1 The need for data converters

Modern electronic systems process and store information digitally. However, due to the analog nature of the world, conversion from analog to digital and/or from digital to analog is, and always will be, inevitable and performed by data converters: Analog-to-digital converters (ADCs) and digital-to-analog converters (DACs).

Depending on a variety of factors, including technical specifications, system partitioning, and market needs, data converters can be either integrated on the same chip, or on the same package, together with other analog or digital blocks, or they can be stand-alone. Most stand-alone data converters are part of the $13–$16 billion standard analog market which also includes amplifiers, comparators, and interface and power management devices. In fact, data converters constitute 16% of this market, but they are growing at a faster rate than the other components, second only to power management devices [1, 2], with unit shipments going from about 2.9 billion units in 2010 to an estimated volume of about 4.7 billion units by 2015. That does not account for the embedded data converters [3] integrated together with digital signal processors (DSPs) in a wide variety of applications ranging from consumer electronics (e.g. audio devices, cell phones, imaging devices, DVD and multi-media players etc.) to automotive (e.g. embedded controllers), process control, and instrumentation.

This chapter will briefly outline some of the applications of data converters, illustrating the pervasiveness and variety of ADCs and DACs in modern predominantly digital signal processing systems. Moreover, as integrated digital systems continue to develop in line with CMOS technology scaling, and systems-on-a-chip (SoC) become ubiquitous, it is inevitable that data converters will need to cope with the needs of the applications and face the challenges and opportunities offered by ever shrinking process technologies. A second part of this chapter will be devoted to a summary of some of the issues, trade-offs, and solutions that data-converter designers face when designing analog circuits in the context of a technology road-map primarily driven by digital circuits needs. The chapter will also point to some of the alternatives to marching down the route indicated by Moore's law.

The concepts summarized in this chapter provide a context and motivation for the advanced data converters covered later in this book and will help us to develop a vision and spot the trends in this exciting field.

1.1 The digital revolution in an analog world: some examples of applications of data converters

As pointed out above, a digital signal processing (DSP)-based system needs ADCs and DACs in order for it to be able to exchange information or to control processes in the physical analog world. The number of applications of signal processing and control is extremely large and constantly expanding, and the aim of this section is to give the reader, by means of a few examples, a sense of the variety of the requirements and how this diversity leads to a wide range of different data conversion architectures, each one aiming to provide the optimal conversion solution to a different set of technical needs and specifications.

As stated before, it is assumed that the reader possesses a good understanding of the traditional A/D and D/A architectures and the fundamentals of data conversion and sampling theory.

1.1.1 Sensing applications

We will begin with precision measurements and sensor signal conditioning applications [4], including the digitization of signals originating from a wide variety of sensors[1] such as pressure, temperature, gas, speed, acceleration, and light-intensity sensors. An important part of this application space concerns temperature sensors used to monitor printed-circuit boards (PCBs) and microprocessor die temperature.

The applications we are referring to are generally characterized by relatively low-bandwidth signals ranging from DC to kHz at most. On the other hand, such signals can require high-precision digitization and/or may coexist with undesired signals due to, for example, electromagnetic interference, noise etc. Moreover, the electric variables transducing the sensed physical variables vary quite a bit. For instance, strain gages, flow meters, and pressure sensors are often variable resistors. Other sensors, such as some position sensors, behave as variable capacitors and so on. Depending on the interface circuitry between the actual ADC and the sensor, the input to the ADC may be a voltage or a current. Examples of interfaces between the sensor and the ADC include a resistor or a capacitor bridge, a force-sense system etc. There are many other important considerations, which go beyond the scope of this book. However, the common elements of many of these applications can include the following:

- very high nominal output resolutions up to even 24 bits with no missing code;
- very low noise level, for example up to 19 bits of noise-free code resolution;
- some level of on-chip digital filtering; for instance allowing excellent 50 Hz/60 Hz power supply rejection;
- low offset;
- low gain error;
- low temperature dependence;
- low pin count package.

[1] Possibly after signal conditioning including analog filtering, amplification, level shifting etc.

