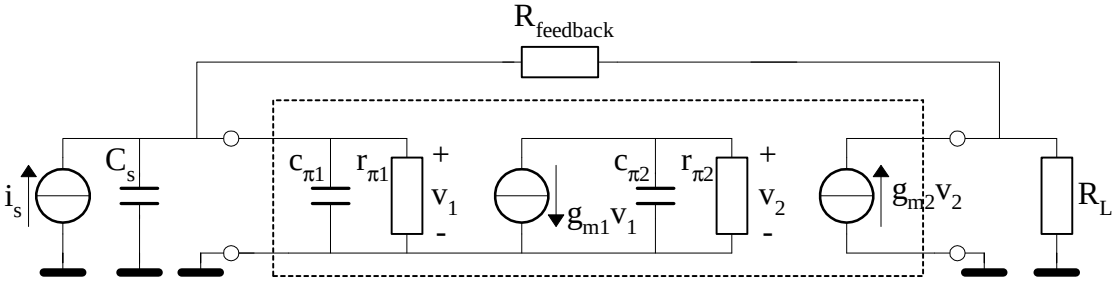

Figure 6.6: Small-signal diagram of the amplifier of figure 6.4

used. The derivation is depicted in figure 6.7. At the left side the small-signal diagram of the differential pair is depicted by drawing for both transistors their corresponding small-signal equivalent. For the case that both base currents are equal (which is a good assumption for a differential pair), no current flows through connection X. So, it can be removed from the diagram. The circuit which remains in that case is easily reduced to the diagram at the right side of figure 6.7. For the

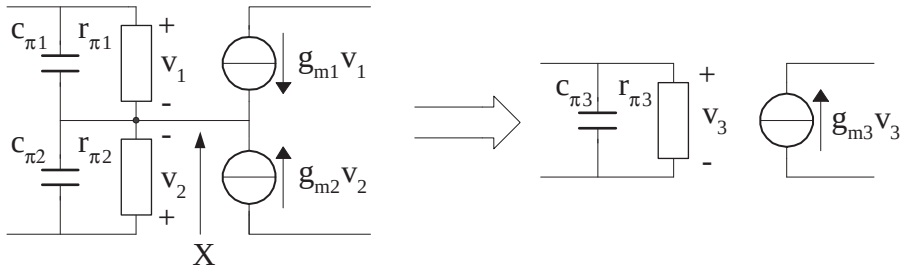

Figure 6.7: The small-signal diagram for a differential pair.

resulting diagram the input and output port do not have a terminal in common and thus an inverting as well as a non-inverting transfer can be realized with this stage. The relations between the elements are:

$$r_{\pi 3} = r_{\pi 1} + r_{\pi 2} = 2r_\pi \tag{6.11}$$

$$c_{\pi 3} = \frac{c_{\pi 1} c_{\pi 2}}{c_{\pi 1} + c_{\pi 2}} = \frac{c_\pi}{2} \tag{6.12}$$

$$g_{m3} = \frac{g_{m1} g_{m2}}{g_{m1} + g_{m2}} = \frac{g_m}{2} \tag{6.13}$$

in which the last equalities hold for the case of two identical transistors biased at the same current.

For calculating the loop gain, a dependent source has to be made independent. In this example, the dependent current source with current $g_{m1}v_1$ is assumed to be independent with output current i_x (see figure 6.8). The loop gain is found easily now. First the transfer from i_x to v_1 needs to be calculated. As in the real amplifier the relation between "i_x" and v_1 is g_{m1}, multiplying by g_{m1} yields the loop gain:

$$\text{loop gain} = \frac{V_1}{I_x} \cdot g_{m1} \tag{6.14}$$

It should be noted that the dependent source which is assumed to be independent for calculating the loop gain, cannot be chosen arbitrary. The criterion is that by making the dependent source independent, the overall loop should be broken. This can easily be checked by making the corresponding constitutive parameter infinite. In that case the loop gain should become infinite and at the input of the nullor implementation the nullor conditions arise. When an incorrect source is used, a local loop is broken and not necessarily the overall loop. In that case the loop gain of a different loop is calculated and of course also a different asymptotic gain is related to that loop.

