1 Fundamentals of Discrete-Time RF Receivers

We start the book with the basics. In this chapter, we first present the motivation and fundamentals of discrete-time (DT) radio-frequency (RF) signal processing, and an overview of zero/low intermediate-frequency (IF) and superheterodyne receiver architectures. Then, different sampling schemes present in the state-of-the-art zero-IF DT receivers are studied using a simplified DT receiver. At the end, a $4\times$-sampling concept is introduced for use in DT high-IF receivers [1].

1.1 Why Discrete Time?

While the main motivations of CMOS scaling have been to reduce transistor cost and to improve digital performance, conventional RF/analog designs have not benefited significantly. A finer process node produces shorter digital gate delays while a lowered supply voltage and gate capacitance reduce power consumption. As shown in Fig. 1.1, from 180 to 28 nm CMOS, the $V_{DD}$ supply is reduced more than 30% while the MOS threshold voltage ($V_{th}$) has not changed considerably. Therefore, the precious available voltage headroom for RF/analog design is now reduced dramatically [2]. Considering also the reduced MOS intrinsic gain [2] and its saturation linearity in scaled CMOS [3], continuous-time (CT) RF/analog design is becoming generally more difficult. In this way, the power consumption and area of the traditional RF/analog designs are not directly process scalable.

On the other hand, the majority of cellular and wireless standard frequency bands are allocated from 400 MHz to 6 GHz, and have not significantly changed for many years. Meanwhile, the transistor cutoff frequency ($f_T$) has improved dramatically with scaling, as shown in Fig. 1.1. For example, the period from 1999 to 2011 has seen $f_T$ increasing from 20 GHz in 0.35 $\mu$m to more than 400 GHz in 28 nm process. This suggests that conventional CT techniques that were optimized for the older technology do not effectively use the ultrahigh speed of transistors of scaled CMOS to improve the performance of RF/analog designs.

In contrast, the newly introduced discrete-time (DT) RF/analog blocks (Fig. 1.2) avoid using complicated traditional analog components such as opamps, and most of signal processing and filtering are done using passive switched-capacitor circuits [4, 5]. Waveforms required for driving the switches are also generated using digital logic. To provide signal gain, DT techniques use inverter-based $g_m$-cells that avoid...
1 Fundamentals of Discrete-Time RF Receivers

transistor stacking and are always compatible with digital technology. As the technology scales, MOS switches become faster and tinier with lower parasitic capacitances. Digital waveform generator becomes also faster and more power efficient. Moreover, the metal capacitor density improves from one process node to the next, resulting in a reduced area. In addition, the inverter-based \( g_m \)-cell structure is fully scalable with improved \( g_m \) over its bias current. In this way, DT receivers directly benefit from scaling similar to that in digital circuits. References [6–15] are examples of DT process-scalable receivers.

1.2 Overview of Wireless Receiver Architectures

The pioneers of RFIC integration [16] have quickly realized the superiority of operating receivers at zero-IF (ZIF) and low-IF (LIF) rather than at high-IF (HIF): simpler architecture, and a much higher level of monolithic integration as a result of using low-frequency low-pass filters (LPF) for channel selection (see Fig. 1.3(a)). This was despite the many issues associated with ZIF/LIF receivers: time-variant dc offsets, sensitivity to 1/f (flicker) noise, large in-band local oscillator (LO) leakage and the second-order nonlinearity [17–22]. The LO leakage to the low-noise amplifier (LNA) input is amplified and then mixed with the LO again, creating a dc offset. This offset could be up to two to three orders of magnitude larger than the wanted signal at the mixer output [23]. By considering the LO leakage to antenna, it could be radiated and subsequently reflected from a moving subject back to the antenna. In this case, the dc offset is time varying and thus much harder to be canceled. In general, all ZIF/LIF issues were viewed rather as an inconvenience and handled through various calibrations. However, high-performance cellular ZIF/LIF receivers now require extensive
1.2 Overview of Wireless Receiver Architectures

Figure 1.3 Comparison of conventional receiver architectures: (a) zero-IF/low-IF and (b) superheterodyne.

calibration efforts. For example: an intensive IIP2 calibration needs to be concurrently run in the background with dc offset and harmonic rejection (HR) calibration [24, 25].

