

Chapter 2

Incremental-Charge-Based Operation

2.1 Introduction

To ease the understanding of the underlying concept of charge-based operation, let's take a mechanical system as an example. Imagine a watermill, which has to be controlled in a precise and timely manner following a wanted input command (Fig. 2.1).

Controlling the rotating wheel and hence the amount of power transferred to the structure is done in this case by controlling the amount of water that flows through the water pipes.¹ A larger water flow implies more energy being transferred to the watermill, and vice-versa.

Such control mechanism could be implemented in many different ways. For instance, let's consider a voltage controlled water pump, connected to the structure using pipes and taking water directly from the water main, as depicted in Fig. 2.2.

To assure the wanted functionality, first this specially designed pump should be strong enough to deliver the required flow to drive the rotating wheel at full power. Second, it should be able to resolve significantly small flow levels, so that the amount of water transferred to the wheel can be precisely controlled, and third, the output water flow should be—desirably—a linear function of the input control parameter, meaning that if $V_{CONTROL}$ is doubled, the water flow should also be doubled. At last, when transitioning from one flow level to another, the water pump should respond timely and consistently.

As one may expect, all of this requirements have a direct impact on how well the watermill can be operated, and therefore how close the output power resembles the

¹For the sake of simplicity, a linear relationship between output power and the water flow is considered.

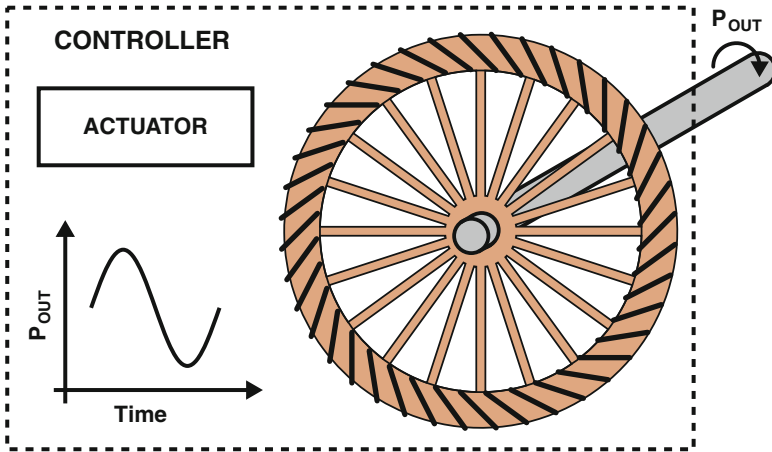


Fig. 2.1 Watermill mechanical system. The amount of power transferred to rotating wheel should follow a wanted input command

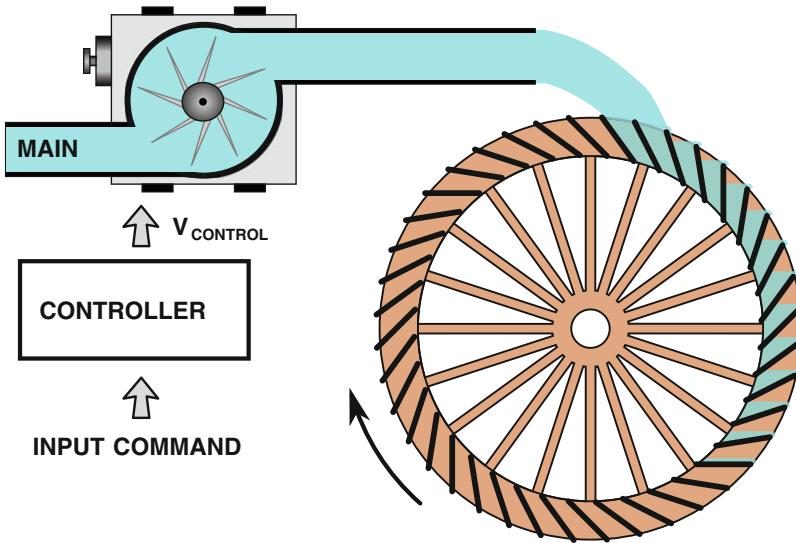


Fig. 2.2 Example control mechanism. A voltage-regulated water pump is used to control the amount of water pushed through the rotating wheel

wanted input command. Providing altogether a high power capability with a good degree of linearity and small enough intermediate steps is a tough challenge in itself, made even tougher when high power efficiency is required.

Since power consumption in typical cases can be dominated by fixed contributors that do not scale with the wanted output power, working at a fraction of the full capacity can lead to a considerable degradation of the system power efficiency

[Eq. (2.1)]. In our analogy, it means that most of the time the mechanical system will be running at poor efficiency levels, degraded by the large power consumption required to operate the strong water pump, compared to the reduced output power delivered.

$$Efficiency = \frac{P_{OUT}}{P_{PUMP}} = \frac{P_{OUT}}{\underbrace{P_{FIXED}}_{\text{DOMINANT}} + P_{VAR}} \quad (2.1)$$

Now let's assume the situation shown in Fig. 2.3. Instead of using a power hungry electrical pump to drive the watermill, the water pipes leading to the rotating wheel are now connected to a reservoir, whose water level can be directly manipulated with a featured controller device. Though the power transfer mechanism remains unaltered, now the control of the rotating wheel is implemented through the modification of the water level inside the reservoir. If the water level is increased, more water is pushed through the pipes—increasing the output power. Similarly, a decrease in water level incurs less power being transferred.

A top-level description of the system operation is depicted in Fig. 2.4. Based on the instantaneous input command $P_{OUT}[k]$, the water level ($W_{RES}[k]$) corresponding to the required flow is first calculated. The amount of water that should be added

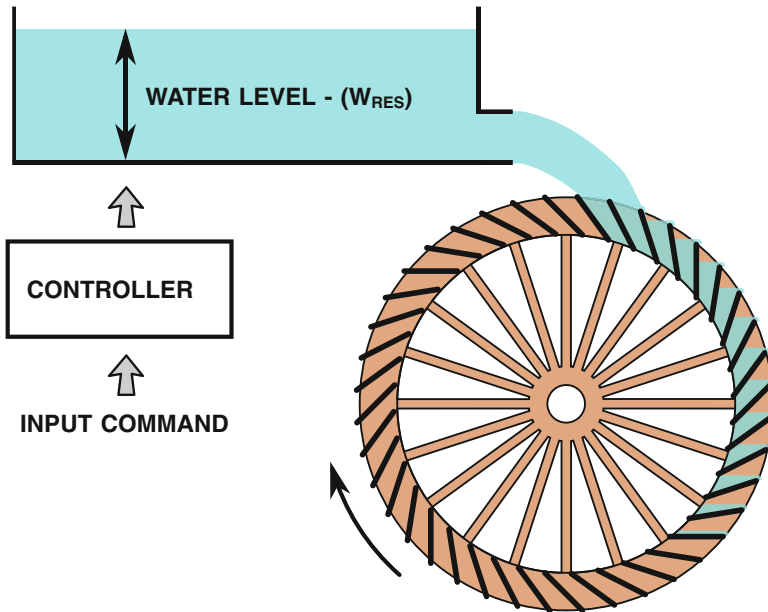


Fig. 2.3 Alternative control mechanism. Instead of using a water pump, the water flow is controlled by changing the water level in a large reservoir

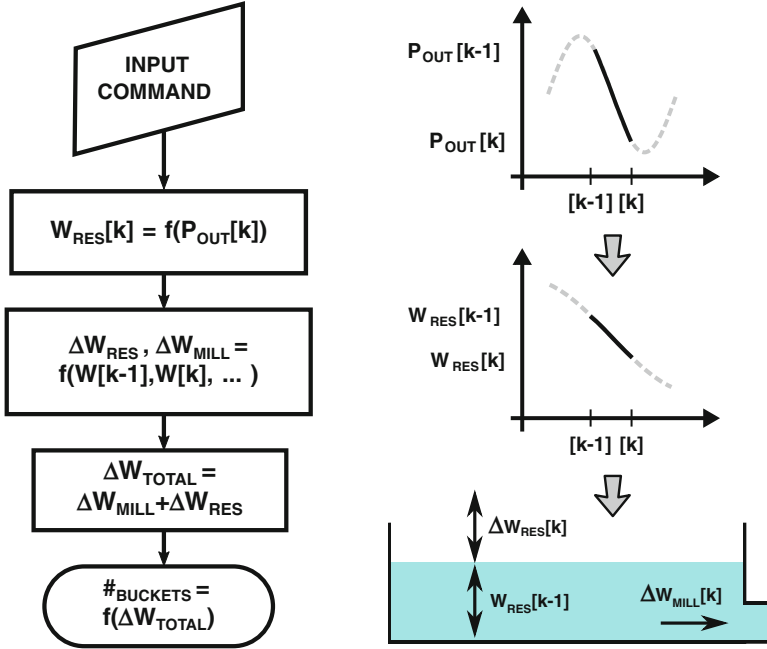


Fig. 2.4 Top-level description of the reservoir-based operation. Following the input command, the control system determines how much water should be added or subtracted from the reservoir

to (or subtracted from) the reservoir ($\Delta W_{TOTAL}[k]$) will depend on two parameters in this case: How much water will be subtracted while driving the rotating wheel (ΔW_{MILL}) during the given period of time, and the amount of water required to bring the reservoir level—and hence the water flow—to its next value ($\Delta W_{RES} = W_{RES}[k] - W_{RES}[k-1]$) as defined by the input command. Note that in this particular operation mode the water level pre-existing in the reservoir is always taken into account, and the water taken from the main corresponds to the required increment only.

The same reasoning can be applied in the electrical domain, more specifically into an alternative transmitter architecture. What in this example represents a rotating wheel, in the transmitter application it could be exchanged for an output (RF) load. The physical quantity being manipulated would be electrical charge—instead of water, in which case the role of a water reservoir would be perfectly matched by a charge accumulator, or capacitor in other words. Applying the already mentioned incremental operation, this transmitter architecture could be therefore denominated an incremental-charge-based implementation, whose exploration and conceptual validation were first presented in [Par15a].

