Chapter 2
Digital Transmitter Architectures: Overview

This chapter gives an overview of different fully or partially digital transmitter architectures that have been presented in literature. These are classified according to different properties, and the main advantages and disadvantages for each type are investigated. This will motivate the use of continuous-time PWM-based polar transmitter architectures, which are the subject of the remainder of this work.

Section 2.1 explains several ways in which analog or digital information can be modulated onto an RF carrier which can be transmitted through the air. A general expression is introduced which covers all the different types of modulation. This expression defines the requirements for a multistandard transmitter.

Any transmitter contains some type of power amplifier (PA). These form a separate field of research, which is not the subject of this work. However, some insight in the operation of PAs is required since the RF modulator inevitably has a large influence on the efficiency of the PA, which largely determines the overall power consumption of the transmitter. For this reason, a short summary of the targeted PA architectures is given in Sect. 2.2. The discussion is limited to switched-mode PAs (SMPAs) since these have the highest efficiency and can amplify an input that comes directly from a digital circuit without requiring digital-to-analog conversion.

Section 2.3 compares different types of modulators, which all transform a baseband signal to a modulated RF carrier. In order to produce a signal suitable to drive an SMPA, all information needs to be encoded into a two-level signal, which can be done in many different ways. The most important ones are discussed and evaluated in Sect. 2.4. Finally, Sect. 2.5 concludes this chapter.

2.1 Modulation

Modulation is the process of converting information one wants to transmit into one or more varying properties of a sinusoidal wave called the carrier.

This information can be an analog signal, such as a varying voltage or current, or a digital signal, which consists of a sequence of bits, i.e. 0 and 1 values. An analog
signal can result from an audio signal, a temperature or pressure measurement, etc., that is converted to a voltage or current using some kind of sensor. A digital signal can be a digitized version of such an analog signal, or it can be inherently digital information such as a file stored on a computer. Many digital signals consist of multiple parallel bit streams (e.g. 8 parallel bit streams can be used to represent a series of numbers from 0 to $2^8 - 1 = 255$). However, for the purpose of modulating a carrier, this is not important: all bits are serialized somehow into a single bit stream.

Modulation schemes for both analog and digital signals will be discussed in this section. It will be shown that regardless of the modulation scheme and the analog or digital nature of the signal, the modulated carrier can be expressed as a sinusoid with a varying phase and amplitude. The rest of this work will then assume such a carrier without taking the exact nature of the modulating signal into account.

2.1.1 Traditional Analog Modulation Schemes

A general, unmodulated sinusoidal wave can be represented as

$$s(t) = a \cdot \cos(2\pi f_c t + \phi),$$

(2.1)

where the variable $t$ represents time. This expression has three parameters: the amplitude $a$, the frequency $f_c$ (the index $c$ is commonly used to indicate that $f_c$ is the carrier frequency) and the phase $\phi$. Each of these parameters can be varied to modulate information onto the wave, which then becomes a carrier.

2.1.1.1 Amplitude Modulation

Amplitude modulation (AM) consists of varying the amplitude of a carrier to modulate information onto it. Thus, the amplitude now becomes a function of time, called the AM signal $a(t)$.

Thus, the modulated carrier is given by

$$v(t) = a(t) \cos(2\pi f_c t + \phi),$$

(2.2)

where $\phi$ is an arbitrary constant and is usually normalized to 0.

2.1.1.2 Phase Modulation

Phase modulation (PM) consists of varying the phase of a carrier to modulate information onto it. Now, the phase becomes a function of time, called the PM signal $\phi(t)$.

The modulated carrier is now given by
\[ v(t) = a \cdot \cos(2\pi f_c t + \varphi(t)), \quad (2.3) \]

where \( a \) determines the amplitude, and thus the power, of the transmitted electromagnetic wave.

### 2.1.1.3 Frequency Modulation

In *frequency modulation* (FM), the *frequency* of the sinusoid is slightly varied in order to modulate information onto it. Since a communication channel is usually a small frequency band centered around a nonzero carrier frequency \( f_c \), the frequency of the sinusoid can be written as \( f_c + \Delta f(t) \), where \( \Delta f(t) \) is the FM signal, and \(|\Delta f(t)| \ll f_c \ \forall t\). Thus, the modulated carrier becomes

\[ v(t) = a \cdot \cos(2\pi (f_c + \Delta f(t)) t + \varphi(t)) \]

\[ = a \cdot \cos(2\pi f_c t + \varphi'(t)), \quad (2.4) \]

where

\[ \varphi'(t) = \varphi + 2\pi \Delta f(t)t \]

\[ (2.5) \]

can be seen as a PM signal. This shows that FM is just a special case of PM, and therefore it does not need to be considered separately.

### 2.1.2 General Modulated Signal and Complex Representation

AM can be combined with PM or FM in order to transmit more information at the same time. Since FM is a special case of PM, this means a general modulated carrier can be written as [8]

\[ v(t) = a(t) \cos(2\pi f_c t + \varphi(t)). \quad (2.6) \]

This is called the *polar representation* of the modulated carrier. Alternatively, \( v(t) \) can be expressed as

\[ v(t) = i(t) \cos(2\pi f_c t) - q(t) \sin(2\pi f_c t), \quad (2.7) \]

where

\[ i(t) = a(t) \cos \varphi(t) \quad (2.8) \]

is the *in-phase signal* or *I signal* and

\[ q(t) = a(t) \sin \varphi(t) \quad (2.9) \]
is the quadrature signal or \( Q \) signal. Equation (2.7) is called the quadrature representation of the modulated carrier.

Finally, \( v(t) \) can also be written as [8, Sect. 4–1] [24, Sect. 2.12]

\[
v(t) = \text{Re} \left\{ g(t) \cdot e^{j2\pi f_ct} \right\},
\]

(2.10)

where \( \text{Re} \{x\} \) represents the real part of a complex number \( x \) and

\[
g(t) = i(t) + j\varphi(t)
\]

(2.11)

\[
g(t) = a(t) \cdot e^{j\varphi(t)}.
\]

(2.12)

Equation (2.10) is called the complex representation of the modulated carrier, and \( g(t) \) is called the complex envelope. It is a complex signal that does not physically exist anywhere but is a convenient mathematical representation.

The Fourier spectrum (see Sect. A.1) of \( v(t) \) is given by [8, Sect. 4–3]

\[
V(f) = \frac{1}{2} \left[ G(f - f_c) + G^*(-f - f_c) \right],
\]

(2.13)

where \( G(f) \) is the spectrum of \( g(t) \) and \( G^*(f) \) is the complex conjugate of \( G(f) \).

### 2.1.3 Single-Carrier Digital Modulation Schemes

Digital modulation schemes are all based on AM, PM, FM or combinations of these modulation types. The most straightforward modulation schemes modify one of the characteristics of the carrier (frequency, amplitude or phase) depending on the bit that is transmitted. The signal switches between two predefined values of this characteristic, where one value corresponds to a 0 and the other one to a 1. This value is kept constant during one symbol period. Since one bit is transmitted per symbol period, the bit rate is equal to the symbol rate.

The case where frequency modulation is used is called frequency shift keying (FSK) and is shown in Fig. 2.1a. If amplitude modulation is used and the amplitude is switched between 1 and 0 as in Fig. 2.1b, the modulation scheme is called on-off keying (OOK). The case where phase modulation is used is called binary phase shift keying (BPSK) and is shown in Fig. 2.1c. In this case, the phase shifts between 0 and \( \pi \) depending on the transmitted bit.

OOK and BPSK signals can be represented as a complex envelope that switches between two different complex numbers: For OOK, these are 0 and 1, while in the case of BPSK they are \(-1\) and 1. These values can be plotted in the complex plane, which leads to a constellation plot, as shown in Fig. 2.2a for OOK and Fig. 2.2b for BPSK. Thus, the complex envelope is constant over each symbol period.
While an FSK signal is an FM signal and thus can be represented by a varying complex envelope, the envelope will change continuously rather than switching between a discrete number of complex values, so that no constellation plot can be created for FSK.

If the complex envelope switches between more than 2 different values, multiple bits can be transmitted per symbol. An example of this is *amplitude shift keying* (ASK), where the amplitudes are equally spaced between $-1$ and $1$ (note that since...
there are also negative amplitude values, this can be considered a combination of AM and PM). Figure 2.2c shows 4-ASK; 2-ASK is equivalent to BPSK.