Table 1.1. Typical specifications for digital audio converters [7]

Application	THD+N (dB)	f_s (kS/s)
Telecom	60–70	8
FM stereo	60–70	32
Speech analysis	70–80	8–48
Computer audio	80–90	48
Stereo CD, DAT, etc.	>100	44.1, 48, 88.2, 96
DVD audio	>100	48, 96, 192

The converters that have traditionally been used in applications like these are the so-called "serial" ADCs such as *single-slope* and *dual-slope* ADCs [5]. The principle of operation of such ADCs consists of comparing an internally generated accurate voltage ramp with the input signal (assuming that this does not appreciably vary while the conversion is ongoing) and counting how many clock cycles elapse between the start of the ramp and the instant at which the ramp reaches the input signal. These integrator-type converters must rely on very linear ramp generators, stable clock sources etc., and, as will become clear from the second part of this chapter, have fundamental issues with the low-voltage supplies of advanced CMOS processes.

For these and other reasons, in recent years, high-resolution $\Delta\Sigma$ ADCs and incremental ADCs[2] have gained considerable popularity in these applications. Moreover, these types of converters are easily implemented in CMOS, and greater flexibility is obtained by adding many on-chip digital programmable functions including filtering, variable throughput rate, programmable gain, calibration modes etc. while requiring few or no external components. In addition to that, since these are oversampled converters, the requirements on the anti-aliasing filters are considerably relaxed due to the large difference between the signal bandwidth and the Nyquist frequency, making the entire front end simpler, smaller, and cheaper.

1.1.2 Digital audio

This market includes audio circuits in portable applications such as wireless handsets or CD and MP3 players as well as in fixed applications such as high-fidelity (HiFi) systems and professional/studio audio systems [7].

In the case of speech digitization (e.g. in cell phones) the standard sampling frequency is $f_s = 8\,\text{kS/s}$, and 16-bit linear $\Delta\Sigma$ ADCs and DACs have replaced the traditional logarithmic converters used in the early systems. In fact, DSP-based speech compression algorithms reduce the overall data rate to acceptable levels. More than resolution, total harmonic distortion plus noise (THD+N) is an important parameter and Table 1.1 shows typical values of the THD+N and sample frequency f_s for various digital audio applications [7]. $\Delta\Sigma$ ADCs and DACs also dominate in the professional audio area, and,

[2] An incremental ADC [6] is a particular type of $\Delta\Sigma$ ADC that is re-set after the conversion of each sample. Better coverage of this type of ADC is deferred to subsequent chapters.

once again, offer the advantage of simplifying the anti-aliasing and reconstruction filters owing to oversampling. Furthermore, oversampled DACs have the additional advantage of suffering less from the $\sin(x)/x$ spectral shaping issue[3] thanks to both oversampling and the use of digital interpolators/filters preceding the DAC itself.

There are also other important specifications that are very specific to audio systems, such as containing the power of specific harmonics, but that is well beyond the scope of this chapter.

1.1.3 Wireless communication infrastructure

Mobile telephony, despite going through market cycles, has overall seen a significant steady increase in demand for mixed-signal products, including data converters. Basic cell phones have been adopted widely in Western countries as well as in Japan and Korea, and they are quickly penetrating populous countries and emerging markets such as China and India as well as many South American countries [8], hence driving increasing production. Furthermore, compounding this increasing pool of users, the emergence of smart phones with their associated higher demand for bandwidth and services means that the wireless communication infrastructure is going to require a significant expansion that is somewhat similar to what has been experienced by the wireline infrastructure as a result of the boom in wired internet access [9].

In wireless communication infrastructure applications, the data converters are critical blocks of base transceiver stations (BTSs), namely the wideband transceivers that handle the communication with the many handsets active in the corresponding phone cell [10, 11, 12]. Such converters, in order to acquire or synthesize multiple user channels, at present have signal bandwidths of the order of a few tens of MHz and are steadily headed toward the 100 MHz mark. BTS systems can be required to be able to handle multiple communication standards such as GSM, CDMA, LTE, wideband CDMA, and WiMax (or IEEE 802.16) [13]. This means that higher flexibility, integration (hence also reduced cost and power consumption per digital function), and lower overall costs should be achieved by placing the ADCs and DACs as close as possible to the antenna [14], hence the shifting of some of the less flexible, expensive, high-count parts, which are traditionally analog functions, to the digital domain. Some of these analog functions include filtering, using surface acoustic wave (SAW) filters, ceramic filters, or crystal filters, with tight tolerance and matching requirements for frequency-sensitive components such as inductors and capacitors, mixing down/up-converters, oscillators etc. Because of that there is an increasing demand, where sensible, to move from zero-intermediate-frequency (IF)/low-IF schemes whereby the ADCs or the DACs convert baseband signals to high-IF digitizing/synthesizing schemes whereby the IF-to-baseband frequency down-conversion/up-conversion stage (which may be a quadrature scheme to take care of image rejection), possibly with some filtering, is moved from the analog to the digital domain. Ideally, the final goal would be to put the ADC right after the

[3] Because the desired output signal resides at a frequency that is much lower than the first frequency null of $\sin(x)/x$ due to the much higher f_s.

low-noise amplifier (LNA), aside, perhaps, for some filtering, and in front of a receive signal processor (RSP), and the DAC right before the power amplifier (PA), aside, again, for some filtering, and following a transmit signal processor (TSP), fully realizing what is commonly referred to as software radio (SR) [10, 15, 16, 17].