The benefits of such modifications in the way the system is operated are clarified once the implementation of the featured water level controller is presented: As depicted in Fig. 2.5, imagine that instead of providing the reservoir with discrete

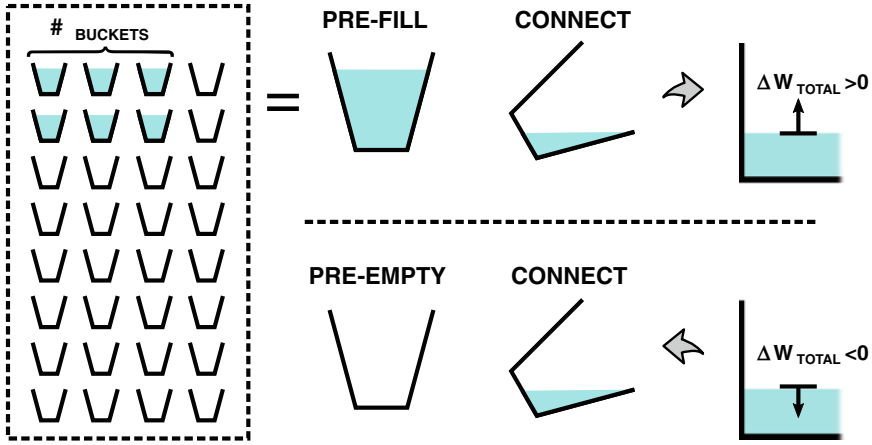


Fig. 2.5 Example structure used to control the reservoir level. According to required ΔW_{TOTAL} , the corresponding number of buckets ($\#_{BUCKETS}$) are pre-filled or pre-empted before being connected to the reservoir, therefore increasing or decreasing the total amount stored

quantities of water using a water pump, a sufficiently large set of equally sized buckets are made available, which can be independently selected and pre-filled (or pre-empted) before being connected to the actual reservoir.

If the water level is to be increased, the “configurable”-size bucket is first filled with water, which is later transferred to the reservoir. Similarly, when the level is to be decreased, the buckets are first emptied before connecting. Whenever a different input command is received, the corresponding bucket size is determined, as well as whether it should be pre-filled or pre-empted, so that the required amount of water is provided. Notably, using buckets to change the water level in the reservoir can have a very interesting impact in the overall system operation and performance:

First, the minimum change in the water level that can be produced in the reservoir—and hence in the water flow—is determined by and scales with the smallest bucket available in the water manipulator. If the smallest bucket size is divided by 2, the minimum increase/decrease producible in the water level is also halved. In fact, the resulting change in water level is also inversely dependent on the reservoir size itself, meaning that for a fixed minimum bucket size, the increment in water level can also be halved by doubling the size of the reservoir. Therefore, rather than an absolute dependence on the minimum bucket size, the water flow resolution is proportional to the ratio between the smallest bucket and the reservoir size, as depicted in Fig. 2.6 and demonstrated in Eq. (2.2).

$$\Delta W_{MIN} \propto \frac{\text{Minimum Bucket Size}}{\text{Reservoir Size}} \Rightarrow \left| S_{n_{QUANT}}^2 \right| \quad (2.2)$$

Compared to the pump-based implementation, a ratio-dependent minimum step simplifies immensely on achieving the required water-flow resolution. In the

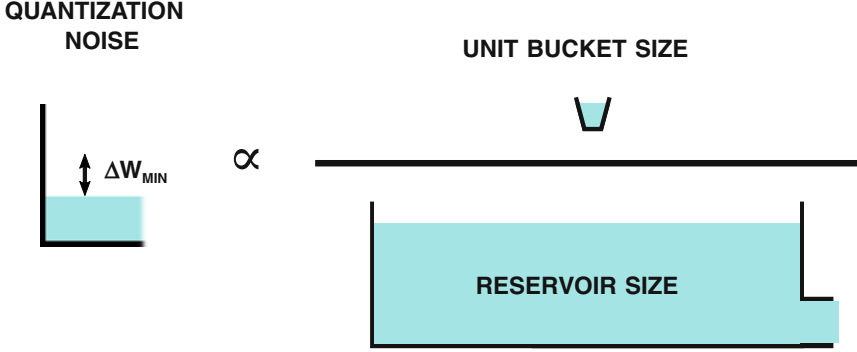


Fig. 2.6 Minimum achievable water level in(de)crease. In this alternative implementation, the water level resolution (and hence quantization noise) can be improved by either increasing the reservoir size, or reducing the minimum bucket size

transmitter domain, it represents the ability to reduce the quantization noise floor ($S_{n_{QUANT}}^2$) below the stringent out-of-band noise requirements of advanced wireless systems by simply reducing the ratio between two capacitors. This feature is further explored in Sect. 3.2.3.2.

Second, feeding the watermill through a reservoir provides intrinsic filtering capabilities to the structure. Whilst in the pump-based solution variations of multiple sources (e.g. changes in the pump supply voltage, water main pressure, etc.) will have a direct impact on the output water flow, the application of a water reservoir provides an intrinsic buffering effect where all sorts of disturbances and fluctuations (“noise”) are naturally averaged and therefore attenuated by the structure. In fact, the larger the reservoir the more damping is provided, further attenuating fluctuations in the observed water flow. In the context of transmitters for advanced wireless communication systems, this characteristic is a key enabler to achieve a sufficiently low out-of-band noise emission, filtering out from supply-coupled to quantization and thermal noise. A more complete explanation of the intrinsic noise capability of the charge-based structure is given in Sect. 2.2.1.

Third, different from typical architectures the charge-based operation can also provide utmost efficiency when operating at lower power levels. First of all, there is no “fixed” consumption, and water is only consumed when the reservoir level should be increased. Second, since the amount of water pre-existing in the reservoir is always taken into account, only the required increment (ΔW_{TOTAL}) is taken from the main. The same reasoning can be applied to the incremental charge-based TX architecture. The amount of charge taken from the supply—and hence the power consumption—scales with the signal amplitude, inherently improving power efficiency at backoff conditions.

Finally, one may argue that driving a large water reservoir (or charge accumulator) can also impact the overall system power efficiency. However, the benefits

provided by the intrinsic noise filtering capability allows the reduction of both system complexity and power consumption related to signal noise filtering. In other words, in the charge-based operation the trade-off between noise performance and power consumption is significantly leveraged, as further clarified in the following sections.

2.2 Incremental-Charge-Based Operation

Suppose that instead of driving a watermill, now an electrical load Z is to be driven following a given input command (Fig. 2.7). Again, the power transfer from the supply to the output load has to be done in a precisely-controlled and timely manner, demanding sufficient linearity and bandwidth.

In a conventional voltage-output driving stage, more bandwidth is necessarily provided with a corresponding increase in transconductance. Circuit linearity, on the other hand, is typically leveraged by increasing the voltage headrooms and supply voltages whenever possible. In both cases, when stringent bandwidth and linearity requirements apply, the resulting DC power consumption implied by large bias currents and/or supply voltages can significantly impact the overall system efficiency.

Looking at the electrical nature of Z , a clear trend can be noticed though. In many cases, the output load to most blocks in any particular signal path is purely capacitive, given by layout parasitics and gate capacitances of the succeeding block's input transistors. As in the watermill example, these capacitances function as reservoirs whose charge content could be manipulated incrementally. Therefore, instead of driving the output load with a conventional voltage-output Digital-to-Analog Converter (DAC), the output signal in this architecture would be constructed by adding or subtracting charge from the output load as shown in Fig. 2.8, taking into account the previous amount of charge existing at all times.

The operating principles are very simple. Assuming for a first example a purely capacitive load (C_{LOAD}), it is possible to calculate how much charge should be moved to the output (per sampling period) based on the input command, as shown in

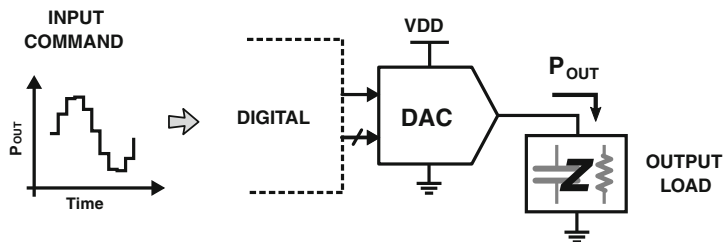


Fig. 2.7 Simplified representation of a typical system where a load Z has to be driven following a wanted (digital) input command

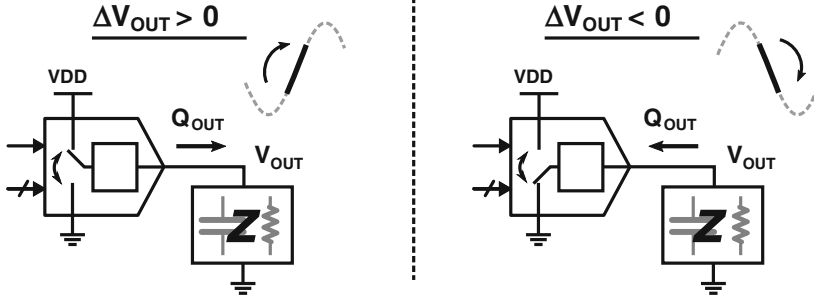


Fig. 2.8 In the incremental charge-based architecture the output voltage is changed by adding and subtracting charge from the output load