Another, more common example is quadrature phase shift keying (QPSK), shown in Figs. 2.1d and 2.2d, which uses only phase modulation. Like 4-ASK, it encodes 2 bits per symbol. This can be extended to more bits, which leads to e.g. 8-PSK, shown in Fig. 2.2e.

Even higher data rates can be achieved by combining AM and PM, which is called quadrature amplitude modulation (QAM). Usually, the number of bits per symbol is even so that the constellation points fit in a square, such as in 16-QAM (Figs. 2.1e and 2.2f) or 64-QAM (Figs. 2.1f and 2.2h), but 32-QAM (Fig. 2.2g) is also used sometimes. 4-QAM is equivalent to QPSK and does not have any amplitude modulation.

The minimal distance between two constellation points (relative to the maximal distance from any constellation point to the origin) determines the robustness against noise: Noise causes the signal to deviate from the ideal constellation points, so that the obtained constellation plots consist of clouds centered around the ideal points. The normalized average deviation on the constellation points is called error vector magnitude (EVM) and is defined in Sect. A.8. If a symbol deviates too far from its ideal constellation point, it can be misinterpreted as representing a different constellation point, in which case bit errors are made. Thus, the choice of the constellation implies a trade-off between data rate and robustness: If there are more constellation points, they will be closer to each other so that the margin for noise becomes smaller.

All the above modulation schemes are referred to as single-carrier schemes. This name will become clear when comparing them to OFDM signals, which are discussed in Sect. 2.1.4.
Many more single-carrier modulation schemes exist, all of which can be represented using (2.10) and thus using (2.6) or (2.7). They all have their advantages and disadvantages [8, Sects. 5–9 to 5–11] [24, Chap. 8]. Often, schemes without AM are preferred as they do not require a linear PA. However, much higher data rates can be achieved if AM is also used, so that QAM is very often used in modern communication standards.

In this work, single-carrier signals are used for the RF PWM measurements presented in Chap. 6. These are carried out using QPSK, 16-QAM, and 64-QAM signals.

Figure 2.1 shows that digital modulation schemes cause very abrupt transitions in the signals, which lead to very high bandwidths. For this reason, bandpass filtering is usually applied, which smooths the transitions.

### 2.1.4 OFDM

Orthogonal frequency-division multiplexing (OFDM) [8, Sect. 5–12] is a digital modulation scheme that uses multiple equally spaced subcarriers within the signal band. Each carrier is modulated using one of the single-carrier modulation schemes, such as 16-QAM or 64-QAM. This way, the number of bits per symbol is multiplied by the number of subcarriers. For example, the WLAN standard [71] uses 64 carriers of which 48 carry information. At the maximal data rate, each of these carriers is modulated using 64-QAM and thus carries 6 bits per symbol (since $2^6 = 64$). Thus, the maximal total number of bits per OFDM symbol is $48 \times 6 = 288$.

Consider an OFDM signal with $2N + 1$ subcarriers which are located at equally spaced frequencies $f_i$ where $i$ is an integer with $-N \leq i \leq N$. The frequency $f_i$ can be written as

$$f_i = f_c + i \Delta f,$$

where $f_c = (f_{-N} + f_N)/2$ is the carrier or center frequency and $\Delta f$ is the spacing between the frequencies.

An OFDM symbol with length $T$ centered at $t = 0$ can then be written as

$$s_0(t) = \sum_{i=-N}^{N} a_i \cdot \cos(2\pi f_i t + \varphi_i) \cdot \Pi \left( \frac{t}{T} \right),$$

where the rectangular function $\Pi(t)$ is defined in Sect. A.3.2. This rectangular window ensures each symbol is zero outside its symbol period. During the symbol period, the amplitude and phase of each subcarrier are constant. It follows that the spectrum of $s_0(t)$ is
Fig. 2.3 Illustration of the orthogonality of the subcarriers in OFDM

\[ S_0(f) = \sum_{i=-N}^{N} \frac{a_i}{2} \left[ \left( e^{j\varphi_i} \delta(f - f_i) + e^{-j\varphi_i} \delta(f + f_i) \right) \right] * [T \text{sinc}(fT)] \]

\[ = \sum_{i=-N}^{N} \frac{a_i T}{2} \left[ e^{j\varphi_i} \text{sinc}((f - f_c - i\Delta f)T) + e^{-j\varphi_i} \text{sinc}((f + f_c + i\Delta f)T) \right]. \]

(2.16)

where \(*\) denotes convolution (see Sect. A.2) and the sinc function is defined in Sect. A.3.1. Thus, the spectrum consists of sinc pulses centered at frequencies \( \pm f_i \), which means their centers are spaced by the subcarrier spacing \( \Delta f \). Since the sinc pulses have infinite bandwidths, they overlap. However, if \( \Delta f = 1/T \), then the sinc functions are centered at each other’s zeros, so that there is exactly one subcarrier which has a nonzero contribution at each frequency \( f_i \). This is illustrated in Fig. 2.3 for \( N = 4 \) for subcarriers which all have the same amplitude. Thus the amplitude \( a_i \) and phase \( \varphi_i \) of each subcarrier can still be measured exactly. This means that even though the different modulated signals have infinite bandwidth and are located close to each other, they theoretically do not interfere with each other.

In [8], it is shown that this property can also be expressed in the time domain by noting that all subcarriers are orthogonal over the symbol period \( T \) provided that \( \Delta f = 1/T \). This explains the word orthogonal in the name OFDM.

Since one symbol represents many bits, the symbol rate can be lowered which makes the signal more robust to multipath effects (different versions of a signal arriving at different times in the receiver, which results in intersymbol interference) [8]. This can also be seen in the frequency domain: Multipath effects result in frequency-selective fading, which means that different subcarriers suffer from different phase and amplitude distortion, and some subcarriers may disappear completely.
2.1 Modulation [47, Chap. 14]. If some redundancy is added in the encoded bits, distortion that affects only a limited number of carriers can often be tolerated as most bits are still received correctly. In the case of a single-carrier standard with the same bandwidth, fading typically affects all transmitted bits. For this reason, OFDM is used in several modern communication standards, such as WLAN.

However, an important disadvantage of OFDM is its large peak-to-average power ratio (PAPR): The maximal instantaneous signal power occurs when the peaks of all subcarriers occur at the same time while all subcarriers are in the constellation points with the highest amplitude. This case is very unlikely, and most of the time some of the carriers are in counterphase and partially cancel out, while some of the remaining carriers have a lower amplitude. This causes the PAPR to be 10 dB or higher in the case of WLAN [12].

WLAN-based test signals are used in Chaps. 3, 5, and 6 to evaluate the performance of transmitters for high-bandwidth high-PAPR signals.

Even OFDM signals can be represented using the general representation given by (2.10). This can be seen as follows. Since the $i$th carrier is modulated using a single-carrier modulation scheme, it can be represented using the complex representation (2.10) as

$$v_i(t) = \Re\left\{g_i(t) \cdot e^{j2\pi ft}\right\}$$

$$= \Re\left\{g_i(t) \cdot e^{j2\pi ft} \cdot e^{j2\pi \Delta ft}\right\},$$

where $g_i(t)$ is the complex envelope of the $i$th carrier. The complete OFDM signal is then given by

$$v(t) = \Re\left\{e^{j2\pi ft} \sum_{i=-N}^{N} g_i(t) \cdot e^{j2\pi \Delta ft}\right\}$$

$$= \Re\left\{g(t) \cdot e^{j2\pi ft}\right\},$$

where

$$g(t) = \sum_{i=-N}^{N} g_i(t) \cdot e^{j2\pi \Delta ft}.$$  

Equation (2.20) is identical to (2.10). Thus, OFDM signals can also be represented as a single modulated carrier, so that (2.6) and (2.7) are also valid for OFDM.
2.1.5 Conclusion

All modulation schemes considered above can be represented using (2.6) and (2.7). Thus, an RF modulator is a modulator that implements one of these equations. This conclusion will be used in Sect. 2.3 where different types of modulators are discussed. However, first the PA will be discussed, since it has a large impact on the modulator types that can be used.

2.2 Power Amplifier

Even though the PA is not part of the digital modulator, it is essential to take its properties into account when designing the modulator. This is because the way the digital signal is encoded has a large influence on the PA’s efficiency.