A superheterodyne architecture, shown in Fig. 1.3(b), pushes the IF frequency much higher such that the aforementioned problems are not a major concern anymore. Despite the obvious advantages, the superheterodyne radios were abandoned decades ago because it was extremely difficult to integrate a high quality (Q)-factor BPF for image rejection and channel selection in CMOS using continuous-time (CT) circuitry [16].

Furthermore, conventional multiband, multistandard cellular receivers (RXs) require many external duplexers, surface acoustic wave (SAW) filters, and switches, typically one per band, to attenuate out-of-band (OB) blockers before they reach the sensitive LNA’s input. In time-division duplexing (TDD) systems, external SAW filters can be eliminated if the RX chain could handle large interferers (e.g., 0 dBm at 20 MHz away from a GSM channel of interest [26]). On the other hand, for frequency-division duplexing (FDD) systems, the external SAW filters are responsible for not only the filtering of out-of-band blockers but also for duplexing, that is, separation of concurrent transmit (TX) and RX operations. To reduce the cost and size of the total system solution, in which the external antenna interface network is nowadays the largest contributor, the recent trend is to eliminate SAW filters and switches by using a highly linear wideband RX [27, 18–22]. As a consequence, the isolation of TX-to-RX, and the suppression of TX interferers are worsening, which all further increase the RX linearity requirements in FDD systems.
The resulting reduction in out-of-band filtering implies tough IIP2 requirements (e.g., 90 dBm [25, 22]) for ZIF and LIF receivers. The IIP2 performance of such receivers depends mainly on the second-order nonlinearity of LNA and RF mixer in the receiver chain, as shown in Fig. 1.3(a). Since the typical IIP2 of an RF mixer is between 50 and 70 dB [28], ZIF/LIF receivers require highly sophisticated calibration algorithms [29–33, 22, 34] to be frequently executed to account for variations in, $V_{DD}$, [19, 35, 36, 24, 37, 38], process corner [38], temperature [39], mixer transistor’s gate bias [35], RF blocker frequency [33, 36–38], LO frequency [36–38], LO power [38], and channel frequency [39]. Also, the IIP2 calibration time is rather very slow to find optimum setting for the mixer, and it needs to be run repeatedly due to environmental and operational changes [35].

Most of the filtering and amplification in a zero-IF receiver are done after the mixer, at low frequency. In CMOS implementations, the flicker noise of devices at low frequencies corrupts the wanted signal, leading to a higher noise figure (NF) of the receiver. In contrast, filtering and amplification in a superheterodyne are done normally at higher frequencies than the device’s flicker corner.

Superheterodyne or HIF architectures, on the other hand, can have a theoretically infinite IIP2. As shown in Fig. 1.3(b), the desired signal and modulated blocker at the RF input will be downconverted to a higher IF and dc, respectively; thus the modulated blocker can be completely filtered out by a band-pass filter (BPF) [40, 41]. For this reason, there is an increasing interest in uncalibrated high-IIP2 SAW-less superheterodyne RXs with integrated blocker-tolerant BPFs that are amenable to CMOS scaling.

1.3 Discrete-Time Concepts

1.3.1 Direct-Sampling Mixer

The basic idea of the current-mode direct-sampling mixer [42, 43] is illustrated in Fig. 1.4(a). The low-noise transconductance amplifier (LNTA) converts the received RF voltage $v_{RF}$ into $i_{RF}$ in current domain through the transconductance gain $g_m$. The current $i_{RF}$ gets switched by the half-cycle of the local oscillator (LO) and

![Figure 1.4 Temporal MA operation at RF rate: (a) single-ended and (b) pseudo-differential configurations.](image-url)
integrated into the sampling capacitor $C_s$. Since it is difficult to switch the current at RF rate, it could be merely redirected to an identical sampler that is operating on the opposite half-cycle of the LO clock, as shown in Fig. 1.4(b) for a pseudo-differential configuration.