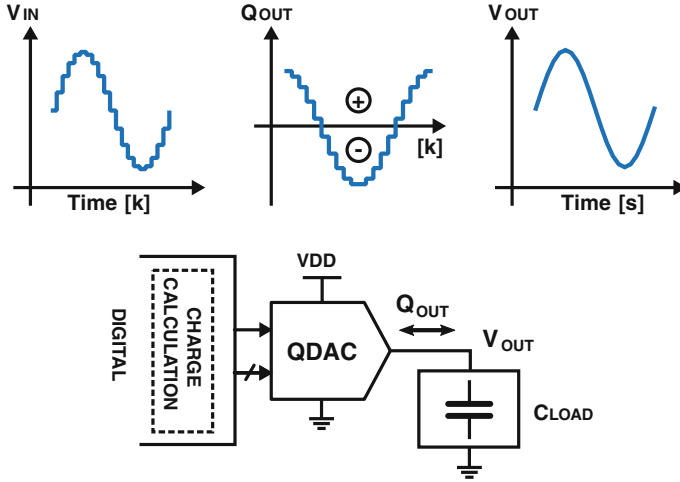


Fig. 2.9 Charge-based operation. A charge calculation block defines how much charge should be moved to/from the output load

Eq. (2.3). If the total amount of charge stored in C_{LOAD} is to be increased, charge is pulled from the supply, otherwise the excess charge is drained into the ground. These calculations are performed by a calculation block that first translates the digital input command into charge values (Fig. 2.9), and then converts it into the specific DAC input quantity (current, conductance, capacitance, etc.). Due to the simple nature of the required calculations, the power consumption and area overhead due to the additional calculation block can be negligible.

$$Q_{OUT}[k] = \underbrace{(V_{IN}[k] - V_{IN}[k-1])}_{\Delta V_{IN}} \cdot C_{LOAD} \quad (2.3)$$

The required charge Q_{OUT} can be delivered in different ways through what is called a charge-based DAC—or *QDAC*, either using discrete packets of charge conveyed to the output at fractions of the sampling period, or in continuous-time by constantly charging and discharging the output load throughout time.

Now imagine that this charge-based DAC is capable of providing the required bandwidth and linearity without requiring any biasing. Without a fixed bias current, the incremental operation assures that the power consumed to drive the output load is directly proportional to the output signal's amplitude. In other words, the DC power consumption of the charge-based architecture scales with the useful output power, leading to an improved power efficiency even at backoff conditions. In fact, for the simple example shown in Fig. 2.9, if the power spent with the QDAC operation (say due to digital switching) is made negligible, the charge-based DAC current consumption for a given output capacitance C can achieve the theoretical limit of fCV coulombs per second (or ampères)—where f and V are the output signal's frequency and amplitude, respectively.

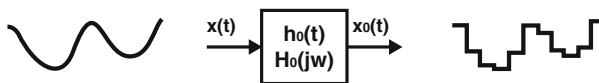
The charge-based operation is not limited to pure capacitive loads, however. As discussed in Chap. 4, it is also possible to apply the same reasoning when driving a resistive output load and still observe all the benefits of charge-based operation, including improved efficiency and intrinsic noise filtering capabilities, as discussed below.

For the sake of a better understanding, the charge-based architecture and its intrinsic benefits are discussed here using a generic “black-box” charge-based DAC implementation. Further details about the two different implementations studied in this book are given in Chaps. 3 and 4.

2.2.1 Noise and Alias Performance

2.2.1.1 Alias Attenuation

In typical Digital-to-Analog Converters (DAC), the output voltage (or current) is a simple translation of the digital input code into a fraction of the converter's output full-scale. In a conventional implementation, the output signal remains fixed when the digital input word is not changing, corresponding to a zero-order hold (ZOH) [Opp97].



$$H_0(j\omega) = e^{-j\omega T/2} \left[\frac{2 \sin(\omega T/2)}{\omega} \right]$$

$$\begin{aligned}
&= T e^{-j\omega T/2} \left[\frac{\sin(\pi \omega T/2\pi)}{\pi \omega T/2\pi} \right] \\
&= T e^{-j\pi(\omega/\omega_0)} \operatorname{sinc}\left(\frac{\omega}{\omega_S}\right)
\end{aligned} \tag{2.4}$$

where $\omega_S = \frac{2\pi}{T}$ and $\operatorname{sinc}(x) = \frac{\sin(\pi x)}{\pi x}$.

As demonstrated in Eq. (2.4), the output spectrum of a converter applying zero-order hold is shaped by a $\operatorname{sinc}(x)$ function, notching at every multiple of the sampling frequency. The sampling aliases, centered around $n \cdot \omega_S$ (where $n = [1, 2, 3 \dots]$), are also shaped by the same $\operatorname{sinc}(x)$ function, as shown in Fig. 2.10.

More involved converter architectures apply different interpolation schemes in order to further attenuate the sampling aliases. In most cases, either the input data is oversampled and interpolated using digital filters [Yon07, Yij03b], or multiple copies of the same converter are interleaved and operated at different phases of the sampling clock [Chi10]. While the former solution increases power consumption with digital operation at high clock frequencies, the latter can have a significant impact in area consumption.

Consider the case of a charge-based architecture where the charge transfer between supply and output load does not happen instantly, but rather uniformly distributed along the entire sampling period. If a constant amount of charge is delivered to a purely capacitive load, the output voltage increases linearly over time, corresponding in this case to an inherent first-order interpolation.

Different from a zero-order hold system, the application of linear interpolation has a beneficial impact on the reconstructed output spectrum. Demonstrated in Eq. (2.5), instead of a $\operatorname{sinc}(x)$, the aliases in this case are shaped by a $\operatorname{sinc}(x)^2$ transfer function, significantly reducing the sampling aliases as shown in Fig. 2.11.

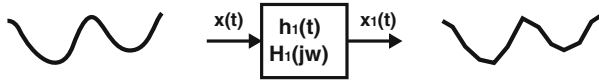


Fig. 2.10 Using zero-order hold the reconstruction spectrum is shaped by a $\operatorname{sinc}(x)$ function

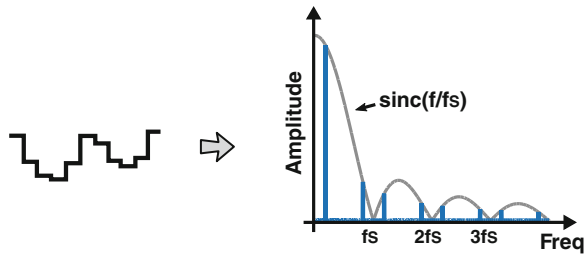


Fig. 2.11 $\text{sinc}(x)^2$ alias attenuation

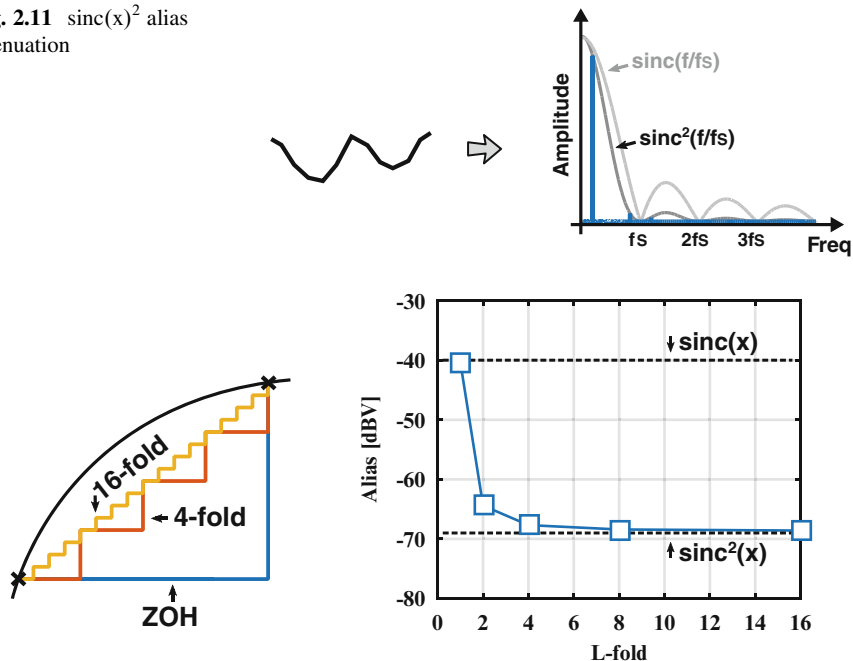
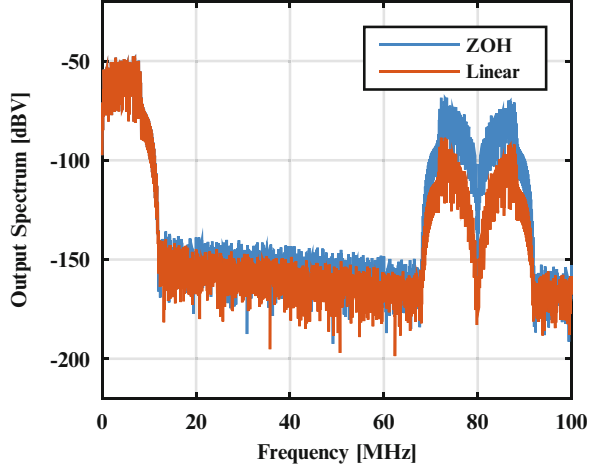


Fig. 2.12 L-fold linear interpolation for different values of L

$$\begin{aligned}
 H_1(j\omega) &= \frac{1}{T} \left[\frac{\sin(\omega T/2)}{\omega/2} \right]^2 \\
 &= \frac{1}{T} \left[\frac{T \sin(\pi \omega T/2\pi)}{\pi \omega T/2\pi} \right]^2 \\
 &= T \text{sinc}^2 \left(\frac{\omega}{\omega_s} \right)
 \end{aligned} \tag{2.5}$$

Even when the charge conveyed to the output node is not uniformly delivered over time, but rather in multiple sub-steps at a fraction of the sampling period (as in a L-fold linear interpolation [Yij03b]), yet the output spectrum shaping approaches the $\text{sinc}(x)^2$ transfer function for values of L above 2, as demonstrated in Fig. 2.12 for 20 MHz output single-tone sampled at 500 MS/s. It should be noted that transmitting charge at a higher speed in this case does not imply oversampling the input signal or interleaved operation. The calculation engine and the corresponding interface toward the DAC input are all driven at the same speed as the digital input data (F_S).