For this reason, this section presents a short overview of the types of PAs that are targeted and the aspects that are relevant in the context of this work.

2.2.1 Switched-Mode Power Amplifiers

Switched-mode power amplifiers (SMPAs) are PAs that are driven by a full-swing signal. They do not draw static bias current and therefore they can achieve a theoretical efficiency of 100%. The most important SMPA classes, classes D and E, are described and compared below.

2.2.1.1 Class-D Amplifier

A class-D amplifier can be implemented as an inverter with large transistors in order to be able to drive the low-resistive load $R_L$, as shown in Fig. 2.4a. Ideally (i.e. in the absence of parasitic capacitance and resistance), a class-D PA achieves 100% efficiency over an infinite bandwidth when driven with a full-swing square-wave input: At each time, one transistor is on while the other is off, so that current can only flow through the load resistor $R_L$. Thus if the PA conversion efficiency $\eta_{\text{conv}}$ is defined as the ratio of the total output power to the total power drawn from the supply [55], then $\eta_{\text{conv}} = 100\%$.

However, in this case the output signal is also a square wave. Generally, the desired signal is a sinusoid, namely the fundamental component of the square wave. The higher-order harmonics are only included in the input signal in order to turn it into a square-wave so that it is suitable to drive the class-D PA. Hence, all power dissipation at harmonic frequencies is undesired and should be taken into account in the efficiency. This is done by using the drain efficiency $\eta_d$, which is defined as
2.2 Power Amplifier

Fig. 2.4 Class-D amplifier: (a) wideband amplifier; (b) tuned version to prevent harmonic power dissipation

Fig. 2.5 Normalized drain-source voltage and current for the nMOS transistor in the tuned class-D PA in Fig. 2.4b [55]. The voltage is normalized with respect to $V_{dd}$ and the current with respect to $V_{dd}/R_L$, i.e. the plotted waveforms are $V_{ds}/V_{dd}$ and $I_{ds}R_L/V_{dd}$. The waveforms for the pMOS transistor are the same but are shifted by half a carrier period.

It follows that the drain efficiency of the class-D PA shown in Fig. 2.4a is significantly lower than 100%. This can be solved by adding an $LC$ bandpass filter as shown in Fig. 2.4b [55], which has zero series impedance at the desired carrier frequency and ideally an infinite impedance at the harmonic frequencies. In this case, no current can flow at the harmonic frequencies so that the output current through $R_L$ is purely sinusoidal and $\eta_d$ becomes 100%. However, the PA now has a relatively narrow bandwidth rather than the infinite bandwidth mentioned before.

Figure 2.5 shows the drain-source voltage and drain-source current waveforms for the nMOS transistor in an $LC$-tuned class-D PA. During the first half of the carrier period, the nMOS transistor is switched off, so that no current flows through it and the pMOS pulls the voltage to the supply voltage $V_{dd}$. During the second half, the nMOS is turned on so that its drain-source voltage goes to zero. Since the pMOS is now turned off and since the current through $R_L$ is purely sinusoidal, the current through the nMOS transistor is now half a period of a sinusoid. During the first half of the period, the other half of the sinusoidal current flows through the pMOS transistor.
It can be shown that the amplitude of the sinusoidal current is equal to $2V_{dd}/(\pi R_L)$, which in the figure was normalized to $2/\pi$.

In practice, CMOS transistors exhibit significant parasitic capacitances and resistances. Apart from putting an upper bound on the operating frequency (regardless of the $LC$ filter), these also reduce the efficiency: The resistance dissipates some of the power that flows through the transistors, and the capacitance of the nMOS is charged to the supply voltage $V_{dd}$ when the pMOS transistor is conducting. When the input signal switches, the charge on this capacitance is drained to the ground and the corresponding energy is wasted [55, 68]. This is known as hard switching, which refers to the fact that there is a nonzero voltage over the transistors at the time they are turned on. This last problem can be solved by tuning out the parasitic capacitance using a narrowband resonant network, but this further restricts the operation frequency range of the PA [68].

### 2.2.1.2 Class-E Amplifier

A basic class-E amplifier [60] is shown in Fig. 2.6. When the transistor is turned on, this creates a voltage over the inductor $L_d$ which causes an increasing current through $L_d$ and the transistor. When the transistor is turned off, this current flows into the capacitor $C_d$ and the load resistance $R_L$.

The combination of $L_d$ and $C_d$ causes an oscillation which causes the output voltage to go up to $3.5V_{dd}$ and then decrease again as shown in Fig. 2.7 [55]. As the voltage decreases, the capacitor $C_d$ is discharged into $R_L$ rather than into the transistor as in a class-D amplifier, so that a theoretical conversion efficiency of 100% is reached. When the voltage reaches zero, the transistor can be turned on while there is no voltage across it. This is known as soft switching or zero-voltage switching. Because of this, there is no switching loss as in class-D amplification. This is an important advantage of class-E PAs. Since the switching losses caused by hard switching are proportional to the frequency, this advantage becomes more significant for high operating frequencies.

However, it should be noted that soft switching only occurs when the PA is switched at the frequency determined by $L_d$ and $C_d$ [49]. For this reason, a class-E PA is tuned to a fixed frequency and must be driven at or close to this frequency in
order to be efficient. Ideally, this is not the case for class-D PAs, but if their parasitic capacitance needs to be tuned out, these PAs also become tuned to a certain frequency as explained before. In class-E amplifiers, the parasitic capacitance of the transistor can be used as part of $C_d$ so that no tuning is required to compensate it. This is another advantage of class-E PAs.

Finally, a class-E PA requires only one power transistor compared to two transistors for a class-D PA. Furthermore, the pMOS transistor is usually larger due to its higher on resistance compared to an nMOS transistor of the same size. Because of this, a class-D PA presents a significantly higher input capacitance than a class-E PA [55]. Therefore it typically needs more driver stages, which also consume power. Thus the overall efficiency $\eta_{oa}$ [55], which includes the power consumed by the drivers, will typically be higher for a class-E PA.

While a class-E PA does not produce a square wave at its output, the output waveform is not a pure sinusoid and therefore it still contains a lot of harmonic power, which reduces the drain efficiency. Similar to a class-D PA, this can be solved by preventing the harmonic power from being dissipated using an $LC$ filter at the output, as shown in Fig. 2.6 [55].

### 2.2.1.3 Conclusion

While both class-D and class-E PAs are highly efficient, class-E PAs are more suitable for high-frequency operation due to their soft-switching property [51]. However, their operation relies on the input consisting of a square wave with fixed or nearly fixed pulse widths and frequency. This reduces the degrees of freedom for the digital design.

The test chip presented in Chap. 5 is designed to drive a class-E amplifier as this poses the most stringent requirements on the digital modulation scheme. Thus,
designing for a class-E PA will result in signals that can also be used to drive a class-D PA.

The test chip presented in Chap. 6 contains an RF PWM modulator (see Sect. 2.4.4) to tackle the problem of the high harmonic distortion peaks produced by baseband PWM systems (Sect. 2.4.3). Since an RF PWM modulator produces pulses with varying widths, it is less suitable to drive a class-E PA, and for this reason the PAs included on this chip are class-D PAs.

### 2.2.2 Differential PA and Power Combining

Due to the low supply voltages, achieving an output power on the order of 1 W in standard CMOS is generally not possible using a single PA [56]. In order to increase the output power, a differential PA can be used. In such a configuration, two identical PAs are connected to both sides of the load and are driven in counterphase. This way, the current flows from one PA to the other and the output voltage swing is doubled. Furthermore, the DC component and all even harmonics resulting from the switched-mode behaviour or from other sources are cancelled out.

The output power can be further increased using a power combiner, which combines the power from multiple, often differential PAs [55]. Many types of power combiners exist; as an example, an inductive power combiner [13, 54, 56] using two differential PAs is shown in Fig. 2.8.

When implementing digital transmitters using SMPAs, the often inevitable need to use power combining can be used as an advantage for the digital modulator design: Each additional PA provides an extra signal level, so that multibit digital signals can be used to drive the PAs.

