

MASTER THESIS:

**“A HIGH EFFICIENCY SWITCHED-CAPACITOR
POWER AMPLIFIER WITH REAL-TIME
CALIBRATION FOR HARMONIC
SUPPRESSION”**

CHARALAMPOS PANAGIOTIS CHARALAMPIDIS

MASTER THESIS:

**“A HIGH EFFICIENCY SWITCHED-CAPACITOR
POWER AMPLIFIER WITH REAL-TIME
CALIBRATION FOR HARMONIC
SUPPRESSION”**

CHARALAMPOS PANAGIOTIS CHARALAMPIDIS

M.Sc. Electrical Engineering, TU Delft

Supervisors:

Dr. Morteza Alavi (TU Delft)
Ao Ba (Dialog Semiconductor)

Thesis Committee:

Prof.dr. L.C.N. de Vreede (TU Delft)
Dr. Qinwen Fan (TU Delft)
Dr. Morteza Alavi (TU Delft)



Work performed at Dialog Semiconductor
Het Zuiderkruis 53
5215 MV 's-Hertogenbosch
The Netherlands

ACKNOWLEDGEMENTS

It is with equal parts pride and relief that I present this completed thesis report. None of this would be possible without the guidance and encouragement of my supervisors Dr. Morteza Alavi and Ao Ba. These gentlemen were always eager to provide support, both creative and moral, whenever it was needed the most. Thank you for your patience and kindness.

I would also like to thank my parents, Androula and Nikolaos, for providing me with every opportunity in the world and believing in me every step of the way for my entire life. My love and admiration for you could never be enough.

Special shout-out to my brother Georgios, as well as to all my friends back home who have kept me company throughout these last couple of decades. We are all going to make it.

ABSTRACT

A switched-capacitor power amplifier (SCPA) is a very desirable solution to the problem of enabling amplitude modulation while utilizing high-efficiency switching PA topologies for non-constant envelope wireless RF signals, due to its highly linear AM curve, low complexity, and high scalability. However, drawbacks exist in the form of low efficiency at lower input amplitude levels and strong harmonic distortion.

In this master thesis, an SCPA is designed in an advanced technology node for wireless signal transmission. Special considerations are taken to explicitly suppress the harmonics of second and third order by manipulating multiple waveforms. A feedback loop is designed in order to facilitate proper suppression in the presence of variations. The issue of power back-off efficiency is addressed through the method of impedance matching utilized.

The system operates between the frequencies of 2.4 GHz and 2.5 GHz. For a carrier frequency of 2.4GHz, a peak system efficiency of 42.75% and a -6dB back-off system efficiency of 22.52% are achieved. The system presents high AM curve linearity with an IIP3 of 42.3 dB. Second- and third-order harmonics remain well below the specification of -41 dBm in all process corners, provided that proper tuning is performed.

Table of Contents

1: INTRODUCTION	1
1.1: Background	1
1.2: Design Goals	1
1.3: Target Specifications	2
2: SCPA THEORY	3
2.1: Operating Principle	3
2.2: Efficiency	4
2.3: Harmonic Suppression	6
2.3.1: Duty Cycle Tuning	6
2.3.2: Phase Tuning	7
2.4: Waveform Combination	8
2.5: Explicit Cancellation	11
2.6: Dual-Array	13
2.7: Transformer	14
2.7.1: Mirrored Waveforms	15
2.8: Design Proposal	16
3: DESIGN	17
3.1: Matching	17
3.1.1: "No Shunt" Matching	18
3.1.2: Comparison of Matching Techniques	19
3.1.3: "No Shunt" Regions	20
3.1.4: Efficiency Comparison & Simulation	22
3.2: Tuner Design	24
3.2.1: Current Starve	25
3.2.2: "Coarse/Fine" Configuration	27
3.2.3: Phase Tuner	31
3.2.4: Tuner Diagrams	34
3.3: Tuning Loop	34
4: IMPLEMENTATION	38
4.1: Transformer	38
4.2: PA Arrays	40
4.3: Tuners	41
4.3.1: Phase Detector	42
4.3.2: Tuner Control	43
4.3.3: Tuner Performance	44
4.3.4: Phase Tuners	45
4.3.5: "Phase/Duty Coupling"	46
4.3.6: Harmonic Performance	47
4.4: Feedback Loop	48
4.4.1: Logic	49
4.4.2: Parameters	50
4.4.3: Operation	51
5: SIMULATION RESULTS	54
5.1: PA Performance	54
5.2: AM Performance	55
6: CONCLUSION	57
6.1: Thesis Summary	57
6.2: Comments and Recommendations for Future Work	57
7: REFERENCES	59
8: APPENDIX	60

1: INTRODUCTION

1.1: Background

Switching power amplifiers (PAs) have become increasingly popular for integrated RF applications due to their superior efficiency. Class-D amplifiers, in particular, are inverter-based designs which can benefit greatly from the increased switching speed and decreased power consumption provided by modern CMOS scaling [1]. Thus, such topologies are desirable for high-frequency wireless transmitters. Considering the fact that many wireless communication devices such as smartphones, wearables, and sensors are battery-powered, the demand for high power efficiency becomes clear.

However, despite their high efficiency, class-D amplifiers suffer from a lack of linearity; the drain voltage of an inverter can only fluctuate between its two supply rails. Although this is no issue for constant-envelope modulation, modern *nonconstant*-envelope schemes also employ amplitude modulation. A challenge thus arises in the linear operation of a class-D-based design.