Fig. 2.13 Output spectrum of a wideband output signal sampled at 80 MS/s



At least 30 dB of additional alias suppression can be achieved with the proposed charge-based architecture for a 20 MHz bandwidth output signal sampled at 80 MS/s, as exemplified in Fig. 2.13.

2.2.1.2 Intrinsic Noise Filtering

Besides the inevitable quantization noise, multiple noise sources also impact the output signal in a typical DAC implementation. Depending on the actual implementation, thermal and flicker noise from the internal DAC components, or even noise coupled to the supply and/or reference voltage can also contribute to a Signal-to-Noise Ratio (SNR) degradation. In cases where the noise spectral density at specific parts of the output spectrum should be kept at a minimum, or when stringent spectral masks should be met, typically a reconstruction filter is applied in order to attenuate both output noise and sampling aliases after digital-to-analog conversion (Fig. 2.14).

An important drawback of using reconstruction filters is the large area typically taken by these blocks. In cases where strict spectral requirements apply, it is not rare to see a large cut of the chip footprint being taken by the reconstruction filter [Oli12]. Another downside of using these bulky filters in a transmitter's signal path, for example, is the bandwidth limitation imposed to higher frequency components intentionally added to the baseband signal in order to compensate for circuit's nonlinearities, reducing the effectiveness of digital pre-distortion (DPD). Anyhow, to any sort of application a DAC implementation that can provide intrinsic noise filtering capabilities along with the already discussed alias attenuation would be very much welcome.

One obvious implementation of the charge-based DAC would comprise a current DAC, where the current delivered to the output load (I_{OUT}) would be readjusted

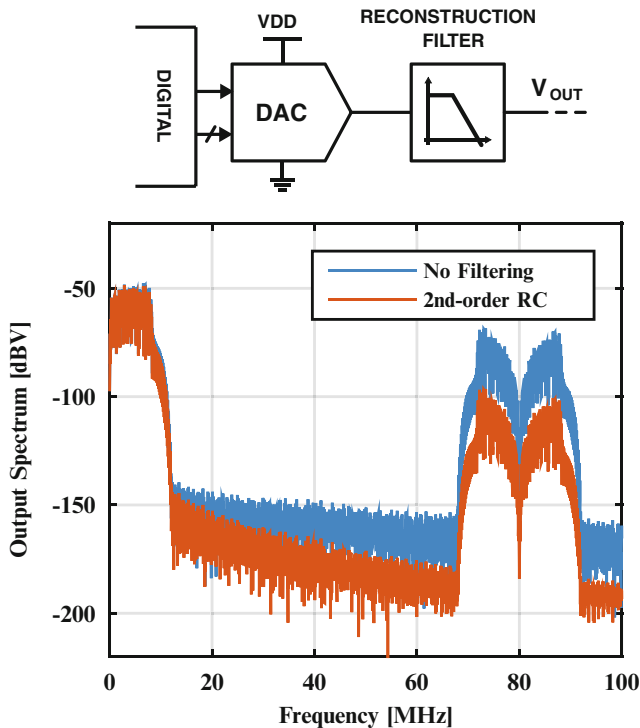
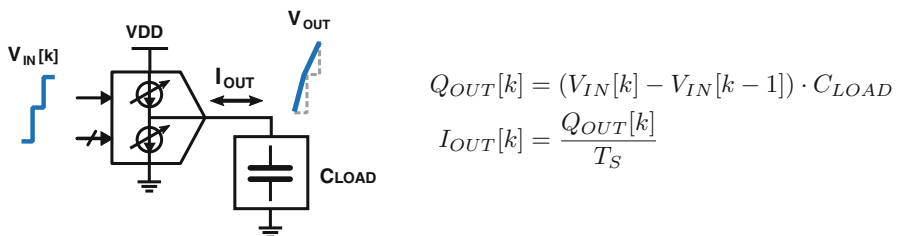


Fig. 2.14 Noise and Alias filtering achieved with a second-order reconstruction RC filter with a cutoff frequency of 15 MHz

every sampling period in order to provide the wanted $Q/\Delta T_S$, as defined by the charge calculation block. In this way, it is clear that the required $Q_{OUT}[k]$ would be uniformly distributed over time, thus leading to the aforementioned $\text{sinc}(x)^2$ alias attenuation.



The same functionality can also be achieved with an alternative implementation where the QDAC is supplied with a fixed voltage source, and the same I_{OUT} is provided with a variable conductance that is changed over time according to the digital input command (Fig. 2.15). In the ideal case where the output voltage is an

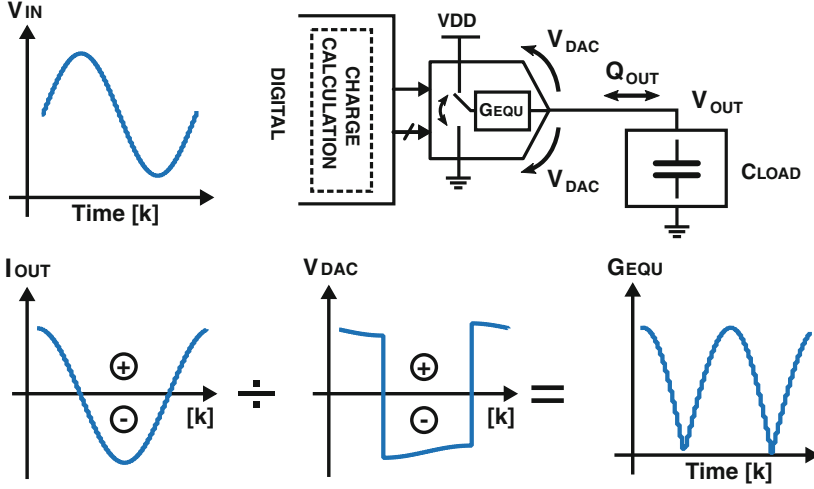


Fig. 2.15 The voltage drop across the QDAC terminals in combination with the output current can be translated into an equivalent DAC conductance

unaltered analog copy of V_{IN} , the equivalent conductance G_{EQU} can be derived as the ratio between the wanted I_{OUT} and the voltage drop across the DAC terminals (V_{DAC}) at time k [Eq. (2.6)].

$$V_{DAC}[k] = \begin{cases} V_{DD} - V_{IN}[k-1], & \text{if } Q_{OUT}[k] > 0 \\ V_{ND} - V_{IN}[k-1], & \text{if } Q_{OUT}[k] < 0 \end{cases} \quad (2.6)$$

$$G_{EQU}[k] = \frac{I_{OUT}[k]}{V_{DAC}[k]}$$

Notably, the combination between the equivalent DAC conductance (G_{EQU}) and the capacitive component of the output load creates an intrinsic single order low-pass filtering effect. What is very particular however, is the fact that the “resistive” component of this RC filter is signal dependent. When a large difference between two consecutive digital input samples is observed (large $\Delta V_{IN}/\Delta T_S$), the increased Q_{OUT} required leads to an also large equivalent G_{EQU} and hence small RC time constant. On the other hand, when a small amount of charge is required (small $\Delta V_{IN}/\Delta T_S$), the instantaneous RC constant is increased (Fig. 2.16). Therefore, from the signal point-of-view no attenuation is implied by the time-varying RC filter, since its resistive component is automatically adjusted according to the wanted output swing.

On the other hand, noise sources which are not correlated to the input signal do not observe the same “on-demand” bandwidth adjustment, and as a result they are filtered by the structure. As demonstrated in Fig. 2.17, noise contributors (such as supply-coupled and DAC-generated noise—including quantization) are intrinsically

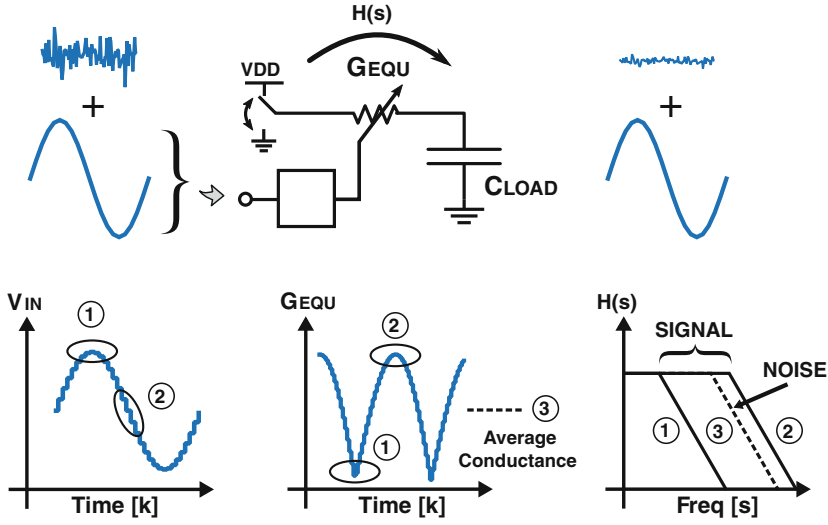


Fig. 2.16 Intrinsic noise filtering mechanism. The signal dependence of the DAC $GEQU$ makes sure that both slow (1) and fast (2) transitions of the desired signal are properly constructed at the output, facing no attenuation. Uncorrelated noise, on the other hand, is filtered by an equivalent cutoff frequency given by the average conductance (3)