However, it should be noted that a power combiner is much more complicated than the ideal analog adder one would like to have: The output swing and efficiency of a PA depend on its load impedance, but through the power combiner, the other PAs’ outputs contribute to this impedance and therefore the load for one PA depends on the input signals to all other PAs. This is known as electronic load pull [55] and can

![Fig. 2.8 Two differential (push-pull) PAs with inductive differential power combiner](image-url)
cause distortion on the output signal. In order to control this, careful PA and power combiner design is required, which is outside the scope of this work. Possibly, the analog circuits can be digitally assisted by precompensating some distortion.

Controlling the output impedances is feasible when the carriers amplified in the different PAs are in phase and only the AM is different. However, if the carriers are not in phase, as is the case in e.g. outphasing transmitters (see Sect. 2.3.3), this becomes much more difficult [55].

2.3 Modulator Types

This section presents the three main types of RF modulators, which differ in the way they represent the complex envelope using real signals and the way they convert this to the desired RF signal. These types are the quadrature modulator (Sect. 2.3.1), the polar modulator (Sect. 2.3.2), and the outphasing modulator (Sect. 2.3.3).

2.3.1 Quadrature Modulator

A quadrature modulator is based on (2.7): It has two input signals $i(t)$ and $q(t)$, both in the range $[-1, 1]$, which are multiplied with two versions of the RF carrier that are $90^\circ$ out of phase, and then summed, as shown in Fig. 2.9a. Using the complex representation, this means that the complex envelope $g(t)$ is created from the real signals $i(t)$ and $q(t)$ according to (2.11), which is shown in Fig. 2.9b.

Quadrature modulators have several advantages compared to polar or outphasing modulators. First, the bandwidth of the signals $i(t)$ and $q(t)$ is equal to the bandwidth of the complex envelope $g(t)$ (measured from DC to the maximal frequency). This bandwidth is specified by the communication standard and enforced either during the generation of the signals or using filters. As will be explained below, polar and outphasing modulators use different signals which have a nonlinear relationship to $i(t)$ and $q(t)$ and hence usually a significantly wider bandwidth.

Another advantage is the symmetry of the architecture: The I and Q paths are identical and therefore relatively easy to match.
Several ways of digitally implementing a quadrature modulator exist in literature. One implementation is purely digital and performs all operations, including mixing, in the discrete-time digital domain. For a carrier frequency \( f_c \), the sampling rate \( f_s \) of the RF output should be at least equal to the Nyquist rate which is equal to \( 2(f_c + B) \) in order to represent the RF signal without aliasing. Since in this work, carrier frequencies from 1 to 3 GHz are targeted, this means the required sampling rates are at least 2–6 GHz, which is very high. Furthermore, some margin is required on the Nyquist rate which leads to even higher sampling rates.

However, if \( f_s \) is exactly \( 4f_c \) as in [15, 16, 37, 69], the I and Q carriers correspond to the sequences \( \{1, 0, -1, 0, \ldots\} \) and \( \{0, 1, 0, -1, \ldots\} \), respectively, as shown in Fig. 2.10. Thus, the adder becomes a trivial component since at any moment one of its input signals is 0. Furthermore, the multiplier only needs to be able to multiply with 1 and \(-1\). Finally, since every second sample will be multiplied by 0, these samples need not be produced, so that the multipliers and preceding circuits can operate at \( 2f_c \) rather than \( 4f_c \) [15].

The input signals \( i(t) \) and \( q(t) \) can be encoded into two-level signals with levels \( \pm 1 \) using several coding schemes such as baseband pulse width modulation (PWM) [37] (Sect. 2.4.3) or baseband \( \Delta \Sigma \) modulation [15, 16, 69] (Sect. 2.4.1). In this case, the multiplier only needs to be able to either replicate its input signal or invert it. Furthermore, the output signal \( v(t) \) will also be a two-level signal and can directly drive a switched-mode PA (SMPA). Thanks to these simplifications of the adder and the multipliers, carrier frequencies around 1 GHz are feasible [15, 16]. However, higher carrier frequencies are more difficult to achieve.

Due to the interleaving of the I and Q signals, the RF output signal has quite variable pulse widths as can be seen in Fig. 2.11. Therefore it is not suited to be amplified efficiently by a class-E PA, which requires pulses with approximately constant width and frequency as was explained in Sect. 2.2.1.2. Thus, either a wideband class-D PA or a linear PA preceded by a bandpass filter is needed to amplify this type of signals.

Several other implementations exist in order to solve the high sampling rate problem. For example, [30, 35, 36] use the above principle to upconvert the signal to an intermediate frequency \( f_{IF} \), which is much lower than \( f_c \). The sampling rate \( f_s \) is then equal to only \( 4f_{IF} \). After this upconversion, the signal in [35, 36] is further
2.3 Modulator Types

Fig. 2.11  Signals in digital quadrature modulator. a I input (dashed line) and upconverted I input; b Q input (dashed line) and upconverted Q input; c modulator output towards PA (sum of upconverted I and Q inputs). Ticks on the x-axis indicate periods of the unmodulated carrier

Fig. 2.12  DRFC-based quadrature transmitter. LO indicates the local oscillator which generates the carrier frequency; $Z_L$ is a load impedance to convert the output current to a voltage

upconverted to $f_c$ using a DAC and analog components. Hence this is not a fully digital transmitter and the PA needs to be linear.

In [12, 19, 30], DAC and mixer are combined in a so-called digital-to-RF converter (DRFC), which converts a digital signal to a sinusoidal current where the amplitude is determined by the digital signal. The addition of the I and Q paths is done simply by adding currents as shown in Fig. 2.12. Such current summing can also be used to add the currents from multiple binary-weighted DRFCs to implement a multibit DRFC, which enhances the resolution. Parikh et al. [45] present a similar architecture where the I and Q paths are summed digitally inside a single quadrature DRFC. Gaber et al. [18] even use the current summing in combination with clocked delay elements to implement a finite impulse response (FIR) bandpass filter within a DRFC.

DRFC-based architectures avoid needing to sample any RF signals as the signals are directly converted from digital baseband to analog RF signals. They are very practical in cases where low output power is used since then the DRFC can be used
as a PA. However, if higher output power is needed, an additional PA must be added. This cannot be an SMPA since its input is a multilevel analog signal.

To conclude, while quadrature modulators have many advantages, they either require very high sampling rates or linear power amplifiers (if DRFCs are used), which makes them unsuited for the applications targeted in this work. Furthermore they are not compatible with class-E PAs. For these reasons, polar modulators were preferred in this work. These are discussed in Sect. 2.3.2.

2.3.2 Polar Modulator

A polar modulator is based on (2.6): It has an amplitude input \( a(t) \in [0, 1] \) and a phase input \( \varphi(t) \) in the range \( [0, 2\pi) \) (or equivalently \( [-\pi, \pi) \)). The phase signal \( \varphi(t) \) is phase-modulated onto the RF carrier, which is then multiplied by \( a(t) \), as shown in Fig. 2.13a. Using the complex representation, this means that the complex envelope \( g(t) \) is created from the real signals \( a(t) \) and \( \varphi(t) \) according to (2.12), which is shown in Fig. 2.13b.

The first polar transmitter was introduced by Kahn [31] under the name envelope elimination and restoration (EER), which was later known as a Kahn transmitter. It is a completely analog transmitter, which separates the amplitude and phase information of the incoming analog RF signal. This way, the phase-modulated carrier (PMC) does not have any AM and can be amplified using an SMPA. The amplitude is restored by modulating the supply power of the SMPA through a second, linear PA. Since this PA operates in baseband, it is much easier to implement.

The Kahn transmitter can be converted to a digital transmitter architecture by coding the amplitude signal into a single-bit signal using e.g. baseband \( \Delta \Sigma \) [1, 7, 25, 40, 65, 70] or baseband PWM [1, 25, 33, 42–44, 67]. In [63, 64], a combination of multi-bit \( \Delta \Sigma \) modulation and ROM-based RF pulse density modulation (PDM) is used. This single-bit signal can then be used to turn the SMPA’s supply voltage on and off as shown in Fig. 2.14a [7], or it can be multiplied with the PMC before the PA as in Fig. 2.14b [1, 25, 29, 33, 40, 42–44, 56, 65, 70]. This way, only an SMPA is required and the linear PA is removed.