A *digital power amplifier* (DPA) [2] is an amplifier architecture consisting of multiple amplifier units connected together. A digital code controls the number of active units so as to provide amplitude modulation; output power is increased as more units are enabled, and a nonconstant-envelope signal can be constructed. This is similar to the operation of a digital-to-analog converter (DAC), and a system utilizing this architecture for high-frequency signals can be referred to as an *RF-DAC*.

A *switched-capacitor power amplifier* (SCPA) [1] is an RF-DAC design which employs class-D amplifier units, each with its own dedicated unit capacitor. Charge is shared among the capacitors, performing voltage division related to capacitor ratios. As these ratios can be very precisely set, the result is a highly linear amplitude modulation curve. The SCPA architecture is ideal for high-performance wireless transmitters, especially when considering the exceptional power efficiency provided by its class-D cells.

1.2: Design Goals

Although SCPA efficiency can be very high at maximum input amplitude, dynamic power dissipation still takes place due to the process of charging and discharging of the capacitor array each cycle. This phenomenon becomes more prominent at lower input amplitudes, or *power back-off* (PBO). When additional factors are considered, such as from CMOS dynamic power dissipation or passive ohmic losses, overall PBO system efficiency can reach very low values.

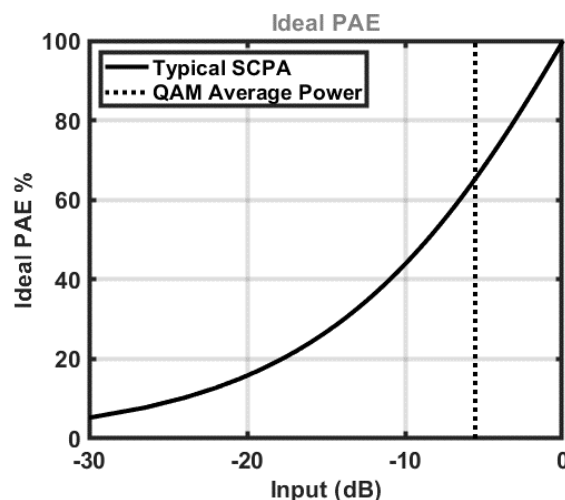


Figure 1.1: Ideal SCPA Power-Added Efficiency vs. Input Power. Also: Average Power for QAM.

Since nonconstant-envelope modulation schemes can display significant *peak-to-average power ratios* (PAPR) [3], it is important that efficiency at power back-off is kept high. Figure 1.1 displays a plot for ideal power-added efficiency of a typical SCPA, as will be described later in this thesis, versus input power. The average signal power

for QAM, a popular nonconstant-envelope modulation scheme, is also pictured. Maximizing efficiency for reduced input amplitudes is an important design goal, although not one unique to this specific architecture.

Minimizing harmonic output is another important aspect of the design process. Transmitted signals must comply to regulations regarding unwanted emissions, meaning that signal components beyond the designated out-of-band frequency range must be suppressed [5]. Class-D-based architectures are highly prone to violating these standards, due to the powerful harmonics present in their pulse-shaped output waveforms, as can be seen from the spectrum of figure 1.2.

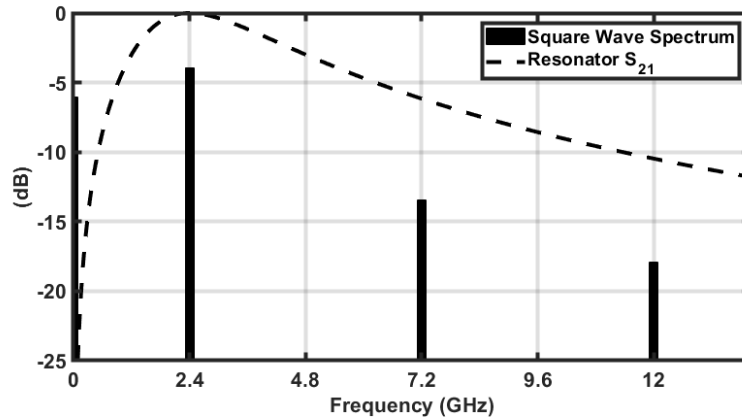


Figure 1.2: Spectrum of Square Wave and S_{21} Plot of Example LC Resonator.

Figure 1.2 also displays a plot for the gain frequency response of an LC resonator, as typically used in SCPA designs [1]. By itself, this network may not provide adequate suppression, and additional filtering can be very costly in terms of efficiency and area, especially for on-chip implementations. Therefore, alternative harmonic suppression methods must be examined.

1.3: Target Specifications

In this thesis, a switched-capacitor power amplifier is proposed, designed, and simulated. The PA must be able to transmit signals between the frequencies of 2.4 GHz to 2.5 GHz at a peak output power level of 13 dBm, or ~ 20 mW. A DAC resolution of 7 bits is desired for amplitude modulation, meaning that 128 distinct power levels should be available.

Output power for harmonics of second and third order must be kept below -41 dBm. As this is the primary goal of the project, the methods, techniques, and design principles employed to achieve this are discussed in great detail throughout the thesis. Additional measures are also taken to ensure proper operation in the presence of process variation.

The above specifications are also displayed in table 1.1. Although not as strictly defined, additional parameters are taken into account during the design and implementation process. Efficiency is considered throughout, as the set specifications must be achieved without inducing excessive power dissipation. A method of increasing efficiency is employed, specifically pertaining to lower input amplitudes.