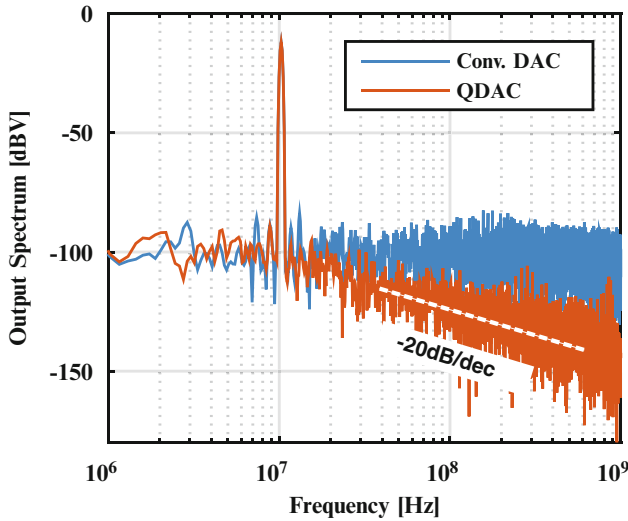


Fig. 2.17 Single-order intrinsic noise filtering, showing a $0.5 V_{pk-pk}$ 10 MHz single-tone

filtered. Interestingly, the equivalent noise cutoff frequency can be numerically approximated by the average QDAC conductance [Eq. (2.7)], as demonstrated with Periodic Steady-State Noise (PNOISE) simulations in Chaps. 3 and 4.

$$f_{-3\text{dB}}(\text{Noise}) = \frac{G_{\text{EQU}}|_{\text{Average}}}{2\pi C_{\text{LOAD}}} \quad (2.7)$$

Moreover, the average QDAC conductance has a strong dependence on the input signal's characteristics, being directly proportional to its amplitude and frequency as approximated by the derivative of the V_{IN} , as shown in Eq. (2.8).

$$\begin{aligned} G_{\text{EQU}}|_{\text{Average}} &= \frac{I_{\text{OUT}}[k]}{V_{\text{DAC}}[k]} \Big|_{\text{Average}} \\ &= \frac{C_{\text{LOAD}}}{V_{\text{DAC}}[k]} \cdot \frac{\Delta V_{\text{IN}}[k]}{T_S} \cdot \left(\frac{dV_{\text{IN}}}{dt} \right) \Big|_{\text{Average}} \\ &\propto C_{\text{LOAD}} \cdot A \cdot \omega \end{aligned} \quad (2.8)$$

Thus,

$$f_{-3\text{dB}}(\text{Noise}) \propto A \cdot \omega \quad (2.9)$$

It can be concluded as a result that: First, for a purely capacitive load the noise cutoff frequency does not depend on the output capacitance, nor the sampling speed. Second, the equivalent noise cutoff frequency scales with the output signal's amplitude and frequency, meaning that for a fixed output frequency the noise cutoff frequency is divided by 2 when the output swing is halved (improving noise filtering).

Figure 2.18 shows the noise cutoff frequency [Eq. (2.7)] versus output frequency for various amplitudes. As noted, for swings below 90 % of the supply voltage the noise cutoff frequency can be even smaller than the actual output frequency, with no attenuation implied to the wanted signal.

The same noise scaling effect is observed with respect to signal's amplitude, with only one remark though: when the voltage difference between the supply and the output is reduced, more conductance is required to convey a given amount of charge. As a result, when the output swing approaches the supply rails (roughly above 70 % of the supply voltage) the DAC average conductance starts increasing exponentially, followed by a corresponding increase of the noise cutoff frequency as observed in Fig. 2.19.

Again, these assumptions are all validated on Chaps. 3 and 4 with PNOISE simulations using the actual QDAC implementations. Although a bit counter-intuitive when first considered, this remarkable feature of a time-varying cutoff frequency—that automatically scales with signal frequency and signal amplitude—is a major advantage of the incremental-charge-based architecture.

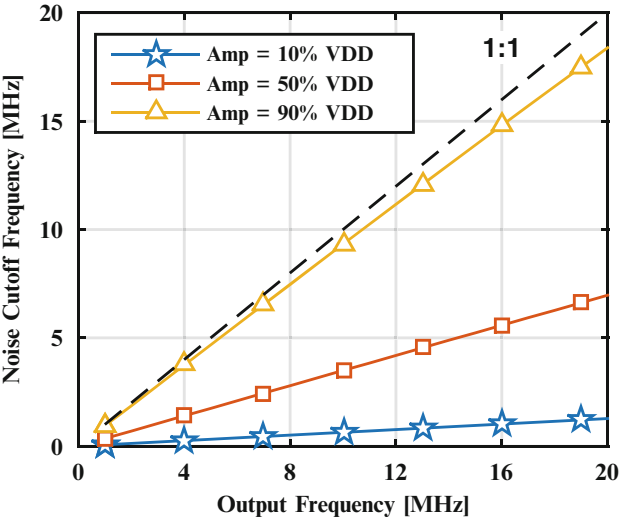


Fig. 2.18 Noise cutoff frequency versus signal output frequency

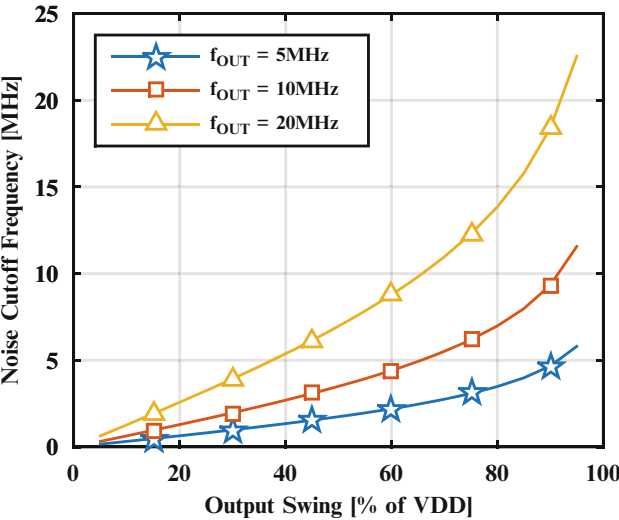
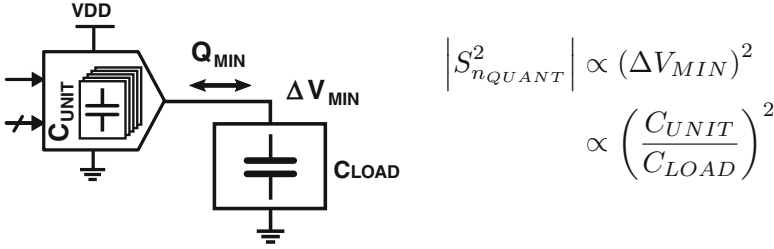


Fig. 2.19 Noise cutoff frequency versus signal amplitude

2.2.1.3 Quantization Noise Scaling

Besides noise filtering given by the signal-dependent RC filter mentioned in Sect. 2.2.1.2, depending on the actual QDAC implementation the quantization noise can be even further reduced.

Chapter 3 introduces a charge-based architecture where quantization noise can be scaled with the ratio between two capacitances, namely an output capacitance C_{LOAD} and a unit capacitance C_{UNIT} with which the QDAC is implemented. Resembling the watermill example, in this charge-based implementation the minimum voltage step resolvable at the QDAC output (and thus quantization noise) can be reduced by either decreasing the unit capacitance C_{UNIT} , or increasing the output capacitance C_{LOAD} , if applicable.



As such, the equivalent number of bits (ENOB) in this architecture can be adjusted by either decreasing C_{UNIT} or increasing C_{LOAD} , based on a required quantization noise level. The DAC number of units, on the other hand, determines what is the maximum amount of charge that can be transferred at once to the output, thus defining the maximum achievable $\Delta V_{OUT}/\Delta T$ that can be produced at the QDAC output.

Further details about the quantization noise scaling feature are given in Chap. 3, Sect. 3.2.3.2.

2.2.1.4 Harmonic Performance

In the charge-based DAC architecture, harmonic performance is strongly dependent on an accurate charge accumulation. Errors in the amount of charge conveyed between the supply and the output load create distortion, hence corrupting signal integrity and affecting harmonic performance. To illustrate the point, for instance the charge-based DAC operation takes into account the existence of a capacitive component to the output load on which charge can be accumulated. For a purely capacitive load, for instance, the output voltage can be represented as follows:

$$\begin{aligned} V_{OUT}[k] &= V_{OUT}[k-1] + \frac{Q_{OUT}[k]}{C_{LOAD}} \\ &= V_{OUT}[k-1] + \frac{(V_{IN}[k] - V_{IN}[k-1]) \cdot C'_{LOAD}}{C_{LOAD}} \end{aligned} \quad (2.10)$$

where C'_{LOAD} is load capacitance value assumed in calculations [Eq. (2.3)].

Since the output voltage is expected to be an unaltered (analog) copy of V_{IN} , an eventual discrepancy between the assumed load capacitance (C'_{LOAD}) and its actual value (C_{LOAD}) can be accounted as:

$$V_{OUT}[k] = V_{IN}[k] + \text{error}[k] \quad (2.11)$$

where the error at time k is given by:

$$\begin{aligned} \text{error}[k] &= V_{IN}[k] \cdot \left(\frac{C'_{LOAD}}{C_{LOAD}} - 1 \right) \\ &\quad + (V_{OUT}[k-1] - V_{IN}[k-1]) \cdot \frac{C'_{LOAD}}{C_{LOAD}} \\ &= V_{IN}[k] \cdot \left(\frac{C'_{LOAD}}{C_{LOAD}} - 1 \right) + \text{error}[k-1] \cdot \frac{C'_{LOAD}}{C_{LOAD}} \end{aligned} \quad (2.12)$$

The analysis of Eq. (2.12) reveals two important aspects: First, a correct charge-based operation relies on a good knowledge of the output load. The more distant the assumed load capacitance (C'_{LOAD}) is from its actual value (C_{LOAD}), the larger the error produced. Second, to make matters worse, the recursive dependence of the error function to its previous value ($\text{error}[k-1]$) demonstrates that the memory effect introduced by the output capacitance not only integrates the wanted charge, but also the errors implied in the charge accumulation. These errors can be caused by multiple reasons, from calculation mistakes to implementation imperfections—including charge leakage.