If the LO oscillating at $f_0$ frequency is synchronous and in phase with the sinusoidal RF waveform, the voltage gain of a single RF half-cycle is

$$G_{v,RF} = \frac{1}{\pi} \cdot \frac{1}{f_0} \cdot \frac{g_m}{C_s},$$

(1.1)

and the accumulated charge on the sampling capacitor is

$$G_{q,RF} = \frac{1}{\pi} \cdot \frac{1}{f_0} \cdot g_m.$$  

(1.2)

In (1.1) and (1.2), the $\frac{1}{\pi}$ factor is contributed by the half-cycle sinusoidal integration. As an example, if $g_m = 30 \text{ mS}$, $C_s = 15.925 \text{ pF}$, and $f_0 = 2.4 \text{ GHz}$, then $G_{v,RF} = 0.25$.

### 1.3.2 Temporal Moving Average

Continuously accumulating the charge as shown in Fig. 1.4 is not very practical if it cannot be read out. In addition, a mechanism to prevent the charge overflow is needed. Both of these operations are accomplished by fixing the integration window length followed by a charge readout phase that will also discharge the sampling capacitor such that the next period of integration would start from the same zero condition. The RF sampling and readout operations are cyclically rotated on both $C_s$ capacitors as shown in Fig. 1.5. When LO$_A$ rectifies $N$ RF cycles that are being integrated on the first sampling capacitor, LO$_B$ is off and the second sampling capacitor charge is being read out. On the following $N$ RF cycles, the operation is reversed. This way, the charge integration and readout occur at the same time and no RF cycles are missed.

The sampling capacitor integrates the half-rectified RF current over $N$ cycles. The charge accumulated on the sampling capacitor and the resulting voltage ($V = Q/C_s$) increases with the integration window, thus giving rise to a discrete signal processing gain of $N$.

![Figure 1.5 Temporal MA operation at RF rate with cyclic charge readout.](image-url)
The temporal integration of $N$ half-rectified RF samples performs a (FIR) operation with $N$ all-one coefficients, also known as moving average (MA), according to the equation:

$$w_i = \sum_{l=0}^{N-1} u_{i-l},$$

(1.3)

where $u_i$ is the $i$th RF sample of the input charge sample and $w_i$ is the accumulated charge. Since the charge accumulation is done on the same capacitor, this formula could also be used in the voltage domain. Its frequency response is a sinc function and is shown in Fig. 1.6 for $N = 8$ (solid line) and $N = 7, 9$ (dotted lines) with sampling rate $f_0 = 2.4$ GHz. It should be noted that this filtering is done on the same capacitor in time domain, resulting in a most faithful reproduction of the transfer function.

Due to the fact that the MA output is being read out at the lower rate of $N$ RF clock cycles, there is an additional aliasing with foldover frequency at $f_0/2N$ and located halfway to the first notch. Consequently, the frequency response of $N = 7$ with decimation of 7 exhibits less aliasing and features wider notches than $N = 8$ or $N = 9$ with decimation of 8 or 9, respectively.

It should be emphasized that the voltage $G_v$ and charge $G_q$ signal processing gains of the temporal moving average (TMA) (followed by decimation) are merely due to the sampling time interval expansion of this discrete-time system (the sampling rate of the input is at the RF frequency): $G_{v, tma} = G_{q, tma} = N$.

In the following analysis, the RF half-cycle integration voltage gain of $\frac{g_m \pi C_s f_0}{2}$ is tracked separately. Since this gain depends on the absolute physical parameters of normally low tolerance ($g_m$ value of the preceding LNTA stage and the total integrating capacitance of the sampling mixer), it is advantageous to keep it decoupled from the discrete signal processing gain of the multi-tap direct-sampling mixer (MTDSM).
1.3.3 High-Rate IIR Filtering

Figure 1.5 is now modified to include recursive operation that gives rise to the IIR filtering capability, which is generally considered stronger than that of FIR.