Some of the issues stemming from bringing the converters toward higher-frequency stages of the transceiver are related to its sensitivity, selectivity, out-of-band (image) rejection, ability to handle undesired powerful "blockers" etc. Moreover, converting such high-frequency signals, with even increased dynamic specifications as the analog–digital interface is moved closer to the antenna,[4] is a formidable challenge and quickly results in prohibitive power consumption levels.

The radio specifications consist of the noise level, measured in terms of its noise spectral density (NSD), and the linearity with which signals are processed, measured typically by means of two-tone or multi-tone intermodulation distortion. Such specifications, once the bandwidth has been specified, result in converter specification. For ADCs in IF digitizing applications these include SNDRs in the range between the low 70s (dB) and the mid-to-high 80s (dB). Often these specifications can be traded, at least to some extent, with the PGA gain placed in front of the ADC and the selectivity of the preceding filters. These converters are often "SNR-limited" (namely, the RMS noise is comparable to or greater than the LSB size) and have very low distortion. The latter is often quantified by the spurious-free dynamic range (SFDR) for wideband cases and by the intermodulation distortion (IMD) for narrowband cases.

The sampling clock phase noise (PN) is also a critical specification due to its negative effects on the digitized noise floor (particularly at frequencies close to that of the digitized signal) as a result of sampling aperture uncertainty, and it becomes increasingly worse as a function of the increasing digitized signal frequency.

For converters used in handsets, in addition to the particular engineering performance specifications enabling the communication standard, some of the key specifications result from the need for products (1) to be small in size, (2) to be inexpensive, and (3) to have a long stand-by time (long battery life). This is achieved through employing higher integration levels (as pointed out before, digitizing earlier in the processing chain reduces the number of passive components and the overall part count) while, at the same time, striving to improve power efficiency [18].

On the other hand, in BTS applications, converters capable of meeting the associated engineering performance specifications can be very power hungry, demanding up to 1 W per converter core or more. However, BTSs are large and power-hungry fixed systems, and, although about 30% of the operation cost is electric power consumption, the contribution of the converters to the overall power consumed by the transceiver is not very significant (as opposed to, for example, the PA or some of the driving amplifiers that make up for the loss in filters and/or drive the inputs of ADCs). Therefore obtaining the desired noise and distortion performance is typically the most important goal, and higher power consumption is an acceptable cost to pay for it. Also, their size

[4] Which means that the signal now includes a lot of undesired content that would otherwise have been filtered out and that now interferes with the desired content.

is not necessarily a very critical factor (and, although it impacts production costs, the average selling prices are typically large enough as a result of the level of performance delivered by these parts) since they take up a relatively small amount of space compared with some of the filters and the system as a whole, and, again, these are not mobile systems.

Somewhat conceptually similar to BTS applications are wireless LAN (sometimes referred to as *WiFi* or, more generally, as the IEEE 802.11 standard) applications, where the *access point* (which is homologous to the BTS in this application) is required to handle multiple "clients" (e.g. laptops, workstations etc.) in a relatively small environment such as an office space, a public library, a train or an airplane, a coffee shop etc.

Converters for such applications are somewhat similar to those just cited in the BTS case, but with considerably more relaxed dynamic specifications. It should be noted that there is at present a trend toward convergence between WiFi, its larger-range sibling WiMAX, and cell phone BTSs (in fact, in the case of so-called pico-cells, a cell is extremely constrained in physical space). This is, again, something that could be made to become more and more real by taking advantage of the flexibility offered by SRs. The catch, of course, is in the details of its implementation.

A slightly different category of wireless communication applications is constituted by satellite communication and military communication applications. In the former case, while the dynamic specifications tend to be more relaxed than in the BTS case, the bandwidth of the signal is considerably wider (converters with $f_s > 1$ GS/s are not unusual). A peculiar requirement in space applications, however, is the ability of the converter to be radiation-tolerant since the chip is meant to be part of a satellite system in space and therefore is going to be exposed to radiation and high-energy particles. Such exposure, if proper design measures/techniques are not employed, can rapidly lead to latch-up (transient or permanently destructive) internal to the chip being induced by radiation and bombardment of ions and particles. Packaging is another challenge, and special packages are often required.