To illustrate the severeness of error accumulation in the charge-based architecture, Fig. 2.20 shows the impact of a one-time only event, where the amount of charge transferred to the load Z at a particular sampling period is mistakenly reduced by 10 %. If the subsequent charge blocks are kept unchanged, the system operation would be quickly compromised since the accumulated error would make the output voltage drift to one of the supply rails.

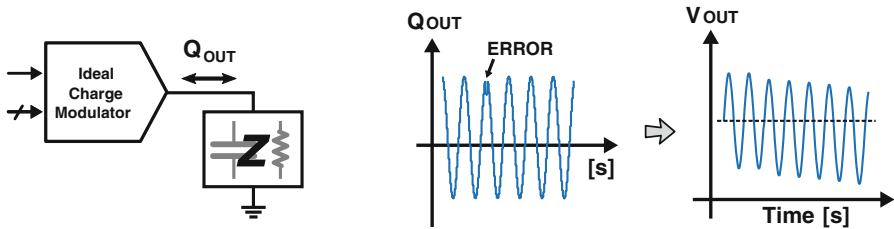
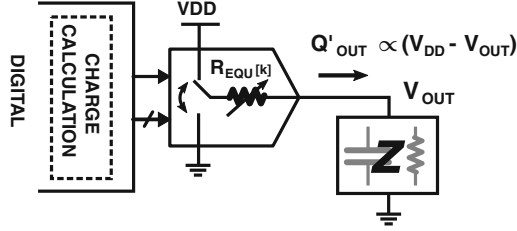


Fig. 2.20 Error accumulation can quickly compromise the charge-based operation if subsequent blocks are not re-adjusted to account for a wrong packet of charge

Basically, the charge-based operation simply would not be feasible if it was not for the “self-healing” effect introduced by the fact that the output charge is also dependent on ΔV 's. In the proposed architecture, the amount of charge transferred to Z during a given sampling period is always proportional to the voltage difference between the supply voltage and the previous V_{OUT} . If an actual time-varying G_{EQU} would be connected between the supply and the output load in order to control the charge output, the amount of charge conveyed to Z in this case would be:



$$Q'_{OUT}[k] = \overbrace{(V_{DD} - V_{OUT}[k-1])}^{\Delta V} \left(1 - e^{(-T_s/\tau[k])}\right) \cdot C_{LOAD} \quad (2.13)$$

where $\tau[k] = R_{EQU}[k] \cdot C_{LOAD}$ and $T_s = 1/F_s$.

The ΔV dependence functions here as an intrinsic feedback where Q'_{OUT} is automatically adjusted when $V_{OUT}[k-1]$ deviates from the expected $V_{IN}[k-1]$, avoiding thus the infinite charge error accumulation.

The principle works as follows: Suppose that equal amounts of charge were to be transferred to the output load at both times A and B ($Q_{A,B}$), but the first charge “block” (Q'_A) is unexpectedly reduced, making the output voltage reach V'_A instead of the wanted V_A . As shown in Fig. 2.21, the increased voltage difference between the “wrong” V'_A and the supply voltage V_{DD} will act in this case increasing the amount of charge transferred at time B (Q'_B).

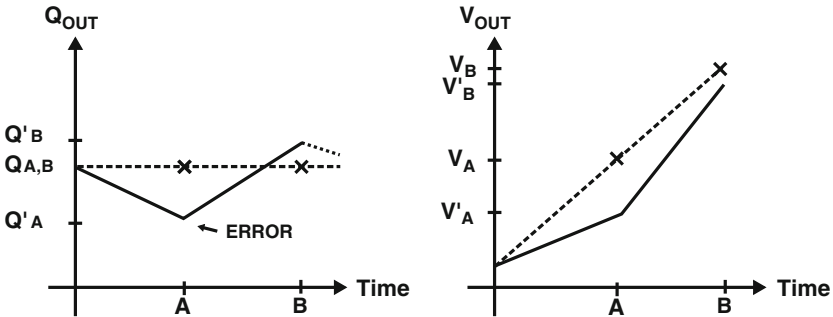


Fig. 2.21 “Self-healing” mechanism. The increased voltage difference implied by a reduced charge “block” increases the amount of charge transferred during the following steps

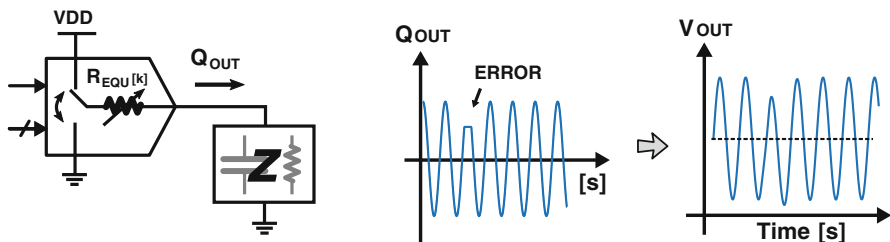
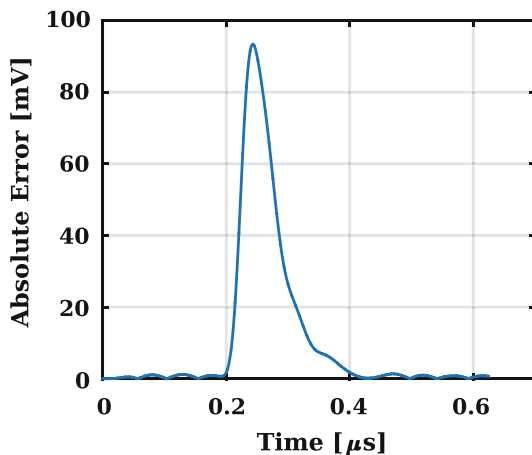


Fig. 2.22 Thanks to the “self-healing” mechanism, the charge-based operation is completely resilient even to large errors in the charge accumulation

Fig. 2.23 Even when large errors are implied, the accumulated error is quickly dissipated



The improved resilience to error accumulation is shown in Fig. 2.22 for a 10 MHz output single-tone, where a total of 50 charge blocks are clipped to a single value representing almost 40 % reduction in the total amount of charge required during that period. As noted, only a localized swing reduction is observed, which is quickly restored once the disturbance is removed.

The architecture’s fast error “dissipation” capability can also be seen in Fig. 2.23. Even in the impracticable situation of 40 % charge reduction, the output voltage error falls to a minimum in less than two cycles.

Though error accumulation in the proposed architecture does not prevent a correct operation, it still affects the charge-based architecture by introducing signal distortion and harmonic degradation. While specific details of each proposed implementation are shown in Chaps. 3 and 4, a broad analysis of charge-based architecture’s harmonic performance is performed here.

Figure 2.24 shows the output spectrum for several degrees of mismatch between the actual output load capacitance (C_{LOAD}) and its assumed value (C'_{LOAD}). As noticed, a clear impact in the odd harmonics is observed, leading to the conclusion (based on Fig. 2.25) that, in order to achieve a third-order harmonic distortion below -60 dBc, an accuracy better than 2 % is necessary.

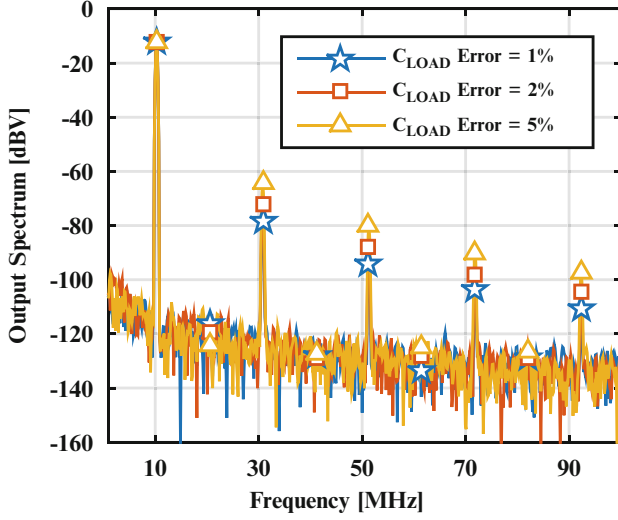


Fig. 2.24 Output Spectrum for several degrees of mismatch between C_{LOAD} and its assumed value

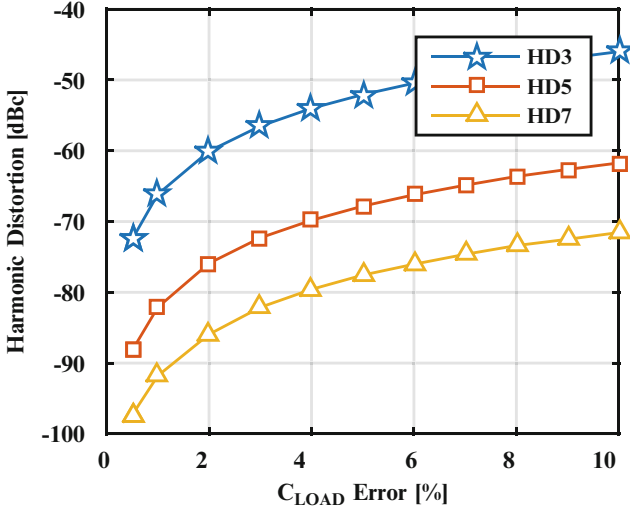


Fig. 2.25 Harmonic distortion versus C_{LOAD} mismatch

Figure 2.26 demonstrates the impact of gain error in the output spectrum. This situation applies when the QDAC unit cell deviates from its expected value. Similarly, a large mismatch incurs increased harmonic distortion, and again less than 2 % mismatch is required to achieve an improved harmonic performance.