A “history” sampling capacitor $C_H$ is added in Fig. 1.7. The integration operation is continually performed on the history capacitor $C_H = a_1 C_s$ and one of the two rotating “charge-and-readout” capacitors $C_R = (1 - a_1) C_s$ such that the total RF integrating capacitance, as seen by the LNTA, is always $C_H + C_R = C_s$. When one of the $C_R$ capacitors is being used for readout, the other is being used for RF integration.

The IIR filtering capability comes into play in the following way: The RF current is being integrated over $N_{RF}$ cycles, as described before. This time, the charge is being shared on both $C_H$ and $C_R$ capacitors proportionately to their capacitance values. At the end of the accumulation cycle, the active $C_R$ capacitor, that stores $(1 - a_1)$ of the total charge, stops further accumulating in preparation for charge readout. The other rotating capacitor joins the $C_H$ capacitor in the RF sampling process and, at the same time, obtains $1 - a_1$ of the total remaining charge in the “history” capacitor, provided it has no initial charge at the time of commutation. Thus, the system retains $a_1$ portion of the total system charge of the previous cycle.

If the input charge accumulated over the most-recent $N$ RF samples is $w_j$ then the charge $s_j$ stored in the system at sampling time $j$, where $i = N \cdot j$, (as stated earlier, $i$ is the RF cycle index) could be described as a single-pole recursive IIR equation:

$$s_j = a_1 s_{j-1} + w_j, \quad (1.4)$$
$$x_j = (1 - a_1) s_{j-1}, \quad (1.5)$$
$$a_1 = \frac{C_H}{C_H + C_R}. \quad (1.6)$$

The output charge $x_j$ is $(1 - a_1)$ of the system charge in the most recent cycle. This discrete-time IIR filter operates at $f_0/N$ sampling rate and introduces a single pole with the frequency attenuation of 20 dB/dec. The equivalent pole location in the continuous-time domain for $f_c \ll f_0/N$ is

$$f_c = \frac{1}{2\pi N} \frac{f_0}{(1 - a_1)} = \frac{1}{2\pi N} \frac{f_0}{\frac{C_R}{C_H + C_R}}. \quad (1.7)$$
8 1 Fundamentals of Discrete-Time RF Receivers

Since there is no sampling time expansion for the IIR operation, the discrete signal processing charge gain is one. In other words, due to the charge conservation principle, the input charge per sample interval is on average the same as the output charge. For the voltage gain, however, there is an impedance transformation of \( C_{\text{input}} = C_s \) and \( C_{\text{output}} = (1 - a_1)C_s \), thus resulting in a gain.

\[
G_{q,iir1} = 1, \tag{1.8}
\]

\[
G_{v,iir1} = \frac{1}{1 - a_1} = \frac{C_H + C_R}{C_R}. \tag{1.9}
\]

As an example, the IIR filtering with a single coefficient of \( a_1 = 0.9686 \), placing the pole at \( f_{c1} = 1.5 \text{ MHz} \) (\( C_R = 0.5 \text{ pF}, C_H = 15.425 \text{ pF} \)) is performed at \( f_0/N = 2.4 \text{ GHz} / 8 = 300 \text{ MHz} \) sampling rate, and it follows the FIR MA = 8 filtering of the input at \( f_0 \) RF sampling rate. The voltage gain of the high-rate IIR filter is 31.85 (30.06 dB).

1.3.4 Additional Spatial MA Filtering Zeros

For practical reasons, it is difficult to read out the \( x_j \) output charge of Fig. 1.7 at \( f_0/N = 300 \text{ MHz} \) rate. The output charge readout time is extended \( M = 4 \) times by adding redundancy of four to each of the two original \( C_R \) capacitors as shown in Fig. 1.8. The input charge is cyclically integrated within the group of four \( C_R \) capacitors. Adding the redundant capacitors gives rise to an additional antialiasing filtering just before the second decimation of \( M \). This could also be considered as equivalent to adding additional \( M - 1 \) zeros to the IIR transfer function in (1.4).