Carrier Frequency Range	2.4 GHz – 2.5 GHz
Fundamental Output Power	13 dBm
DAC Resolution	7 bits
2nd Harmonic Output Power	< -41 dBm
3rd Harmonic Output Power	< -41 dBm

Table 1.1: Targeted Specifications.

2: SCPA THEORY

2.1: Operating Principle

A switched-capacitor power amplifier (SCPA) [1], in its simplest form, is composed of an array of capacitors connected in parallel through their top plates, with each capacitor individually connecting to a switch on its bottom plate. The switches, often implemented as simple CMOS inverters, can bring the bottom-plate voltage to either V_{GND} or V_{DD} , depending on their input voltages.

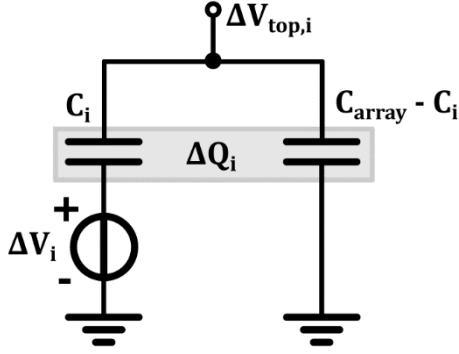


Figure 2.1: Schematic of Voltage Division on Parallel Capacitor Array.

A change in voltage ΔV_i on the bottom plate of a capacitor C_i brings forth a charge difference $\Delta Q_i = C_i \Delta V_i$. This charge is then shared with the entire array capacitance C_{array} , creating a change in the common top-plate voltage, as seen in the illustration of figure 2.1.

$$\Delta V_{top,i} = \frac{\Delta Q_i}{C_{array}} = \frac{C_i}{C_{array}} \Delta V_i \quad (2.1)$$

The capacitor array performs voltage division, with a ratio equal to that of switched capacitance over total array capacitance. Assuming an array composed of N equal unit capacitors C_{unit} , of which n are connected to V_{DD} and $N - n$ are grounded, as seen in the schematic of figure 2.2, the top-plate voltage can be calculated as:

$$V_{top} = \frac{n C_{unit}}{C_{array}} V_{DD} = \frac{n}{N} V_{DD} \quad (2.2)$$

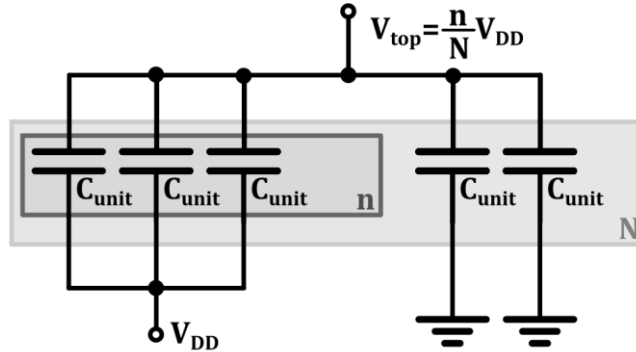


Figure 2.2: Schematic of Voltage Division on Unary Capacitor Array.

If the n active units are switched between V_{DD} and V_{GND} at the LO frequency f_{LO} , the top-plate voltage will display a similar waveform, with an amplitude set by the number of active units. Figure 2.3 presents a schematic of this configuration. If a square wave clock signal is used to trigger the switches, the input voltage v_{in} at the bottom plates of the active unit capacitors will be a rail-to-rail square waveform of frequency f_{LO} .

Through capacitive voltage division, the top-plate waveform v_{top} is also a square, albeit with an amplitude adjusted through a factor of n/N . The fundamental component of this waveform can be calculated as:

$$v_{top,1}(t) = \frac{2}{\pi} V_{DD} \frac{1}{N} n(t) \cos(\omega_{LO} t) \quad (2.3)$$

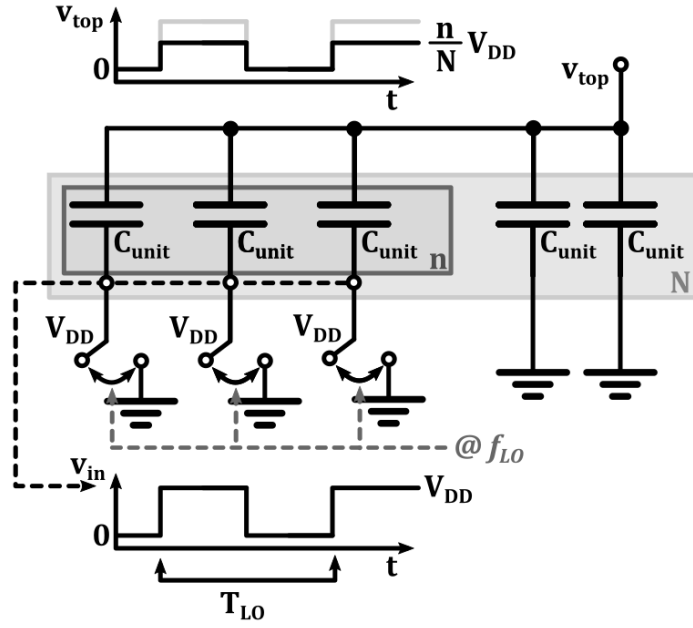


Figure 2.3: Schematic of RF-DAC Operation of Switched-Capacitor Array.

The system operates as an RF-DAC, or mixing-DAC, since a digital word controlling the number $n(t)$ of active units at a given moment is upconverted to the RF carrier frequency.