Though achieving such degree of accuracy may seem challenging at first, each one of these parameters can be calibrated after fabrication, using low frequency

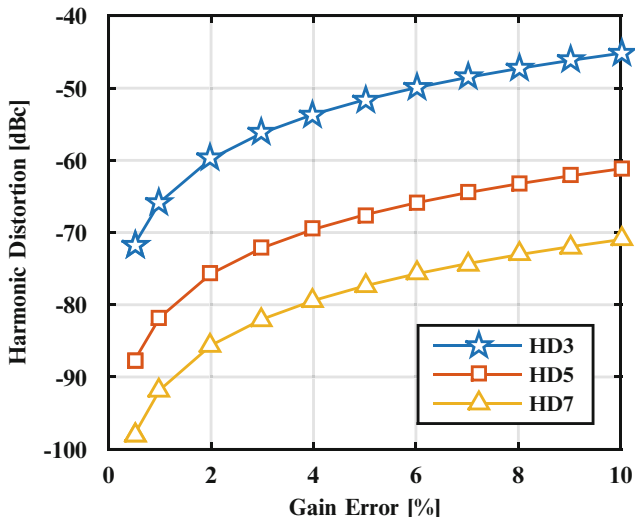


Fig. 2.26 Harmonic distortion versus QDAC gain mismatch

measurements to deduct their exact values. Instantaneous variations of one or more load impedances (such as antenna impedance variations due to changing surrounding conditions) will necessarily require a feedback loop to change the corresponding parameters involved in the charge calculation. Nevertheless, further improvements in harmonic performance can also be targeted by improving the algorithm to account for circuit non-idealities.

2.3 Charge-Based Transmitter

In light of all noise filtering characteristics discussed in Sect. 2.2.1, it seems logical to apply the charge-based operation in order to achieve the stringent noise specifications required to SAW-less transmitters in full duplex.

Starting from a conventional architecture, a typical transmitter implementation is composed of a DAC, followed by a reconstruction filter and a mixer to perform the frequency translation. In most cases, a Pre-Power Amplifier (PPA) is also added as an output stage before driving the PA. In the proposed transmitter architecture, the DAC and the reconstruction filter are exchanged for a QDAC and a baseband capacitance, all the rest remaining the same (Fig. 2.27). Instead of consuming excessive DC power in order to drive the capacitive nodes using strong voltage buffers, now both baseband and RF nodes are going to be driven with charge. Therefore, in the following analysis, the RF impedance is represented as a generic load Z .

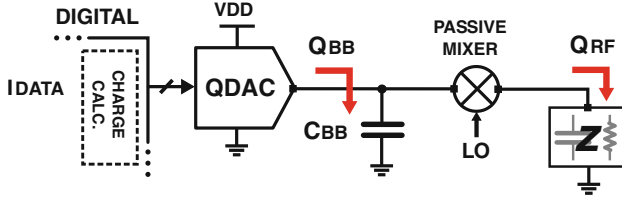


Fig. 2.27 Simplified charge-based TX block diagram

As in the watermill example, the purpose of the baseband capacitance C_{BB} is to provide the functionality of a reservoir, where charge is first buffered before being transferred to the RF output. In combination with the QDAC operation, this charge reservoir should provide all the benefits discussed previously, from intrinsic noise filtering to sampling alias reduction. Compared to the basic QDAC operation, the difference now is that once the RF load is introduced, the charge subtracted from C_{BB} in order to drive the RF load should also be accounted. In this case, the total amount of charge per sampling period required to operate the charge-based structure becomes:

$$Q_{TOTAL} = Q_{BB} + Q_{RF} \quad (2.14)$$

where Q_{BB} corresponds to the amount of charge required to move the baseband voltage across the consecutive input samples, and Q_{RF} stands for the amount of charge subtracted from C_{BB} while driving the RF load Z every time the passive mixer switch is closed.

The operating principles are kept the same. To prove the functionality an ideal passive mixer is added, with a load impedance Z equal to the input capacitance that would be expected from a PPA implementation with a compression point around 10 dBm. Using the same “black-box” QDAC implementation where the output charge Q_{TOTAL} is uniformly distributed over the sampling period, Fig. 2.28 shows the output spectrum of an example 20 MHz multi-tone baseband signal transmitted at 2 GHz. As expected, both noise filtering capabilities and alias attenuation are also clearly noticeable at LO frequency.

The inclusion of a passive mixer and a RF load in the architecture has mainly two effects: First, the total amount of charge consumed (Q_{TOTAL}) is largely increased when a low impedance RF load is used, requiring a similar increase in the QDAC charge capacity. In the two implementations proposed in this work, in both cases the charge capacity is increased by increasing the total number of QDAC elements that can be used to convey charge to the baseband capacitance. In the watermill analogy, it corresponds to increasing the number of buckets available.

Second, since the passive mixer does not provide isolation between baseband and RF nodes, the noise cutoff frequency is also affected by the equivalent RF load impedance translated to the baseband side. As also discussed in [Raz12], when looking from the baseband node at frequencies much smaller than the mixer

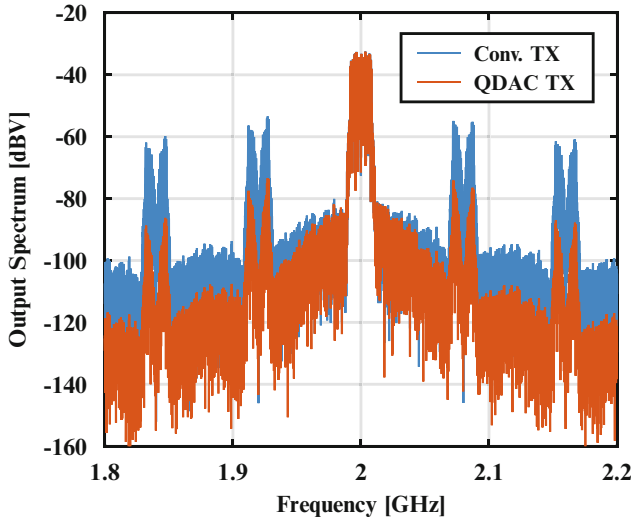


Fig. 2.28 RF spectrum of a charge-based TX driving a 200 fF capacitance at 2 GHz

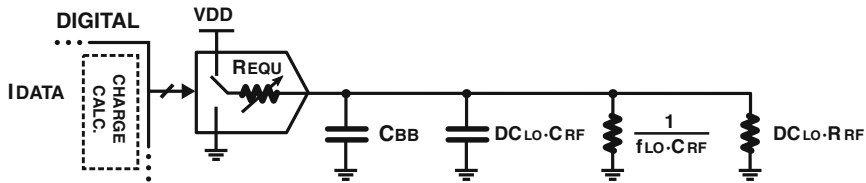


Fig. 2.29 Simplified schematic where the RF load is translated to the baseband side

switching frequency ($\omega \ll 2\pi f_{LO}$), an output RF capacitance can be translated into two components: A capacitive component given by C_{RF} scaled by the LO duty cycle (DC_{LO}), and a resistive component given by the average (DC) current drained by C_{RF} ($R_{C_{RF}} = 1/C_{RF}f_{LO}$). Similarly, an RF resistance is also translated to the baseband by looking at the average current drained by R_{RF} . In a differential implementation where each baseband node is connected twice per LO period to the RF node, the LO duty cycle is doubled. Figure 2.29 shows the equivalent schematic considering the translated RF load.

The resulting noise cutoff frequency considering the RF load is hence given by:

$$f_{-3dB}(\text{Noise}) = \frac{\left(G_{EQU}|_{\text{Average}} + f_{LO}C_{RF} + \frac{1}{DC_{LO}R_{RF}} \right)}{2\pi(C_{LOAD} + DC_{LO}C_{RF})} \quad (2.15)$$

Figure 2.30 shows the impact of a resistive RF load for various baseband capacitances considering a 0.7 Vpp 10 MHz single-tone transmitted at 2 GHz. As

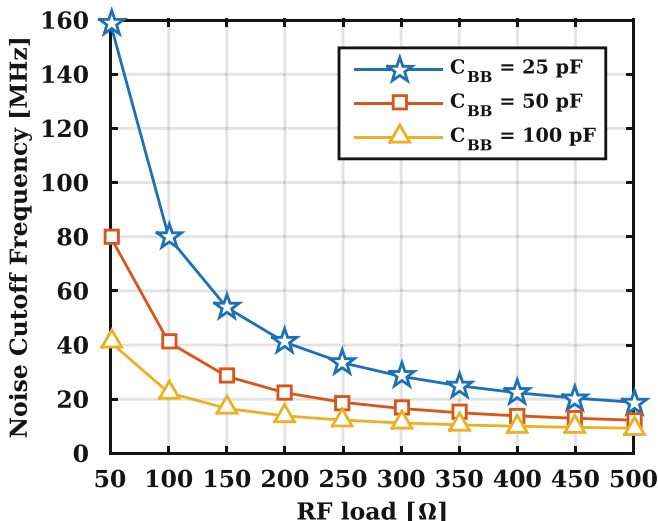


Fig. 2.30 Noise cutoff frequency versus RF load resistance

noticed, the noise cutoff frequency is considerably shifted to higher frequencies when low impedance RF loads are used, which can be addressed by either increasing the baseband capacitance, or applying impedance transformation. This particular issue is addressed in Chap. 4, where the charge-based architecture is used to drive a 50 Ω load representing the PA input impedance.

2.3.1 Power Efficiency

Besides noise reduction, incremental charge-based operation can also provide great improvements in terms of power efficiency. For a more fundamental understanding of the potential improvements and how it trades with noise performance, only the charge intake represented by Q_{TOTAL} is considered in this analysis. The impact of additional contributors including QDAC digital operation and LO buffering are discussed in subsequent chapters.