This error is easily made when, for instance, for the differential pair the small-signal diagram of the left side of figure 6.7 is used and one of the two dependent

sources is assumed to be independent. The two transistors are anti-series connected and thus both have local feedback. In that case the gain of this local loop is calculated. This can easily be seen when the corresponding transconductance is made infinite. In that case the corresponding transistor acts as an ideal voltage or current follower (depending on which transistor is considered) and the loop gain of the overall loop is certainly not made infinite. The correct choice is using g_{m3} of the diagram at the right side of figure 6.7.

For the example we assume that the voltage-controlled current source of the input transistor is made independent. Figure 6.8 shows the small-signal diagram. First step is to calculate the transfer from i_x to v_1. This can be done by the MNA

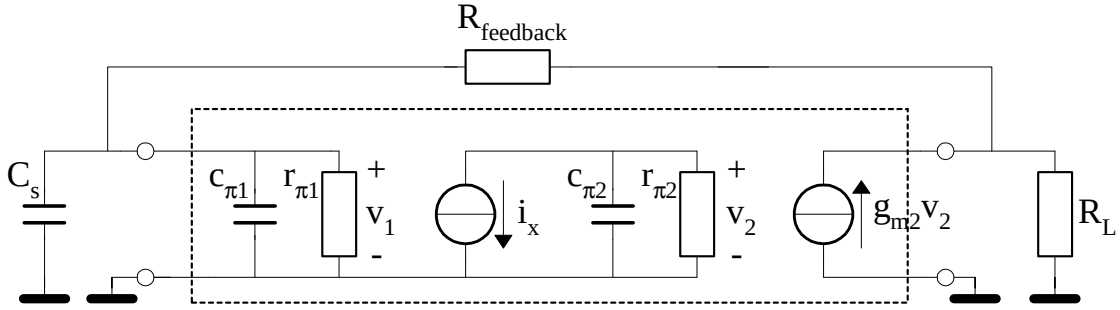

Figure 6.8: Small-signal diagram for calculating the loop gain; the voltage-controlled current source of the input transistor is assumed to be independent and the input signal is made zero

method but also by inspection: use current division at the several nodes and the gain of the transconductances:

$$\begin{aligned} V_1 = -I_x \quad &\cdot \quad \frac{r_{\pi 2}}{1 + s r_{\pi 2} c_{\pi 2}} \cdot g_{m2} \\ &\times \quad \frac{R_L}{R_L + R_{feedback} + \frac{r_{\pi 1}}{1 + s r_{\pi 1}(C_s + c_{\pi 1})}} \cdot \frac{r_{\pi 1}}{1 + s r_{\pi 1}(C_s + c_{\pi 1})} \end{aligned} \tag{6.15}$$

in which s is the Laplace variable. The loop gain is found by multiplying by g_{m1}. Simplifying the resulting expression yields:

$$\begin{aligned} L \quad &= \quad \frac{-\beta_1 \beta_2 R_L}{r_{\pi 1} + R_L + R_{feedback}} \\ &\times \quad \frac{1}{1 + s r_{\pi 2} c_{\pi 2}} \cdot \frac{1}{1 + s \frac{r_{\pi 1}(R_L + R_{feedback})}{r_{\pi 1} + R_L + R_{feedback}} \cdot (C_s + c_{\pi 1})} \end{aligned} \tag{6.16}$$

in which $\beta = g_m r_\pi$. From this expression the DC loop gain $[L(0)]$ and the poles

(p_1, p_2) are readily found to be:

$$L(0) = \frac{-\beta_1\beta_2 R_L}{r_{\pi 1} + R_L + R_{feedback}} \tag{6.17}$$

$$p_1 = \frac{-1}{2\pi r_{\pi 1} c_{\pi 1}} \tag{6.18}$$

$$p_2 = \frac{-(r_{\pi 1} + R_L + R_{feedback})}{2\pi r_{\pi 1}(R_L + R_{feedback})(C_s + c_{\pi 1})} \tag{6.19}$$

6.6 Exercises

1. With negative feedback the quality of passive elements (accuracy) and the quality of active elements (gain) can be combined.

(a) Discuss the problems of error feedforward when trying to design accurate amplification with passive and active elements.

(b) Discuss the problems that arise when error-compensation techniques are used for realizing active amplification.

(c) What is the essential difference between negative feedback and the two previous discussed methods?

2. Assume that the inaccuracy of the passive elements that can be used in the feedback network is 1%. The active part is implemented such that it approximates the nullor *well enough*.

(a) What would you consider in this case a reasonable minimum loop gain? Motivate your selection.