Alternatively, a so-called digital PA or can be used [6, 46, 61, 62], which has a multibit digital amplitude input and an RF input for the PMC, which is a square
2.3 Modulator Types

Fig. 2.14 Fully digital polar modulators: a AM through supply of PA; b AM through multiplication before PA

Fig. 2.15 Signals in digital polar modulator for differential PA. a Phase-modulated square wave carrier; b Amplitude modulated as 0/1 signal (dashed line) and modulator output towards PA (product of phase-modulated carrier and amplitude signal). Ticks on the x-axis indicate periods of the unmodulated carrier

wave. This is similar to the DRFC mentioned in Sect. 2.3.1, but now the PMC is used as RF input rather than an unmodulated reference carrier.

The PMC can be generated in different ways. Some implementations [1, 7, 25, 29] use baseband I/Q signals $i(t)/a(t)$ and $q(t)/a(t)$ which are normalized so that the resulting carrier has a constant amplitude. However, this presents the same problems as the aforementioned quadrature modulator, namely a high $f_s$ and a signal that is not suited for class-E amplification.

Another possibility is to modulate the phase directly onto the carrier. This is often done using a voltage-controlled (VCO) or digitally controlled oscillator (DCO), mostly with feedback to implement a phase-locked loop (PLL) [39–41, 61, 63–65]. In this case, the carrier is phase-modulated while it is created. However, it is also possible to phase-modulate an existing carrier, as is demonstrated in [42–44, 52, 53], and in Chaps. 5 and 6 of this work.

In implementations where the carrier is directly phase-modulated, the RF signal is a square-wave PMC multiplied with the single-bit amplitude signal as shown in Fig. 2.15. Now the PA input consists of bursts of carrier pulses whose width and frequency are only slightly changing due to the phase modulation. This leads to the name burst-mode amplification. These bursts can be amplified efficiently using any type of SMPA including class E. Between the bursts, the PA is turned off and ideally does not consume any power.
This is an important advantage of polar modulators: At zero amplitude, the PA is turned off, and at low amplitudes it is turned off most of the time. This is in contrast to quadrature transmitters, which represent 0 by oscillating between 1 and −1, and to outphasing transmitters, which realize low amplitudes by summing carriers that are nearly 180° out of phase. This leads to a lower efficiency at lower amplitudes.

The efficiency benefit in polar transmitter architectures is confirmed by evaluating the coding efficiency, which is defined as the ratio of the desired in-band power in the RF signal to its total power [2, 4]. As shown by Blocher and Singerl [2], the coding efficiency of a polar modulator is higher than for a quadrature modulator. Combined with the increased PA efficiency, this results in an overall efficiency improvement. For these reasons, the following chapters will focus on polar modulators.

However, polar modulators also have some disadvantages compared to quadrature modulators. First, the input signals $a(t)$ and $\varphi(t)$ as well as the PMC $p(t)$ have significantly larger bandwidths than $i(t)$ and $q(t)$ due to the nonlinear relationships given by (2.8)–(2.9). Typically, the bandwidths are about 3–6 times the I/Q bandwidth $B$ [56]. This limitation can often be overcome without major problems in modern CMOS technologies thanks to the increasing transistor speed.

Another disadvantage is the inherent asymmetry of the architecture: The amplitude and phase paths are fundamentally different and it may be difficult to accurately match their delays, especially in an analog implementation. However, in the baseband PWM test chips presented in Chaps. 5 and 6, both paths are implemented using similar components (mainly delay elements, multiplexers, and logic gates) and it will be shown that the delays of both paths can be matched quite accurately by including a number of dummy components.

The most important disadvantage in the context of this work is the fact that both paths are combined in a multiplier rather than an adder. This is because addition is a linear operation while multiplication is not: If two signals are added, the resulting spectrum is simply the sum of both spectra. Thus, any out-of-band noise or distortion produced by the nonlinear components on both paths remains outside the signal band after addition of both signals. When two signals are multiplied, however, the resulting spectrum is the convolution of both spectra (convolution is defined in Sect. A.2), which can introduce contributions at different frequencies than those in the original signals. Thus, if both paths cause out-of-band spectral components, the multiplication of both signals may move some of these into the signal band. Such effects can be very significant and will be analyzed in detail for the case of baseband PWM in Sect. 3.4.

This observation has important implications: For example, a quadrature modulator may include ΔΣ modulators (see Sect. 2.4) on both paths to quantize the signals while moving most of the quantization noise outside the signal band, where it can be filtered out. In a polar modulation, a ΔΣ modulator can be included on the amplitude path or the phase path, but not both: In this case, both ΔΣ modulators would shape the quantization noise away from the signal band, but the convolution of both spectra would destroy this desirable shape and move a large amount of quantization noise back into the signal band. Even when only the amplitude path is ΔΣ-modulated, the multiplication with the PMC corrupts the desired noise shape since the PMC has nonzero bandwidth [1, 23].
2.3 Modulator Types

Jeong and Wang [29] solved this ΔΣ problem by implementing a polar ΔΣ modulator, which generates both the amplitude signal and the normalized I/Q signals that constitute the PMC. The ΔΣ modulator operates on the original I and Q inputs instead of the amplitude signal only. This avoids distortion resulting from intermodulation of the quantization noise with the phase information, at the expense of a more complicated system where the quadrature-to-polar converter is part of the ΔΣ loop. A similar architecture is presented in [23], where it is called complex ΔΣ modulation.

Another solution is using RF PWM as proposed in [39, 41, 44, 48, 68]. Here, the width of every single RF carrier pulse is modulated, which avoids harmonic peaks close to the signal band as produced by baseband PWM, and instead produces only harmonics of the PMC, making this an interesting idea with regard to spectral mask requirements. Furthermore, the multiplication is removed since the signal is upconverted in the process of pulse width modulation as described in Sects. 2.4.4, 3.5, and 3.6. For this reason, one can debate whether this qualifies as a polar modulator. An RF PWM implementation will be presented in Chap. 6. However, it will be shown that this is currently only efficient for moderate dynamic ranges [68].

Park et al. [46] used RF PWM to refine the resolution of a digital PA. Together, this can be seen as an implementation of multilevel RF PWM (see Sect. 3.7.2).

2.3.3 Outphasing Modulator

Assuming \( a(t) \in [0, 1] \), one can define \( \theta(t) \in [0, \pi/2] \) so that

\[
a(t) = \cos \theta(t)
\]

\[
= \frac{1}{2} \left( e^{j\theta(t)} + e^{-j\theta(t)} \right).
\]

Equation (2.12) can then be rewritten as

\[
g(t) = \frac{1}{2} \left( e^{j(\varphi(t) + \theta(t))} + e^{j(\varphi(t) - \theta(t))} \right),
\]

which leads to

\[
v(t) = \frac{1}{2} \left( \cos (2\pi f_c t + \varphi(t) + \theta(t)) + \cos (2\pi f_c t + \varphi(t) - \theta(t)) \right)
\]

\[
= \cos \theta(t) \cos (2\pi f_c t + \varphi(t)).
\]

This is called the outphasing representation of the modulated carrier. The signal \( \theta(t) \) is called the outphasing angle.

This principle can be used to create an outphasing modulator [5], as shown in Fig. 2.16a. This modulator starts from the inputs \( \varphi(t) \in [0, 2\pi) \) and \( \theta(t) \in [0, \pi/2] \).
and converts these to \( \varphi(t) + \theta(t) \) and \( \varphi(t) - \theta(t) \). These angles are phase-modulated onto two different carriers which are then added up. Using the complex representation, this means that \( g(t) \) is generated from \( \varphi(t) \) and \( \theta(t) \) according to (2.24), which is shown in Fig. 2.16b.

Outphasing transmitters usually use a PA on each phase path and a power combiner to add both PA outputs as shown in Fig. 2.17 [17, 52, 53, 66, 72]. This has the advantage that both PAs are driven by constant-amplitude signals and hence SMPAs can be used. Furthermore, some of the disadvantages of polar modulators are removed: There is no more multiplication, and both signal paths are equal and thus easier to match. The increased bandwidth compared to a quadrature transmitter remains, however.

A more important disadvantage is the fact that the power combiner needs to combine the power from 2 PAs which produce different signals. The varying signal in one PA causes its output impedance to change, which affects the other PA, and vice versa, as explained in Sect. 2.2.2. This makes it hard to achieve linear behaviour for the power combiner [55]. The linearity improves if an isolating power combiner is used, but this type of combiner dissipates the power from both carriers that is not transmitted, so that the efficiency at low amplitudes becomes very low [20, 26]. Finally, when the amplitude is small, it becomes very sensitive to slight errors in the phases of both PMCs, since the operation of subtracting two large numbers to obtain a small one is ill conditioned.