The top-plate voltage of the capacitor array is transferred onto the load R_L through a series inductor L , which resonates with the total array capacitance C_{array} at the desired frequency ω_{LO} , as shown in the schematics of figure 2.4.

$$L = \frac{1}{\omega_{LO}^2 C_{array}} \quad (2.4)$$

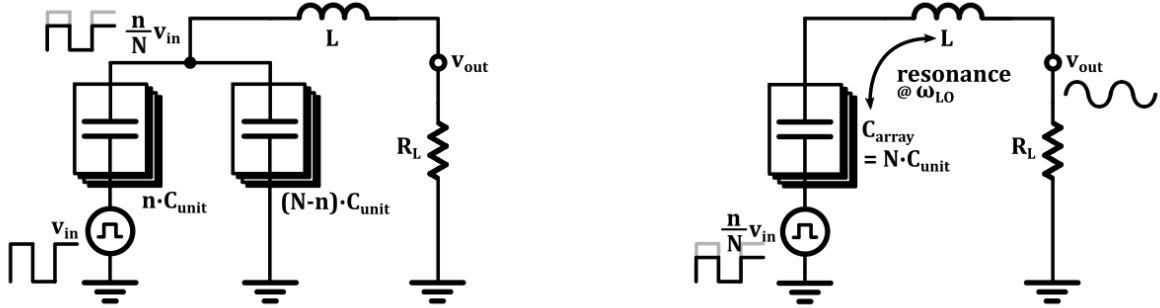


Figure 2.4: Left: Schematic of Switched-Capacitor Array with Resonator and Load. Right: Thevenin Equivalent. Adapted from [1].

It can be seen from the Thevenin equivalent of the array that a constant capacitance is seen at the shared capacitor top plate. Therefore, there is no code dependency on the matching. Additional reactive components may be used in order to transform the load impedance into an appropriate value, so as to achieve the desired output power. For the sake of simplicity, only R_L will be considered for the following examples.

2.2: Efficiency

During the fast switching of the active units, current is drawn from the supply, charging and discharging the entire array capacitance each cycle. This results in switched-capacitor dynamic power consumption [1]:

$$P_{SC} = C_{in} V_{DD}^2 f_{LO} \quad (2.5)$$

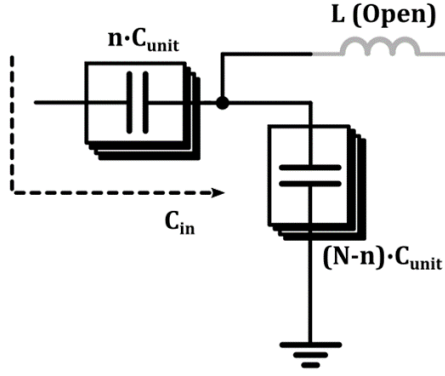


Figure 2.5: SCPA Equivalent Circuit for Calculation of Input Capacitance. Adapted from [1].

In order to calculate the value of this expression, the input capacitance C_{in} seen from the supply must be quantified. Figure 2.5 displays the equivalent circuit as seen from the bottom plates of the active PA units during the switching edges. C_{in} is the series combination of switched and grounded capacitors.

$$C_{in} = \frac{n(N-n)}{N^2} C_{array} \quad (2.6)$$

This expression curiously becomes equal to zero in the case of $n = N$ i.e., when all PA units are active. Indeed, the top and bottom plates of all capacitors are then equal to V_{DD} , with no charge inside any of the capacitors.

The fundamental RMS output power to the load can be derived from the voltage expression produced above in equation 2.3, and is equal to:

$$P_{out} = \frac{2}{\pi^2} \frac{n^2 V_{DD}^2}{N^2 R_L} \quad (2.7)$$

In the ideal case, where only P_{out} and P_{SC} are considered, the power-added efficiency of the SCPA is:

$$PAE_{ideal} = \frac{P_{out}}{P_{DC}} = \frac{P_{out}}{P_{out} + P_{SC}} = \frac{1}{1 + \frac{(N-n)\pi^2 R_L C_{array} f_{LO}}{2n}} \quad (2.8)$$

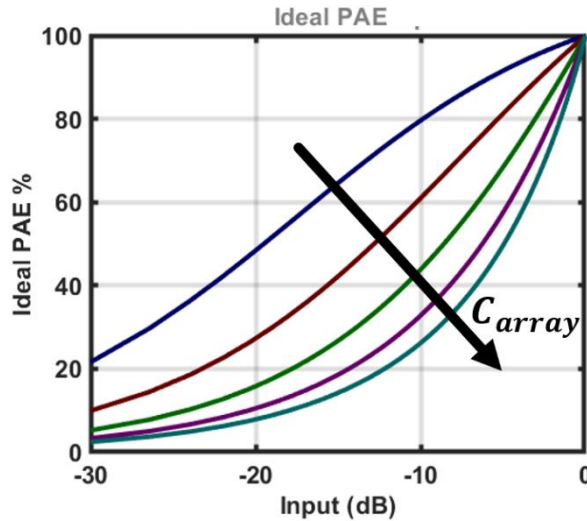


Figure 2.6: Plot of Ideal PAE vs. Input Amplitude and Array Capacitance.