3. Given the three amplifiers in figures 6.9 - 6.10. For each of the amplifiers

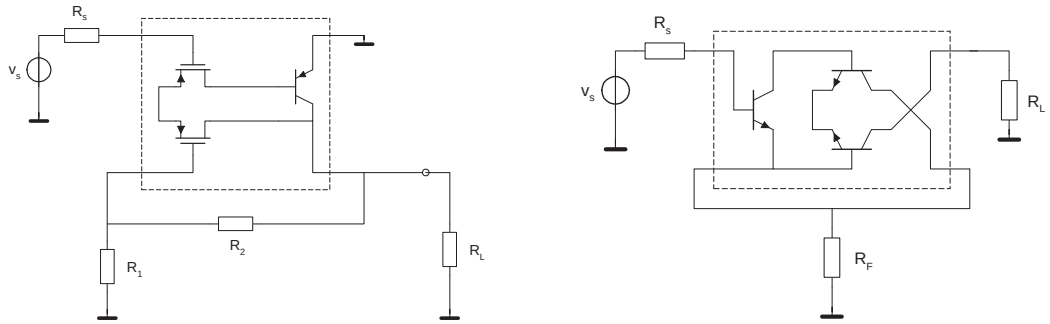

Figure 6.9: Amplifier 1 and 2

calculate (using simple small-signal models, i.e. g_m, r_π, c_π, c_{gs}):

(a) The Asymptotic-gain for the transfer for source to the load quantity

(b) The loop gain in terms of the small-signal parameters, source, load and feedback impedances.

4. Consider the active integrator as depicted in figure 6.11.

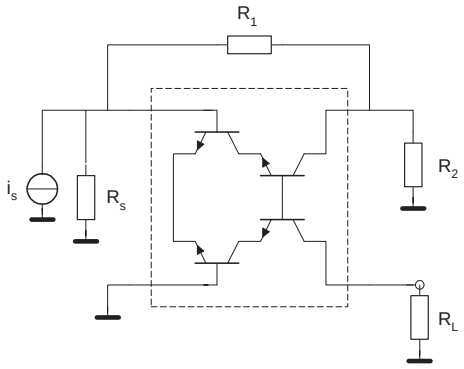

Figure 6.10: Amplifier 3

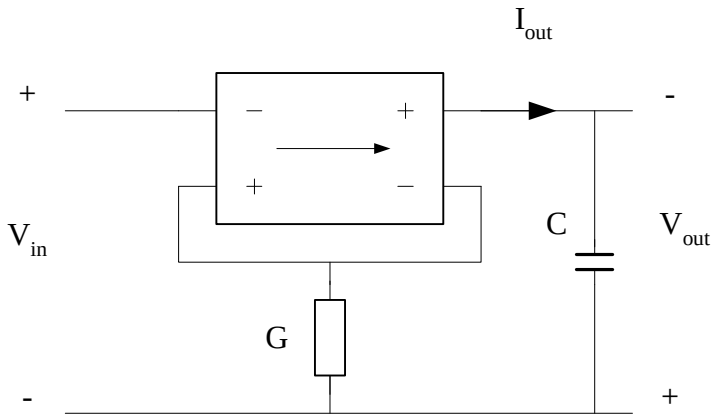

Figure 6.11: An active integrator

(a) Determine $A_{t\infty}$ for the active conductance.

Subsequently, assume that the nullor is approximated by a three stage implementation. The corresponding small-signal diagram is depicted in figure 6.12. For this situation the loop gain is studied.

(b) For calulating the loop gain, the loop needs to be broken somewhere. How and where can you break the loop for the amplfier of figure 6.12?

(c) What should the polarity be of a corret loop gain? Motivate!

(d) What is the dimension of a correct loop gain? Motivate!

(e) What is the effect of the integrator capacitor, C, on the loop gain?

(f) Determine the expression for the loop gain.