For the above reasons, pure outphasing as described here was not further considered in this work. Further research on digital modulators in combination with PAs and power combiner is required in order to determine the feasibility of fully digital outphasing modulators.
However, the outphased PMCs can also be added before the PA, in which case only one PA is required. But since the sum of both PMCs is a carrier with both PM and AM, it can normally not be amplified using an SMPA. There is, however, a special case where both PMCs as well as their sum are square waves. As will be shown in Sect. 3.5.4, this case corresponds to differential RF PWM, where the PWM signal is generated using an outphasing implementation. This principle was demonstrated by Walling et al. [68] and will also be used in Chap. 6 to implement an outphasing modulator [44].

### 2.4 Types of 1-bit Coding

In order to drive an SMPA, the amplitude needs to be encoded in a single-bit signal. This can be done in several ways, which result in different types of out-of-band noise and distortion. The most important coding schemes are discussed in this section.

#### 2.4.1 Baseband Delta-Sigma Modulation

A very simple first-order baseband $\Delta \Sigma$ modulator [1, 7, 58] is shown in Fig. 2.18a. The quantized output $Y(f)$ is fed back and subtracted from the input $U(f)$ to calculate the error that is made from the input to the output. This error is filtered by the loop filter with transfer function $H(f)$. In the case of a first-order baseband $\Delta \Sigma$ modulator, this filter is an integrator, which integrates the error made over time. When the integrated error becomes very large in absolute value, it will cause the quantizer’s output to change, after which the instantaneous error’s sign will change and the integral’s absolute value will decrease again. This way, the output signal $Y(f)$ will constantly switch between its quantization levels to keep the integrated error (i.e. the average error) as small as possible. This is in contrast to a simple quantizer without feedback, which keeps the instantaneous error as small as possible. Thus, while a $\Delta \Sigma$ modulator causes larger instantaneous errors, the error power is mostly located at higher frequencies, as will be shown next. This concept is called noise shaping, which means noise power is reshaped so it becomes less problematic.
Figure 2.18b shows a linearized model of the ΔΣ modulator, where the quantizer is modeled by an independent additive quantization noise source $\Delta_q Y(f)$. While it is important to note that this model is far from exact and should be handled with care [3], it is useful in order to understand the general operation principle of ΔΣ modulation.

From Figure 2.18b it can be seen that the modulator output $Y(f)$ is given by

$$Y(f) = STF(f) \cdot U(f) + NTF(f) \cdot \Delta_q Y(f),$$

(2.27)

where

$$STF(f) = \frac{H(f)}{1 + H(f)}$$

(2.28)

is called the signal transfer function and

$$NTF(f) = \frac{1}{1 + H(f)}$$

(2.29)

is called the noise transfer function.

If $H(f)$ is an integrator, its gain is high at low frequencies, so that $STF(f) \approx 1$ and $NTF(f) \approx 0$ within the signal band, and the in-band quantization noise is reduced.

Outside the signal band, $NTF(f)$ behaves as a highpass filter so that the noise increases when moving away from the signal band. In a first-order ΔΣ modulator, the integrator $H(f)$ is a first-order lowpass filter, so that the quantization noise increases with 20 dB/decade.

Higher-order ΔΣ modulators can be implemented by replacing $H(f)$ with a higher-order lowpass filter. This way, very high signal-to-noise ratios (SNRs) can be achieved. However, in order to avoid instability, feedback or feedforward paths need to be added to internal nodes of the filter [58], so that the model shown in Fig. 2.18 is no longer valid. Apart from the filter order, the SNR also increases with the oversampling ratio $OSR = f_s/(2B)$, where $B$ is the baseband signal bandwidth, and with the number of quantization levels in the quantizer. However, when the ΔΣ is used to produce an SMPA input signal, this number is typically fixed at 2, or at 3 if a differential PA configuration is used.

For an $n$th-order baseband ΔΣ modulator, the quantization noise increases with $20n$ dB/decade. Note that when $H(f)$ becomes smaller than 1, $NTF(f)$ starts to saturate at 1. When this ΔΣ modulator is used in a digital-to-analog converter (DAC), the out-of-band noise can be removed using an analog lowpass filter. For an $n$th-order ΔΣ, an $n$th-order filter is needed to flatten the noise spectrum, and if needed a higher-order filter can be used to further reduce it. The remaining signal then consists of the desired signal and the very low in-band quantization noise contribution.

Single-bit baseband ΔΣ modulation can be used on the I and Q paths of a quadrature transmitter [16, 69] or on the amplitude path of a polar transmitter [1, 7, 25, 40, 65, 70]. In both cases, the single-bit ΔΣ output is first digitally upconverted to the carrier frequency $f_c$ and amplified by an SMPA, as is illustrated in Figs. 2.10
2.4 Types of 1-bit Coding

and 2.14. Afterwards, a bandpass filter is needed to filter out the noise. This causes a major problem [34]: After upconversion, the quantization noise no longer increases with $20n$ dB/decade as can be seen from the following example.

Assume a first-order $\Delta \Sigma$ modulator operating on a baseband signal with a bandwidth from 0 to 10 MHz. Assume further that the in-band quantization noise is at an acceptable level which is normalized to 0 dB. Since the out-of-band noise increases with 20 dB/decade, the noise reaches 20 dB at one decade away from the signal band, i.e. at 100 MHz.

After upconversion to $f_c = 1$ GHz, the signal band ranges from 0.99 to 1.01 GHz, and the noise reaches 20 dB at 0.9 and 1.1 GHz. A second-order bandpass filter centered at $f_c$ has a transfer function that decreases with 20 dB/decade at both sides of the signal bands. This means it reaches $-20$ dB at 100 MHz and at 10 GHz. But in order to reduce the noise to 0 dB, it should reach $-20$ dB already at 0.9 and 1.1 GHz, which is about 0.04 decades away from $f_c$. Thus, a 50th-order bandpass filter (which decreases at 500 dB/decade) is required to accomplish the task. Clearly, implementing such a filter on-chip is not realistic, so that mostly an external ceramic filter is required.

A second problem arises from the fact that upconversion to $f_c$ also implies upconversion to $-f_c$. When moving the signal and the shaped noise to $-f_c$, the noise that was at $2f_c$ moves to $f_c$ and ends up in the signal band, as illustrated in Fig. 2.19a, b. Similarly the noise from $-2f_c$ ends up at $-f_c$. In the above example, where the bandwidth is 10 MHz and $f_c = 1$ GHz, this noise is 46 dB higher than the original in-band quantization noise, which is unacceptable. This can be solved by filtering the signal before upconversion, but this generally causes it to lose its desirable square-wave shape.

However, the $\Delta \Sigma$ modulator is usually implemented in the digital domain at a certain sampling rate $f_s$. This causes the spectrum to be periodical with period $f_s$ as in Fig. 2.19c (see Sect. A.1.2), which implies that the shaped noise is maximal at $f_s/2$ and goes back to 0 at $f_s$. When it is converted to the analog domain using a zero-order hold operation, this is reflected in the frequency domain by a multiplication with a sinc function (see Sect. A.3.1), which is indicated by the dashed line in Fig. 2.19d. Now the effect of upconversion depends on the ratio of $f_c$ and $f_s$. Figure 2.19e shows the case where $f_c = f_s/2$. The upconversion shifts some quantization noise and some spectral replicas into the signal band, but both are small since the quantization noise was already going to 0 and both are filtered by the sinc function. Thus, the in-band effect is much lower.

If $f_c$ is equal to $f_s$, or some higher multiple of $f_s/2$, the effect becomes even smaller since the sinc function decreases with frequency. Thus, in most practical implementations, this effect can be expected to be quite limited as $f_c$ is often equal to $f_s$ [1, 25] or even much larger [59, 70].
Fig. 2.19 Illustration of the image problem in baseband ΔΣ modulation: a Analog baseband ΔΣ signal. b Signal from a after upconversion to $f_c$. c Digital baseband ΔΣ signal: The spectrum is periodic with period $f_s$. d Signal from c after converting to the analog domain using zero-order hold (ZOH). The dashed line indicates the sinc response that results from the ZOH operation. e Signal from d after upconversion to $f_c$ where $f_c = f_s/2$

2.4.2 Bandpass Delta-Sigma Modulation

Bandpass $\Delta\Sigma$ modulation [27, 57] is based on the same principle as baseband $\Delta\Sigma$ modulation, but now the filter $H(f)$ in (2.28)–(2.29) is a bandpass filter, so that $\text{NTF}(f)$ becomes a bandstop filter. The signal can now be upconverted to $f_c$ in the digital domain and is converted to a single-bit signal afterwards, as shown in Fig. 2.20.