A few important deductions can be made from this expression. Most importantly, SCPA efficiency is code-dependent; the plot of figure 2.6 displays the relation of PAE to input amplitude, represented as the ratio of n/N in dB, where it can be seen that the system can become largely inefficient at power back-off. Simply explained, a large amount of capacitance C_{in} is being charged every cycle, with little power to the output at low amplitudes. At

maximum amplitude, $C_{in} = 0$, meaning that the SCPA reaches 100% ideal PAE as all of the expended power reaches the load.

Reducing the load resistance R_L can improve PAE, by virtue of increasing output power with no additional cost to switched-capacitor dynamic power. However, R_L cannot be infinitely small, as ohmic resistance related to active and passive components will disproportionately load the PA, severely limiting efficiency.

The factor $C_{array} f_{LO}$ stemming from the expression of dynamic power dissipation is also very important. Higher frequencies will naturally result in more edges, and thus instances of charge expenditure, per second. The amount of charge per cycle, however, is set by the total array capacitance C_{array} . Limiting factors to this design parameter are related to impedance matching and will be discussed later in this thesis.

2.3: Harmonic Suppression

The capacitor array top plate voltage v_{top} displays a square waveform. As such it contains odd-order harmonic components, with amplitudes of:

$$A_n = \frac{2}{n\pi} V_{DD}, \quad n = 2k + 1 \quad (2.9)$$

The third harmonic (HD3) can be especially troublesome; not only does it have a large amplitude of only 1/3 of that of the fundamental, but it also appears at a relatively low frequency which can infringe on other signals. Therefore, effort must be made to reduce HD3 in particular.

The resonator composed of C_{array} and L does not provide perfect suppression of the third harmonic, as it operates as a band-pass filter of only first order. Additional filtering may be implemented, but at the cost of area and efficiency, as more passive components are involved. Moreover, adequately suppressing HD3 without affecting the fundamental would require a complex high-order filter. A low-pass filter will be recommended as a future improvement upon the design proposed in this thesis, but at significantly lower complexity.

2.3.1: Duty Cycle Tuning

Duty cycle tuning, or conduction angle calibration [4], is another approach that can suppress the third harmonic. It involves modifying the input LO waveform controlling the SCPA switches, and thus the array top-plate voltage, so as to cancel the third harmonic.

A pulse waveform i can be broken down as follows:

$$v_i(t) = \sum_n \frac{2}{n\pi} A_i \sin(n\pi d_i) e^{jn(\omega_i t + \phi_i)} \quad (2.10)$$

Wherein A_i , ω_i , ϕ_i , and d_i refer to the amplitude, angular frequency, phase, and duty cycle of waveform i . The normalized Fourier coefficients for the first three components of a pulse waveform of duty cycle d are:

$$\text{Fundamental: } \sin(\pi d), \quad \text{HD2: } \frac{1}{2} \sin(2\pi d), \quad \text{HD3: } \frac{1}{3} \sin(3\pi d)$$

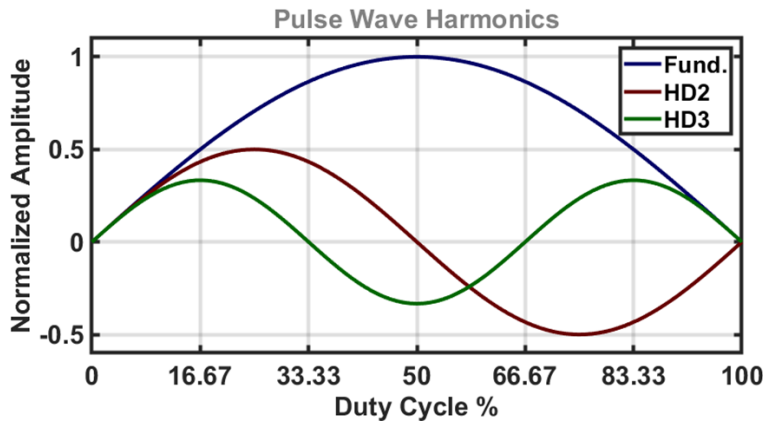


Figure 2.7: Plot of Normalized Amplitude for Fundamental, Second-, and Third-Order Harmonic Components vs. Duty Cycle.

The plot of figure 2.7 provides visualization of the above expressions. Setting $HD3 = 0$ results in a duty cycle of either $d_\alpha = 1/3$ or $d_\beta = 2/3$. Two things are worth noting here. First, the amplitude of the fundamental is reduced to $\sin(\pi/3)$, or 86.6% of its maximum value for either of $d_{\alpha,\beta}$ compared to the original $d = 1/2$. This predictably reduces PAE at $n < N$ as discussed previously, since the switched-capacitor dynamic power dissipation will remain constant regardless of duty cycle.

$$PAE_{ideal}\left(d = \frac{1,2}{3}\right) = \frac{1}{1 + \frac{(N-n)\pi^2 R_L C_{array} f_{LO}}{3n/2}} \quad (2.11)$$

Additionally, setting the duty cycle to any value other than $d = 1/2$ will result in a non-zero $HD2$. This second harmonic appears at a lower frequency, while having a higher amplitude than $HD3$, possibly violating the spectral purity requirements.

A solution to this issue can be found by adding together waveforms of both $d_\alpha = 1/3$ and $d_\beta = 2/3$ with equal amplitudes. As can be seen in figure 2.7, the individual $HD2$ components of these signals are equal in magnitude and opposite in phase, the resulting waveform should be devoid of both second- and third-order harmonics.