By first looking at QDAC independently and considering the amount of power required to drive the baseband capacitance alone, it can be noticed that the charge-based architecture fundamentally behaves as a Class-B topology, since it only drains charge from the supply when the total amount of charge in the system should be increased—thus for only half of the signal period (Fig. 2.31). As a result, when driving the baseband capacitance only, the charge-based DAC can provide a maximum efficiency of 78.2 % as verified in simulations and shown in Fig. 2.32 for various signal amplitudes. However, the maximum efficiency corresponds to an output swing equal to V_{DD} , which cannot be realized since it would require an

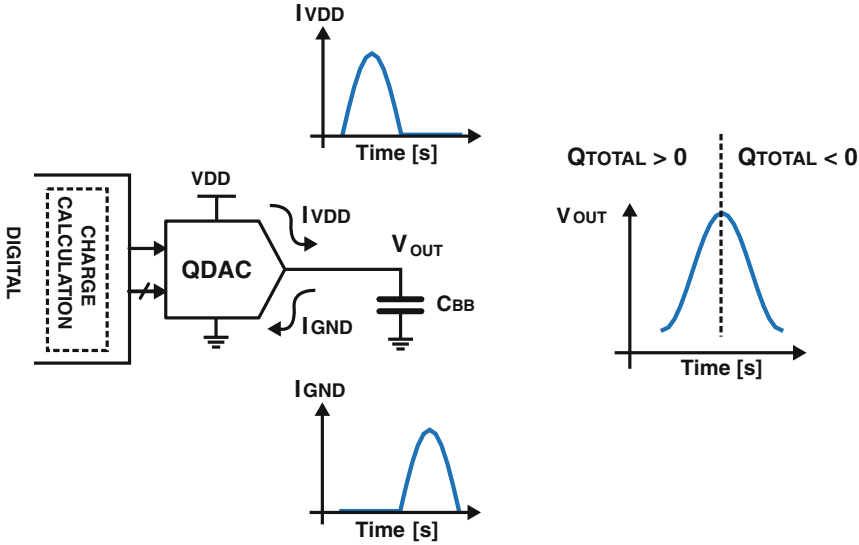


Fig. 2.31 Simplified schematic depicting QDAC Class-B operation

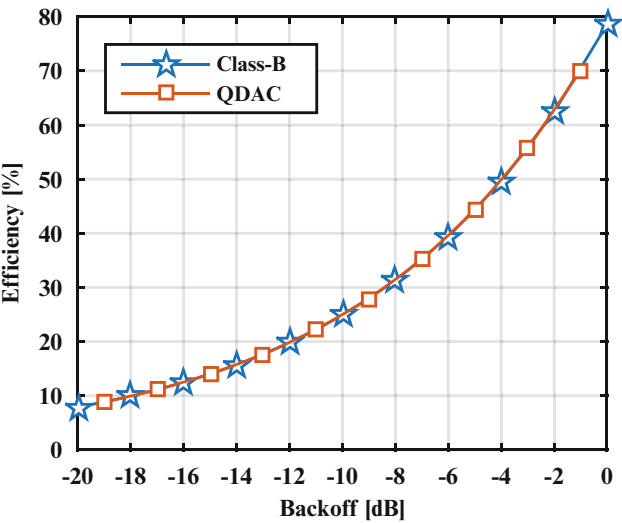


Fig. 2.32 Efficiency versus Backoff considering C_{BB} only

infinite QDAC conductance. Nevertheless, swings of at least 90 % of the supply voltage are completely feasible, corresponding to a QDAC efficiency of 70 %.

When the RF load is included, the power consumption becomes a function of two contributors, namely the baseband and RF charges. Since the Q_{RF} component corresponds to the only fraction of Q_{TOTAL} that is actually transmitted, the maximum

Fig. 2.33 Example charge-based transmitter considered for Efficiency calculations

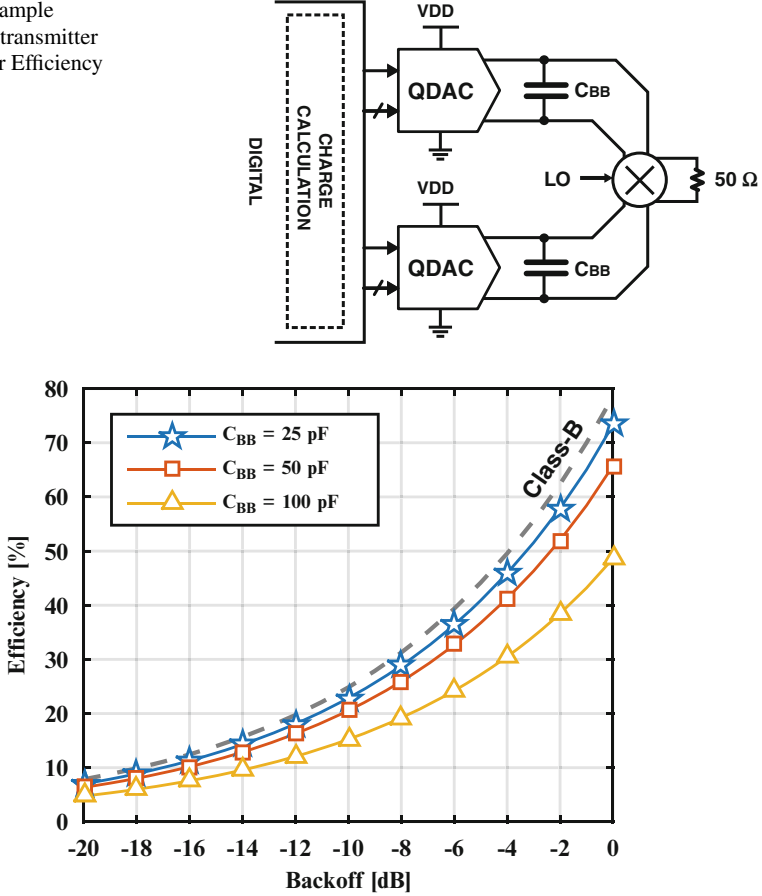


Fig. 2.34 QDAC TX Efficiency versus Backoff considering a 50 Ω load and various C_{BB} values

power efficiency of a charge-based transmitter becomes mainly determined by how much power is spent to drive C_{BB} . On the other hand, the baseband capacitance is also responsible for all the benefits of charge-domain operation, so reducing C_{BB} (in order to improve efficiency) will necessarily come at the cost of reducing the noise filtering capabilities provided by the architecture.

This tradeoff is explored here with a simple exercise considering a differential I/Q charge-based TX driving a 50 Ω load (Fig. 2.33). In this example, the efficiency versus backoff from theoretical maximum swing ($V_{pp} = V_{DD}$) is shown in Fig. 2.34. As noticed, the increasing amount of power spent to drive the baseband capacitances decrease the maximum achievable power efficiency as C_{BB} is increased.

To demonstrate how efficiency trades with noise filtering in this specific case, Fig. 2.35 shows both efficiency and noise cutoff frequency at 2 dB backoff for different baseband capacitances.

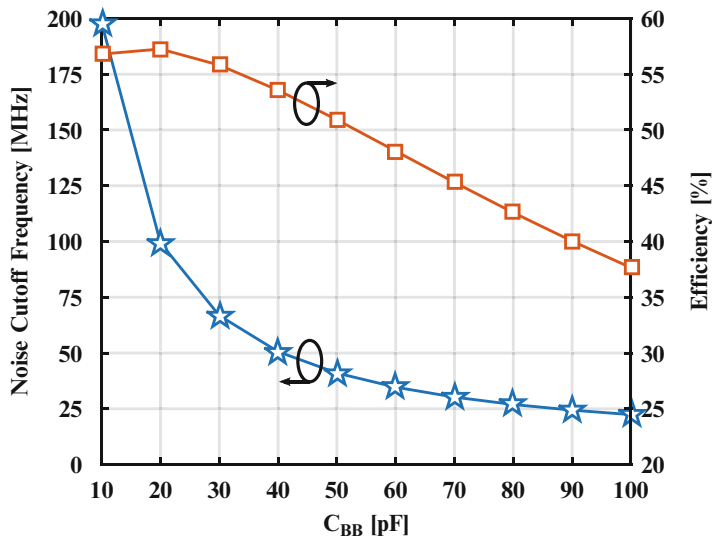


Fig. 2.35 Efficiency and noise cutoff frequency versus C_{BB} at 2 dB backoff

Thus, the charge-based architecture provides a flexible solution where power efficiency can be easily improved in cases where OOB noise requirements are more relaxed, which unfortunately is not the case in our target application. Therefore, the following chapters make a solid case demonstrating the noise filtering capabilities, rather than the considerable improvements in power efficiency that could also be attained with the charge-based architecture.

2.4 Conclusion

In this chapter the fundamentals of incremental-charge-based operation are discussed. Starting with the analogy of a watermill whose output power has to be controlled in a precise and timely fashion, most of the benefits derived from this unconventional operating mode could be implied. The incremental-charge-based operation is based on a controlled charge convey and accumulation at the different nodes existing in the system. The voltage swings are produced by “incrementally” adding or subtracting charge from the several charge accumulators, in a way that only the added charge is drained from the supply.

Using a generic QDAC “black-box” implementation that is capable of delivering controlled amounts of charge defined with simple calculations, three main advantages of incremental-charge-based operation could be demonstrated: First, the continuous (or discrete-time) charge accumulation yields an inherent quasi-linear interpolation that significantly attenuates the sampling aliases. Second, the

combination between the charge accumulator and the QDAC operation provides a single-order low pass filter that attenuates different noise contributors including quantization noise, which can be even further reduced depending on the QDAC implementation. Third, since power is only consumed when the total amount of charge in the system has to be increased, when the noise filtering capabilities are relaxed the charge-based architecture can also provide ultimate power efficiencies similar to Class-B topologies.

However, the incremental operation relies on a precise tracking of the absolute amount of charge existing in the system at all times. Errors implied in the charge accumulation are also integrated over time and produce signal distortion that can degrade the achieved harmonic performance. For a flawless operation, an accurate definition of every baseband and RF component is hence mandatory, as well as observation and reduction of every circuit non-ideality that can affect or distort the system charge balance, such as leakage and switch charge-injection.

<http://www.springer.com/978-3-319-45786-4>

Charge-based CMOS Digital RF Transmitters

Paro Filho, P.E.; Craninckx, J.; Wambacq, P.; Ingels, M.

2017, XXIX, 152 p. 156 illus., 98 illus. in color.,

Hardcover

ISBN: 978-3-319-45786-4