If $H(f)$ is a second-order bandpass filter, its gain decreases at 20 dB/decade on both sides of the signal band, so that the noise will increase at 20 dB/decade. However, the decades now are decades with respect to $f_c$: If $f_c = 1$ GHz, the noise will have increased by 20 dB at 100 MHz and 10 GHz, and hence it can be filtered out using a second-order bandpass filter. This is the main advantage of bandpass $\Delta\Sigma$ compared to baseband $\Delta\Sigma$. 
2.4 Types of 1-bit Coding

The OSR is still defined as $f_s/(2B)$ where $B$ is the original baseband signal bandwidth so that $2B$ is the RF signal bandwidth. Thus a high OSR can still be reached with a moderate $f_s$. However, it is clear that $f_s$ should be at least equal to the Nyquist frequency $2(f_c + B)$ in order to correctly represent the RF signal.

Often, $f_s = 4f_c$ is used as this still allows easy quadrature upconversion using the series $\{1, 0, -1, 0, \ldots\}$ and $\{0, 1, 0, -1, \ldots\}$, respectively, as shown in Fig. 2.20. However, this often means $f_s$ needs to be very high. Alternatively, the complete RF signal can be generated digitally at or slightly above the Nyquist frequency $2(f_c + B)$. Then the carriers cannot be represented by trivial sequences so that multibit signals are required.

In either case, $f_s$ needs to be above $2f_c$, which is very high for GHz-range transmitters. This is an important disadvantage of bandpass $\Delta \Sigma$ transmitters, which is not present in baseband $\Delta \Sigma$ or in PWM transmitters. While some bandpass $\Delta \Sigma$ transmitters are found in literature [28, 32], no references with $f_c > 1$ GHz have been found [40].

2.4.3 Baseband PWM

Instead of using baseband $\Delta \Sigma$ modulation, the single-bit quantizers in Figs. 2.10 and 2.14 can be implemented using pulse width modulation (PWM). Such a transmitter will be called a baseband PWM transmitter since the pulse width modulation occurs entirely in baseband.

A PWM signal has a PWM frequency $f_{pwm}$ and contains exactly one pulse per period $T_{pwm} = 1/f_{pwm}$, with pulse widths varying from 0 to $T_{pwm}$. Ideally, it produces a nearly noise-free copy of the input signal and adds harmonic distortion peaks at all multiples of $f_{pwm}$. This will be shown in Chap. 3, where additional nonidealities will be examined in detail. Thus, the shaped noise which was present in $\Delta \Sigma$ modulation is now replaced with isolated harmonic peaks.

Just as for baseband $\Delta \Sigma$, the output of the PWM modulator is upconverted to $f_c$ as in Figs. 2.10 and 2.14. This causes similar problems as for baseband $\Delta \Sigma$: First, after upconversion some of the harmonics may fall into the signal band, but this can
Fig. 2.21 Different oversampling quantizers representing the same constant level: a baseband $\Delta \Sigma$; b baseband PWM. The gray lines represent the carrier pulses after multiplication with a single-ended carrier.

...be avoided by optimizing the ratio $f_c/f_{pwm}$ as will be shown in Sect. 3.4.4. Second, on a logarithmic axis, the upconversion brings the harmonics much closer to the signal band which makes them difficult to filter out. It will be shown in this work that this is the most fundamental limitation of baseband PWM transmitters.

Nevertheless, baseband PWM modulators have several advantages compared to baseband $\Delta \Sigma$ modulators. First, they do not require any feedback paths which makes them more robust and guarantees their stability. Furthermore, it allows implementing them using continuous-time building blocks such as delay lines (see Chap. 4), so that they can achieve very good resolution (down to about 4 ps in 40-nm CMOS) using a sampling rate $f_s = f_{pwm}$ that is an order of magnitude below $f_c$. This will be demonstrated in the following chapters. While the resulting in-band quantization noise is not as low as with $\Delta \Sigma$ modulation, it will be shown that the achieved resolution is good enough in order not to be a limiting factor in the transmitter performance.

For a $\Delta \Sigma$ modulator, a continuous-time digital implementation is very difficult due to the feedback loop and the fact that multibit adders and accumulators are required. Hence the time resolution is equal to the sampling period $T_s$, which is at least in the order of several hundreds of ps. Thus, $\Delta \Sigma$ transmitters achieve a lower resolution than PWM transmitters while requiring a higher sampling rate. While the in-band effect of the lower resolution is compensated by the higher OSR (see below) and the noise shaping, the total quantization noise power will be larger so that more out-of-band power will appear due to the quantization.

Second, while baseband $\Delta \Sigma$ modulators have a tendency to oscillate between both output levels, as shown in Fig. 2.21a, PWM guarantees that only one rising and one falling edge can occur per PWM period $T_{pwm}$ [67]. In a $\Delta \Sigma$ modulator, up to one edge can occur per $T_s$ period, but since $\Delta \Sigma$ modulators need a high OSR in order to achieve good SNR, $T_s$ will usually be much smaller than the $T_{pwm}$ used for baseband PWM. Thus, on the average, a baseband PWM transmitter generates longer carrier bursts for the same input period, as shown in Fig. 2.21b. Since class-E PAs achieve higher efficiency if the bursts are longer [7, 14, 67], this is a serious advantage.

---

1 Continuous-time $\Delta \Sigma$ modulators have been presented in literature [9–11], but they are used in analog-to-digital rather than digital-to-analog conversion, so that the adder and integrator are implemented in the analog domain.
2.4 Types of 1-bit Coding

Fig. 2.22 Differential baseband PWM (a) versus differential RF PWM (b)

A detailed mathematical analysis of baseband PWM transmitters and their spectral effects will be presented in Sects. 3.3 and 3.4, and two silicon implementations are presented in Chaps. 5 and 6.

2.4.4 RF PWM

As mentioned before, PWM produces harmonics at every multiple of the PWM frequency $f_{\text{pwm}}$. As will be shown in Chap. 3, the amplitude of these harmonics depends on the duty cycle. This fact can be used to implement the upconversion as a part of the PWM process: If one chooses $f_{\text{pwm}} = f_c$, the first harmonic is an amplitude-modulated RF signal. Thus, rather than considering the baseband part of the PWM output as the signal and upconverting it to $f_c$ while considering the harmonics as distortion, one can also consider the component at $f_c$ as the signal, which is already at RF so that no upconversion is required. The fundamental difference between baseband and RF PWM is illustrated in Fig. 2.22. The figure assumes the sampling rate $f_s$ is the same for both modulators, and is also equal to the baseband PWM frequency. However, the sampling rate in RF PWM can go up to $f_c$ (in certain cases even 2$f_c$ or 4$f_c$ as will be explained in Chap. 3).

In Sects. 3.5.4 and 6.1.2 it will be shown that the RF PWM signal can be phase modulated, so that a carrier with both AM and PM is produced.

It may appear as though an RF PWM transmitter is just a special case of a baseband PWM transmitter where the PWM frequency $f_{\text{pwm}}$ is taken to be equal to $f_c$ or 2$f_c$. While there is indeed some similarity, the reality is somewhat more complicated as is explained in more detail in Sect. B.2 in App. B. The conclusion from this section is that $f_{\text{pwm}}$ should be equal to 2$f_c$, not $f_c$, and in addition, the baseband PWM signal should be phase-modulated together with the RF carrier.

RF PWM transmitters are analyzed mathematically in Sects. 3.5 and 3.6, and a silicon implementation is presented in Chap. 6.

The main advantage of RF PWM is the high value of $f_{\text{pwm}}$: Since harmonic distortion only occurs at multiples of $f_{\text{pwm}} = f_c$, the first distortion peak is at 2$f_c$. Furthermore, in Sect. 3.5.4 it will be shown that RF PWM can be implemented
differentially as illustrated in Fig. 2.22b. In this case, all even harmonics cancel out and the first unwanted peak is at $3f_c$.