Adding together the two waveforms $= v_\alpha$ and v_β :

$$HD2: \frac{1}{2} \left[\sin\left(\frac{2\pi}{3}\right) + \sin\left(\frac{4\pi}{3}\right) \right] = 0, \quad HD3: \frac{1}{3} [\sin(\pi) + \sin(2\pi)] = 0$$

A time domain visualization of the resulting waveform can be seen in figure 2.8.

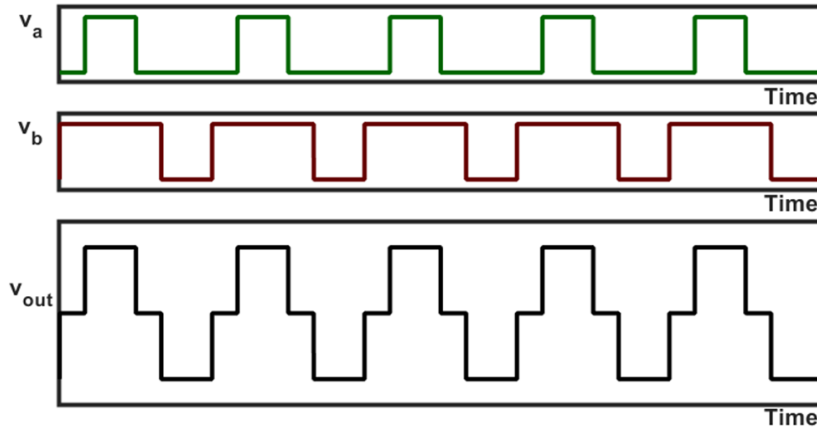


Figure 2.8: Component Waveforms and Output Waveform for Dual-Waveform Duty Cycle Tuning Configuration.

2.3.2: Phase Tuning

It is worth noting that the two component waveforms are considered to be of equal amplitude, frequency, and phase, as established in equation 2.10. These parameters, however, can provide alternate ways to suppress second- and third-order components.

Providing a phase shift $\Delta\varphi$ between waveforms v_α and v_β of equal duty cycle d will result in the following expressions for the harmonics of interest:

$$HD2: \frac{1}{2} (1 + e^{j2\Delta\varphi}) \sin(2\pi d), \quad HD3: \frac{1}{3} (1 + e^{j3\Delta\varphi}) \sin(3\pi d)$$

Setting $\Delta\varphi = \pi/3$, or 60° , and $d = 1/2$ will bring both harmonics to zero [6]. The resulting waveform can be seen in figure 2.9. Compared to the waveform created through duty cycle tuning, no differences can be discerned. Indeed, calculating the fundamental voltage in both of these cases results in a factor of $\cos(\pi/6)$, or 86.6% of that of a typical SCPA.

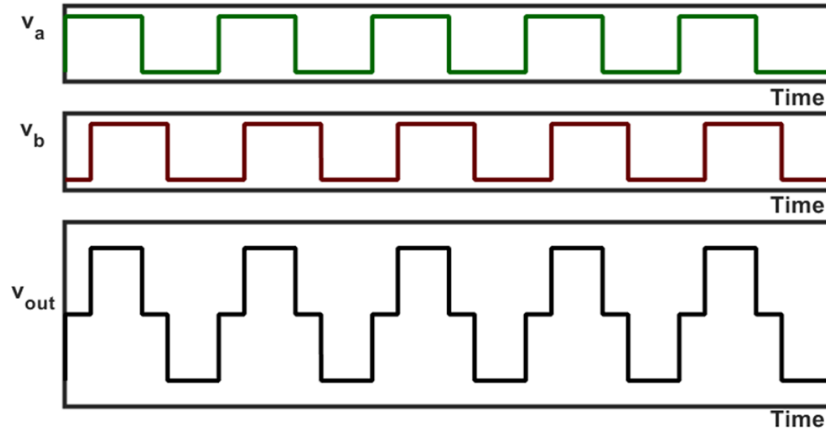


Figure 2.9: Component Waveforms and Output Waveform for Dual-Waveform Phase Tuning Configuration.

2.4: Waveform Combination

In order to compare the effects of these two multi-waveform methods – duty cycle tuning and phase tuning – on efficiency, a method of combination must first be established. The switched-capacitor array already provides linear summation of voltages, and can do so regardless of the phase, frequency, or amplitude of the waveforms controlling the switches. Therefore, driving half the units at phase φ_α with the rest at φ_β will result in the desired top-plate waveform. The same is also true for the combination of d_α and d_β in the multi-duty-cycle configuration.

Figure 2.10 displays a schematic of a single-array, dual-waveform configuration, specifically the multi-phase variant. For a given code n , waveform v_α at phase φ_α drives n units, whereas another n are driven by v_β with φ_β . Thus, each input contributes to v_{top} with a factor of $n/2N$. The top plate voltage is calculated as:

$$v_{top}(t) = \frac{n}{2N} (v_\alpha(t) + v_\beta(t)) \quad (2.12)$$

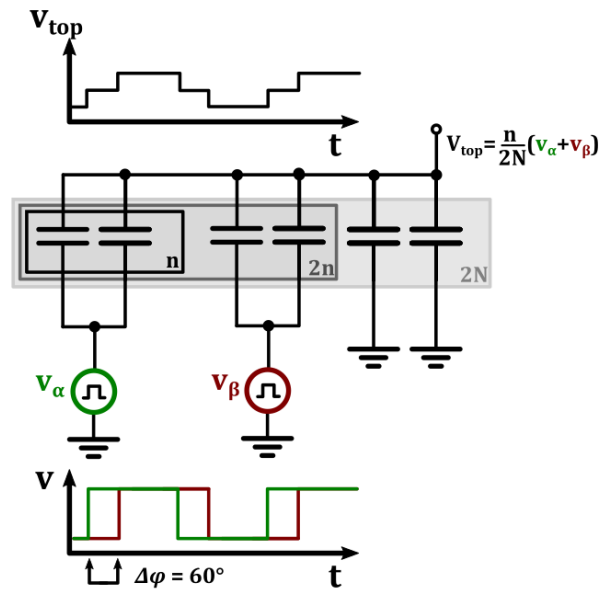


Figure 2.10: Schematic of Single-Array Dual-Phase Waveform Combination.