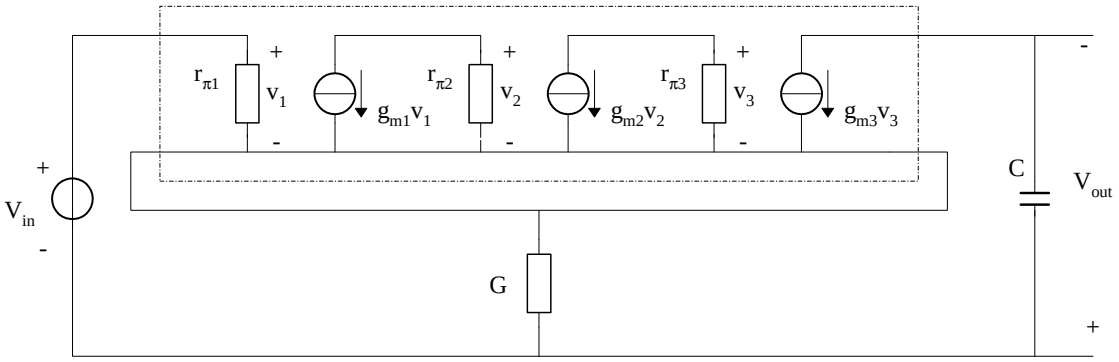

Figure 6.12: The small-signal diagram for the integrator when the nullor is implemented by three stages.

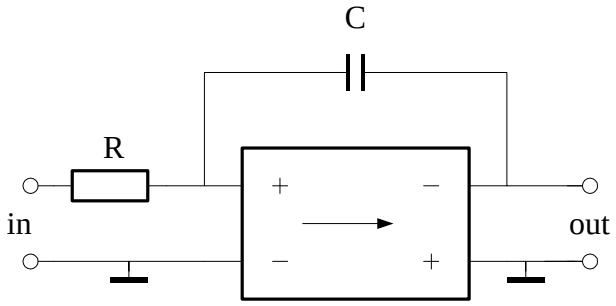

Figure 6.13: An active integrator.

5. Given an active integrator as depicted in figure 6.13.

(a) For what source and load type is this amplifier optimal? Motivate your choices.

(b) Determine $A_{t\infty}$ for this integrator.

The nullor is implemented by means of two amplifying stages. The corresponding small-signal diagram is depicted in figure 6.14.

(b) What is the DC loop gain of the integrator ?

(c) Calculate the loop gain as a function of the frequency.

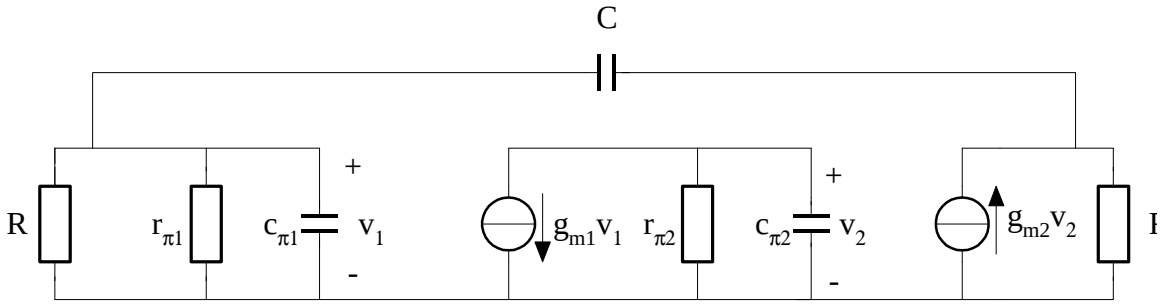

Figure 6.14: The small-signal diagram of the integrator when the nullor is approximated by two amplifying stages.

References

These lecture notes consist of a compilation of work from different authors. Besides newly written text, the following texts where used:

[1] G.L.E. Monna, "Analog IC-Techniques for Low-Voltages Low-Power Electronics", Parts of Chapter 6, editors W.A. Serdijn, C.J.M. Verhoeven and A.H.M. van Roermund, Delft University Press, 1995.

[2] C.J.M. Verhoeven, A. van Staveren, M.H.L. Kouwenhoven and G.L.E. Monna, "Structured Electronic Design, Negative-Feedback Amplifiers", Lecture Notes Structured Electronic Design, TU Delft

[3] A. van Staveren, "Structured Electronic Design of high-performance low-voltage low-power references, Kluwer Academic publishers, 2001

[4] C.J.M. Verhoeven, "Analog IC-Techniques for Low-Voltages Low-Power Electronics", Parts of Chapter 3, editors W.A. Serdijn, C.J.M. Verhoeven and A.H.M. van Roermund, Delft University Press, 1995.

[5] M.H.L. Kouwenhoven, A. van Staveren W.A. Serdijn and C.J.M. Verhoeven, part of Chapter 22 "Trade-offs in Low Noise Design", from Trade off in Analog Design, Editors C. Toumazou and G. Moschytz, Kluwer Academic Publishers, 2001