![Figure 1.8](image-url)
After the first bank of four capacitors gets charged \((S_{A1} - S_{A4} \text{ in Fig. 1.8})\), the second bank \((S_{B1} - S_{B4})\) is in the process of being charged and the charge on the first bank of capacitors is summed and read out \((R_1)\). Physically connecting together the four capacitors performs an FIR filtering described as the spatial moving average of \(M = 4\):

\[
y_k = \sum_{l=0}^{M-1} x_{k-l},
\]

(1.10)

where \(y_k\) is the output charge and sampling time index \(j = M \cdot k\). \(R_A\) and \(R_B\) in Fig. 1.8 are the readout/reset cycles during which the output charge on the four non-sampling capacitors is transferred out, and the remnant charge is reset before the capacitors are put back into the sampling operation. It should be noted that after the reset phase, but before the sampling phase, the capacitors are unobtrusively precharged \([44]\) in order to implement a dc-offset cancelation or to accomplish a feedback summation for the \(\Sigma\Delta\) loop operation.

Since the charge of four capacitors is added, there is a charge gain of \(M = 4\) and a voltage gain of 1. Again, as explained before, the charge gain is due to the sampling interval expansion: \(G_{q,sma} = M\) and \(G_{v,sma} = 1\).

Figure 1.9 shows the frequency response of the temporal moving average with a decimation of 8 \((G_v = 18.06 \, \text{dB})\), the IIR filter operating at RF/8 rate \((G_v = 30.06 \, \text{dB})\), and the spatial moving average filter operating at RF/32 rate \((G_v = 0 \, \text{dB})\) with a decimation of 4. The solid line is the composite transfer function with the dc gain of \(G_v = 48.12 \, \text{dB}\). The first decimation of \(N = 8\) reveals itself as aliasing. It should be noted that it is possible to avoid aliasing of a very strong interferer into the critical IF band by simply changing the decimation ratio \(N\). This brings out the
advantages of integrating RF/analog with digital circuitry by opening new avenues of novel signal processing solutions not possible before.

1.3.5 Lower-Rate IIR Filtering

The voltage stored on the rotating capacitors cannot be readily presented to the MTDSM block output without an active buffer that would isolate the high impedance of the mixer from the required low driving impedance of the output. Figure 1.10 shows the mechanism to realize the second, lower-rate, IIR filtering through passive charge sharing. The active element, the operational amplifier, does not actually take part of the IIR filtering process. It is merely used to sense voltage of the buffer feedback capacitor $C_B$ and present it to the output with a low driving impedance. Figure 1.10 additionally suggests the possibility of differentially combining, through the operational amplifier, the opposite (180° apart) processing path.

The charge $y_k$ accumulated on the $M = 4$ rotating capacitors is being shared during the dumping phase with the buffer feedback capacitor $C_B$. At the end of the dumping phase, the $M \cdot C_R$ capacitors get disconnected from the second IIR filter and their charge reset before they could be reengaged in the MTDSM operation of Fig. 1.8. This charge loss mechanism gives rise to IIR filtering. If the input charge is $y_k$, then the charge $z_k$ stored in the buffer capacitor $C_B$ at sampling time $k$ is

$$z_k = a_2(z_{k-1} + y_k) = a_2z_{k-1} + a_2y_k.$$  \hspace{1cm} (1.11)

$$a_2 = \frac{C_B}{C_B + MC_R}. \hspace{1cm} (1.12)$$

Equation (1.11) describes a single-pole IIR filter with coefficient $a_2$ and input $y_k$ scaled by $a_2$, where $a_2$ corresponds to the storage-to-total capacitance ratio $\frac{C_B}{C_B + MC_R}$. Conversely, due to the linearity property, it could also be thought of as an IIR filter with input $y_k$ and output scaled by $a_2$.

This discrete-time IIR filter operates at $f_0/\text{NM}$ sampling rate and introduces a single pole with the frequency transfer function attenuation of 20 dB/dec. The equivalent pole location in the continuous-time domain for $f_{c2} \ll f_0/(\text{NM})$ is

$$f_{c2} = \frac{1}{2\pi} \frac{f_0}{\text{NM}} \cdot (1 - a_2) = \frac{1}{2\pi} \frac{f_0}{\text{NM}} \cdot \frac{MC_R}{C_B + MC_R}. \hspace{1cm} (1.13)$$