In such a scheme, increasing n by 1 would mean activating two units, one at φ_α and one at φ_β . Thus, one bit of resolution is lost, requiring double the number of units to compensate. A multi-phase array of capacitance C_{array} can be considered as having $2N$ units, with code n having a maximum value of N when all $2N$ units are active.

More importantly, using multiple phases under a single array can severely impact efficiency. In order to calculate the equivalent input capacitance, and thus the PAE of a multi-phase array, an edge-by-edge analysis can be performed.

Figure 2.11 illustrates the equivalent circuits at four moments in the cycle; the rising and falling edges of both waveforms v_α and v_β . The array is segmented into three equivalent capacitors:

$$C_\alpha = C_\beta = \frac{n}{2N} C_{array}, \quad C_{off} = C_{array} - (C_\alpha + C_\beta) = \frac{N-n}{N} C_{array}$$

C_α and C_β refer to the total capacitance of active SCPA units at phases φ_α and φ_β , respectively. C_{off} is the remaining capacitance, whose bottom plate remains grounded throughout the cycle. The inductor connecting the array top-plate to the load acts as an open during the waveform edges, thus it can be considered that the current supplied by the source is exclusively used to charge the capacitors.

Prior to edge 1, the rising edge of waveform v_α , all three capacitors are grounded through their bottom plates. The top-plate voltage is also equal to ground; thus, no charge exists within any of the capacitors.

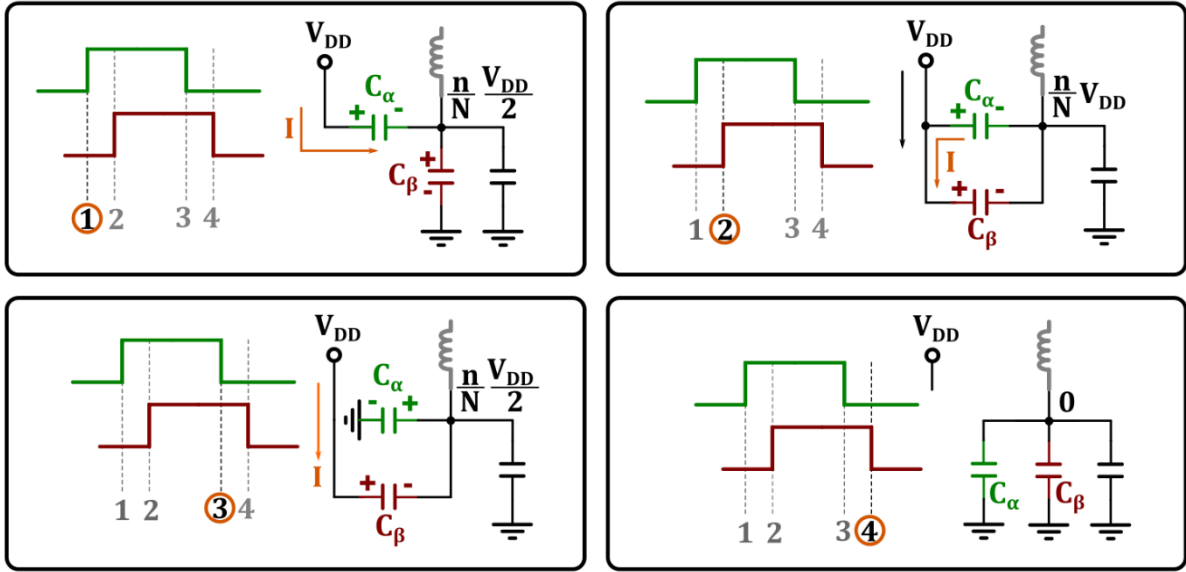


Figure 2.11: Equivalent Circuits for Calculation of Input Capacitance in Dual-Waveform Configuration.

During edge 1, the bottom plate of C_α is brought to V_{DD} . Through capacitive division, the shared array top-plate acquires a voltage of:

$$V_{top,1} = \frac{n}{2N} V_{DD} \quad (2.13)$$

The charge provided by the supply during edge 1 is equal to the change in charge for C_α :

$$\Delta Q_1 = \Delta Q_{\alpha,1} = Q_{\alpha,1} = C_\alpha (V_{DD} - V_{top,1}) = C_\alpha \left(1 - \frac{n}{2N}\right) V_{DD} = \frac{n}{2N} C_{array} \left(1 - \frac{n}{2N}\right) V_{DD}$$

$$\Delta Q_1 = \left(N - \frac{n}{2}\right) \frac{n}{2N^2} C_{array} V_{DD} \quad (2.14)$$

The charge within C_β can also be calculated as:

$$Q_{\beta,1} = \left(-\frac{n}{2}\right) \frac{n}{2N^2} C_{array} V_{DD} \quad (2.15)$$

In this case, the charge is considered to have a negative value, as the bottom plate of C_β is at a lower voltage than its top plate.