This is a very important advantage: In a baseband PWM implementation with $f_c = 1$ GHz and $f_{pwm} = 100$ MHz, the first harmonic peak is at $1.1$ GHz, which is only about 0.04 decades above $f_c$. With RF PWM, it is at $3$ GHz, which is almost half a decade above $f_c$, so that it is much easier to filter out. This makes RF PWM very interesting compared to baseband PWM as well as baseband and bandpass $\Delta \Sigma$, which all produce much more spectral components close to the signal band.

Another advantage is the fact that the multiplication that is normally present in polar modulators has been removed. This implies that there are no intermodulation effects that can bring out-of-band noise back into the signal band.

Apart from these advantages over baseband PWM, RF PWM shares some advantages with baseband PWM: It too can be implemented without feedback and hence does not cause any stability issues. Furthermore, good resolution can be achieved when using continuous-time circuits like delay lines.

However, RF PWM also has a number of disadvantages. First, as will be noted in Sect. 3.5, the RF amplitude is no longer proportional to the PWM duty cycle $d$ but to $\sin(\pi d)$. Hence, predistortion using an arcsine function is required. This is only a minor issue since it can easily be implemented in a DSP core.

Another disadvantage is that RF PWM produces pulses at a fixed frequency but with varying widths, which has a negative effect on the efficiency of tuned PAs such as class-E PAs [49, 50]. In this regard, baseband PWM may be more desirable. However, comparing to bandpass $\Delta \Sigma$, RF PWM can be expected to be superior since bandpass $\Delta \Sigma$ modulators sample at a rate that is higher than $f_c$, so that the output pulses are likely to oscillate at a rate that is not the carrier frequency for which the PA is designed.

Finally, it should be noted that for the same time-domain resolution $T_{res, RF PWM}$, RF PWM transmitters have much less pulse width quantization levels since the PWM frequency is much higher. This results in increased quantization noise, but this is compensated by a much higher sampling rate. Thus the OSR is higher which reduces the in-band contribution of the quantization noise. For this reason, depending on the system parameters, theoretical simulations such as those presented in Sect. 3.6 often show better signal quality than for baseband PWM.

In practice, however, circuit-level effects such as pulse shrinking, pulse swallowing and delay mismatch occur in any continuous-time circuit. These have a much larger effect in RF PWM due to the shorter pulse widths, which often causes them to be dominant over the quantization noise. For this reason, current RF PWM implementations are mainly suited for applications where the required dynamic range for AM is limited, as is explained in [68] and in Chap. 6.
2.4 Types of 1-bit Coding

2.4.5 Other Coding Schemes

All single-bit coding schemes discussed above have the same aim, namely to represent the complete RF signal as a two-level or (for a differential PA) three-level signal so it can be amplified efficiently using an SMPA.

Any such coding scheme inherently introduces noise and/or distortion, since otherwise the output signal would be equal to the input signal. Different coding schemes place the noise and distortion power at different frequencies and furthermore result in different efficiencies for the SMPA.

Several other coding schemes can be used in order to further optimize the spectral characteristics and/or the PA efficiency. For example, Stauth and Sanders [63, 64] designed dedicated sequences of ones and zeros for each amplitude quantization level, in order to accurately control where the noise and distortion power appears. This is an interesting approach but it requires a lookup table which can be read out at high speeds, which means a high sampling rate is needed.

Finally, some implementations avoid the single-bit quantization altogether by using e.g. a multibit digital-to-RF converter (DRFC) [12, 30, 45] or a multibit digital PA [61, 62] as was explained in Sect. 2.3. However, the research presented here started from the a priori goal to design digital modulators which can directly drive a switched-mode PA, which is not the case for these architectures.

2.4.6 Multibit Noise Shaping

Even if baseband or RF PWM is used as the single-bit quantizer, ΔΣ modulators can still be of use: Digital PWM implementations can usually only produce a discrete set of pulse widths, which means that their input signal is effectively quantized to a finite number of quantization levels, which causes quantization noise. This noise can be shaped by implementing the quantizer as a multibit digital ΔΣ modulator rather than just rounding the signal. This can also be done with different single-bit coding schemes such as in [63, 64], or with multibit DRFCs [30] or multilevel PAs [61].

Similarly, digitally implemented phase modulators are often limited to a discrete number of phases, especially when operating on square waves. This implies the phase signal is quantized, which can also be done using a ΔΣ as in [52, 53]. However, this only works well if there are enough quantization levels: Switching between two phases to achieve an intermediate phase also results in an amplitude reduction, which is larger if the phase quantization levels are further apart, as can be seen from the phasor diagrams in Fig. 2.23. This is called PM-to-AM distortion. Note that this effect can be considered as a form of outphasing. Thus, if there are few phase quantization levels, this amplitude reduction has to be compensated digitally by applying a higher amplitude at the input.
Fig. 2.23  Illustration of the amplitude distortion arising from using ΔΣ modulation on the phase path. \( a \) Many quantization levels: small amplitude reduction; \( b \) few quantization levels: large amplitude reduction

As mentioned before, care should be taken with ΔΣ modulation in polar transmitters since the multiplication may corrupt the noise shaping, in particular when both the AM and PM paths include ΔΣ modulators.

Furthermore, as will be seen in Chap. 3, when a ΔΣ modulator is used before a PWM modulator, the PWM causes additional distortion which is not shaped by the ΔΣ modulator. While this distortion is quite low if pseudo-natural-sampling PWM (PNPWM, see Sect. 3.2.2 and [21, 22]) is used, it poses an upper bound on the achievable SNR. Beyond this point, improving the ΔΣ order or the PWM resolution does not improve the SNR anymore. This can be solved by replacing the ΔΣ modulator with an integral noise shaping (INS) block [38], which shapes both the quantization noise and the PWM distortion by analytically integrating the PWM pulses.

2.5 Conclusion

Many partially or fully digital transmitter architectures can be found in literature. The most important types have been discussed and evaluated in this chapter. While other architectures can be more useful in other circumstances, it was found that for the applications targeted in this work, polar transmitters based on PWM are the most promising in terms of circuit complexity and PA efficiency. Therefore, the remainder of this work is dedicated to investigating this type of transmitters.

Two types of PWM-based transmitters are identified. Baseband PWM transmitters offer good resolution and are suited for both class-D and class-E PAs. However, they produce large harmonic distortion peaks close to the signal band. RF PWM transmitters produce a much cleaner spectrum but are more limited in resolution. Furthermore they produce waveforms which disrupt the soft switching property of class-E PAs and hence they are more suited for class-D amplification.

Both architectures will be thoroughly analyzed in Chap. 3. Analytical expressions for their output signals will be derived, and unwanted noise and distortion
terms will be identified and analyzed. The theory is supported and complemented with simulation results.

Achieving sufficient resolution at GHz-range carrier frequencies can be done using a continuous-time digital implementation. While this implies an increased design effort compared to clocked digital circuitry, it provides resolutions in the order of a few ps without requiring any sampling rates or reference frequencies higher than the carrier frequency.

For this reason, all transmitters designed in this work are implemented as continuous-time digital transmitters, which are not frequently found in literature. Chapter 4 gives an overview of the most important aspects of continuous-time digital circuit design and presents some key building blocks. Next, two prototypes will be presented in Chaps. 5 and 6.

References

60. Sokal NO, Sokal AD (1975), Class E—A new class of high-efficiency tuned single-ended switching power amplifiers. IEEE J Solid-State Circuits sc-10(3):168–176
63. Stauth JT, Sanders SR (2008a) A 2.4GHz, 20dBm class-D PA with single-bit digital polar modulation in 90nm CMOS. In: IEEE custom integrated circuits conference, pp 737–740
64. Stauth JT, Sanders SR (2008b) Pulse-density modulation for RF applications: The radio-frequency power amplifier (RF PA) as a power converter. In: IEEE power electronics specialist conference (PESC), pp 3563–3568
71. WLAN (2007) Wireless LAN medium access control (MAC) and physical layer (PHY) specifications. IEEE, std. 802.11-2007
Continuous-Time Digital Front-Ends for Multistandard Wireless Transmission
Nuyts, P.; Reynaert, P.; Dehaene, W.
2014, XXV, 309 p. 164 illus., 3 illus. in color., Hardcover
ISBN: 978-3-319-03924-4