In military applications, similarly to the space applications, a common requirement is a very high sampling frequency and, again, relatively relaxed SNDR.[5] However, here, the environmental conditions (e.g. the temperature range) under which one needs to guarantee the converter's performance are extreme.

1.1.4 Health care and life sciences

Electronics for health care is another application area that is projected to grow considerably in the near future. This is a very diverse application field insofar as it includes new-generation implementations of "traditional" equipment such as ultrasound scanners, CAT scanners, X-ray machines, blood pressure monitors etc. as well as newer types of devices such as diagnostic devices taking advantage of ultrawideband wireless

[5] In reality it is not that it is relaxed but rather that it is not yet feasible to have single ADCs with more than 12–14 ENOB *and* a sample rate in excess of 3–5 GSPS! So the system designer will need to make do with what is available.

technology for breast cancer screening or pocket-sized CPR devices using micro-electro-mechanical accelerometers and able to perform real-time measurements of the rate and depth of CPR chest compressions [19]. The variety of system implementations (and the corresponding bandwidth and dynamic range of the signals being processed) is huge.

For example, ultrasound machines take advantage of phased array signal processing to be able to estimate spatial information together with density information. Correspondingly, large arrays (16, 32, 64, and more) of high-voltage DACs are often required as part of the synthesizer for the ultrasound pulse launched toward the organs being scanned. Likewise, large arrays of ADCs are critical components of the receive path digitizing the returning ultrasound pulse. At present, for example, medium-resolution ultrasound machines use arrays of ADCs with SNR in the low 70s (dB) and input signal bandwidth of the order of tens of MHz [20]. In such applications, due to the large number of DACs/ADCs required, together with the dynamic specifications (SNR, THD, etc.), it is also important that the individual converters have low power consumption and low area in order to allow the highest possible level of integration and, increasingly, low weight and long battery life in order to also enable portable devices. Both pipelined and $\Delta\Sigma$ ADCs are used in this type of application.

A completely different case is, for example, the one where miniaturized arrays for gene-based tests are used in DNA testing and the signals have sub-kHz bandwidth and a complex chip, where the A/D (basically an incremental A/D) is only one of the blocks, using only microwatts of power [21]. Another case is that of a chip to monitor vital signs such as temperature, heart rate, and electrocardiogram (ECG/EKG), where the integrated 10 bit/500 Hz $\Delta\Sigma$ ADC works at 1 V supply and requires 20 μW [22].

1.2 Challenges and opportunities offered by recent technology advancements

Progress in the development of silicon technology continues to allow deeper levels of CMOS transistor scaling and integration [23, 24]. Moore's law [25] has continued to predict the rate of integration fairly well for decades and, actually, the rate has even increased in recent years [23, 26]. That has both been driven by system needs and has made new applications possible in a somewhat symbiotic fashion. Despite the various types of physical and technological challenges, CMOS scaling is going to continue advancing for many years to come thanks to innovation in the area of materials and device design. In addition to higher levels of integration and reduced cost of functionality, scaling benefits digital circuits in terms of power efficiency and speed.

Unfortunately, however, analog circuits such as ADCs and DACs do not necessarily benefit from CMOS scaling in the same way as that in which digital circuits do [27, 28, 29, 30, 31], as will be briefly summarized in Section 1.2.1. However, as discussed above, to continue to exploit the system advantages offered by digital functionality, it is necessary for data converters in many application spaces to deal with the challenges of ever shrinking CMOS processes.

Table 1.2. Examples of voltages associated with CMOS technology nodes

Technology node (nm)	Analog V_{DD}	Digital V_{DD}	V_T
250	3.3	2.5	0.60
180	3.3	1.8	0.45
130	3.0	1.2	0.40
90	2.5	1.0	0.30
65	1.8	1.0	0.30
45	1.2	0.8	0.25

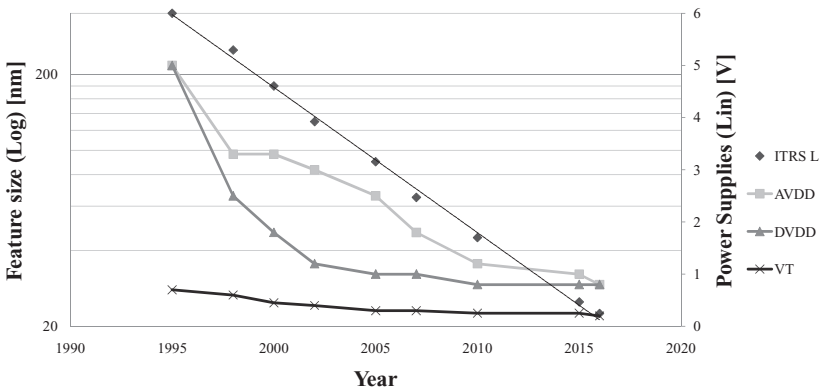

Figure 1.1 CMOS process scaling: feature length, associated supplies, and threshold voltages.

Yet, in some other cases, alternatives to full integration in deep sub-micron CMOS processes do exist. This includes using BiCMOS processes, multi-chip modules, or thoughtfully defined chip sets as discussed in Section 1.2.2.

1.2.1 Main issues with scaling in analog circuits

A first and fundamental issue is the reduction in power supply V_{DD} as the thickness of the thin oxide of the MOS gates T_{ox} is decreased.[6] Table 1.2 and Fig. 1.1 show examples of supplies and MOS threshold voltages for various CMOS process nodes. The available signal swing is reduced accordingly because of the reduction in voltage headroom in critical circuits such as operational amplifiers, comparators etc. Unfortunately, the noise floor does not scale down while the threshold voltages do not scale at the same rate as that of the supplies.[7] Owing to the lower supply, the common-mode range of differential circuits is also reduced, aggravating the swing limitations further.

[6] Assuming a SiO_2 dielectric, the oxide thickness T_{ox} scales proportionally to the minimum feature size L_{min} as $T_{ox} = \lambda L_{min}$, with $\lambda \sim 0.03$ [29].

[7] Lowering V_T leads to excess subthreshold current $I_{OFF} \sim \exp[-V_T q_e/(nk_BT)]$ when the channel is not formed. However, some processes provide different choices of devices with different threshold voltage levels, including, sometimes, zero-threshold devices.

Although, as shown in Table 1.2, mixed-signal CMOS processes offer "analog" MOS transistors with a higher nominal supply than that of the "native" digital MOS devices, the former have higher gate oxide thicknesses (hence, these are sometimes referred to as "dual gate oxide," DGO, devices) and longer minimum channel lengths than the latter. In other words, for a given process node, the designer can take full advantage of the scaled digital circuits while being able to design analog circuits on the same chip with devices from roughly one generation (or more) behind. A considerably more useful process option is the one of using "asymmetric" thin-oxide *high-performance analog* (HPA) transistors, with which a specially designed drain implant area leads to higher output impedance at relatively short channel length with clear benefits for op-amp design [32].

As the voltage headroom is reduced, the use of (especially multiple) cascoding to implement high DC gain in operational amplifiers becomes increasingly difficult. To cope with that, designers replaced multiple levels of cascoding with regulated/boosted cascoding, or ultimately resorted to using multi-stage (e.g. Miller or nested-Miller and then feed-foward compensated amplifiers) amplifiers [33, 34, 35, 36, 37, 38, 39, 40]. However, some of these architectural choices can either lead to the aggravation of introducing multiple poles at lower frequency than in a traditional single-stage cascoded amplifier, therefore ending up with actual lower closed-loop bandwidth designs, or are difficult to use in traditional switch capacitor circuits due to their transient response.

CMOS switches (sometimes referred to as "transmission gates") also suffer greatly from the supply reduction (see Fig. 1.2(c)). The on-resistances (R_{onN} and R_{onP}) of the parallel-connected NMOS and PMOS transistors constituting the transmission gate strongly depend on their respective transistor's overdrive voltage (e.g. $V_{OD} = V_{GS} - V_{TN}$ for the NMOS transistor) (see Fig. 1.2(d)). If the gates of these two transistors are driven to the supply levels (e.g. $V_G = V_{DD}$ to turn on the NMOS transistor) while the passing input signal V_{in}, tied to the source terminal ($V_S = V_{in}$), is at a voltage level that is becoming too close to V_T volts away from the gate voltage level, then the overdrive voltage (e.g. $V_{OD} = V_{DD} - V_{in} - V_{TN}$ to turn on the NMOS switch; see Fig. 1.2(a)) may become too close to zero or even negative. As a result, the corresponding MOS switch will not be able to have a sufficiently small on-resistance or it could be completely switched off while it was desired to be turned on (see Fig. 1.2(b)). So, while a passing signal close to either V_{DD} (leading to low R_{onP} in the PMOS switch) or ground (leading to low R_{onN} in the NMOS switch) will cause at least one of the two parallel transistors to be fully turned on with a large overdrive, on the other hand, a passing signal at a level around $V_{DD}/2$ may lead to insufficient overdrive voltage on, possibly, both NMOS and PMOS transistors, which could even turn off [33, 34] (see Fig. 1.2(e)). The latter will happen, for example, as the absolute sum of the two threshold voltages $||V_{TN}| + |V_{TP}||$ becomes equal to or larger than V_{DD}.

In order to address the issue of insufficient overdrive voltage in MOS switches, techniques involving driving the gate voltage outside the range set by the supply voltages (i.e. to a potential higher than V_{DD} for the NMOS switch and to a potential lower than ground for the PMOS switch, respectively) have been developed. Generally, two types of techniques are used, known as "boosted supplies" and "boosted switches," respectively. With "boosted supplies" an internal voltage level higher than V_{DD} (or lower

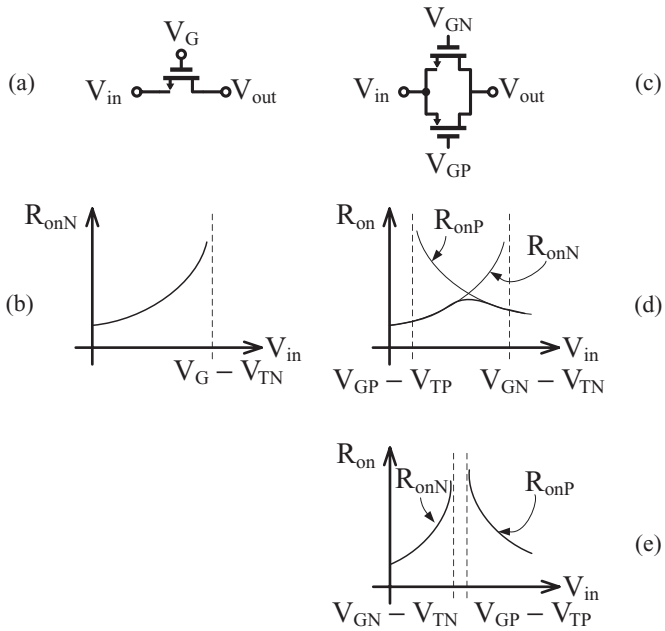

Figure 1.2 MOS switches. (a) An NMOS switch. (b) The on-resistance of the NMOS switch R_{onN} as a function of the input voltage V_{in}; the resistance becomes increasingly higher as the input approaches $V_G - V_{TN}$. (c) A CMOS switch. (d) The on-resistance of the CMOS switch R_{on} as a function of the input voltage V_{in}; since $R_{on} = R_{onN} // R_{onP}$, for sufficiently high overdrive voltages on the NMOS and PMOS transistor, R_{on} is low and relatively constant. (e) The same R_{on}, however, for very low voltage and, hence, insufficient overdrive on the NMOS and PMOS, can become very high or nearly infinite when V_{in} nears the middle of its range; the switch can conduct only when V_{in} is near the supply and ground as the NMOS and PMOS switches separately get sufficient overdrive to turn on.

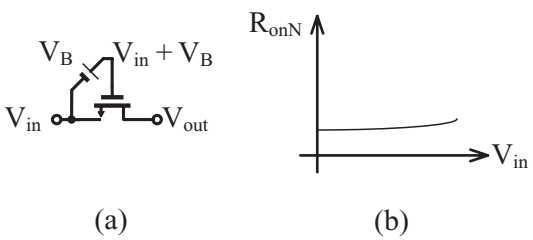

Figure 1.3 (a) An NMOS boosted switch. (b) R_{on} of the boosted switch versus V_{in}; the slight increase of R_{on} for higher V_{in} is due to the "bulk effect," namely the increase of the threshold voltage for increasing V_G.

than ground) is generated (e.g. using charge pumps) to be used to drive the gates of the MOS switches. With "boosted switches" [41], instead, a "floating battery" V_B is generated and then connected across the gate–source of the MOS switch, to turn it on, so that its gate voltage level "rides" V_B volts away from the passing signal as shown in Fig. 1.3. Conceptually, for example, this could be realized by first connecting a capacitor