During edge 2, the rising edge of waveform v_β , the bottom plate of C_β is brought from ground to V_{DD} , whereas C_α maintains its prior position. The top-plate voltage becomes equal to:

$$V_{top,2} = \frac{n}{N} V_{DD} \quad (2.16)$$

The two capacitors now have charge equal to:

$$Q_{\alpha,2} = Q_{\beta,2} = C_a(V_{DD} - V_{top,2}) = (N - n) \frac{n}{2N^2} C_{array} V_{DD} \quad (2.17)$$

The charge provided by the supply is equal to the change in charge for both capacitors:

$$\Delta Q_2 = \Delta Q_{\alpha,2} + \Delta Q_{\beta,2} = Q_{\alpha,2} - Q_{\alpha,1} + Q_{\beta,2} - Q_{\beta,1} \quad (2.18)$$

$$\Delta Q_2 = (N - n) \frac{n}{2N^2} C_{array} V_{DD} \quad (2.19)$$

Similarly, during edge 3, the falling edge of waveform v_α , the bottom plate of C_α is moved to ground, with C_β remaining connected to V_{DD} . The top-plate voltage is now equal to:

$$V_{top,3} = \frac{n}{2N} V_{DD} \quad (2.20)$$

This is a similar state as that of edge 1, only with C_α and C_β swapping roles. The supply provides charge equal to the change in charge for C_β . It can be calculated as:

$$\Delta Q_3 = \Delta Q_{\beta,3} = Q_{\beta,3} - Q_{\beta,2}$$

$$\Delta Q_3 = \left(\frac{n}{2}\right) \frac{n}{2N^2} C_{array} V_{DD} \quad (2.21)$$

During edge 4, all capacitors are grounded. No charge is provided by the supply, and all the existing charge is lost.

$$\Delta Q_4 = 0 \quad (2.22)$$

The total charge drained for the supply per cycle is:

$$Q_{in} = \sum_{i=1}^4 \Delta Q_i = (2N - n) \frac{n}{N^2} \frac{C_{array}}{2} V_{DD} \quad (2.23)$$

The equivalent input capacitance for this configuration can then be calculated as:

$$C_{in} = (2N - n) \frac{n}{N^2} \frac{C_{array}}{2} \quad (2.24)$$

In contrast with the input capacitance of a typical SCPA configuration, as displayed in equation 2.6, this expression cannot be reduced to zero for $n = N$. Even when all units are active, the capacitors still need to be charged due to the difference in timing between the waveform edges.

The same analysis can be performed for any switched-capacitor configuration in which multiple waveforms are used to drive different units of the same array. This includes the aforementioned *multi-duty-cycle tuning* method, in which two waveforms of duty cycles $d_\alpha = 1/3$ and $d_\beta = 2/3$ are used. In that case, the resulting input capacitance C_{in} is equal to that calculated for the phase tuning configuration.

Seeing as both methods also provide the same factor of $\cos(\pi/6)$ for the amplitude of their output fundamental components, it follows that the ideal power-added efficiency for both configurations will also be equal. Considering the following values, PAE can be calculated:

$$PAE_{ideal} = \frac{P_{out}}{P_{out} + P_{SC}}, \quad P_{out} = \frac{2}{\pi^2} \frac{n^2}{N^2} \frac{V_{DD}^2}{R_L} \cos^2\left(\frac{\pi}{6}\right), \quad P_{SC} = C_{in} V_{DD}^2 f_{LO}, \quad C_{in} = (2N - n) \frac{n}{N^2} \frac{C_{array}}{2}$$

$$PAE_{ideal} = \frac{1}{1 + \frac{(2N - n)\pi^2 R_L C_{array} f_{LO}}{3n}} \quad (2.25)$$

Figure 2.12 presents a plot of this expression, along with plots for the power-added efficiency of a typical configuration (equation 2.8), and a single-waveform $d = 1/3$ configuration (equation 2.11). It can be seen that whereas the latter two are able to reach 100% ideal PAE at maximum input code $n = N$, this is not the case for the dual-waveform configurations discussed above.

Since both dual-waveform configurations result in the same output waveform and have identical ideal PAE, they can be considered equivalent methods of harmonic suppression. Selecting either the multi-phase or multi-duty-cycle solution thus becomes a matter of implementation.

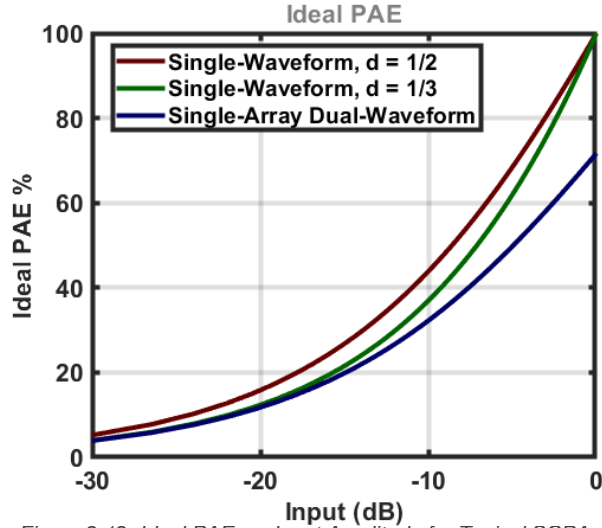


Figure 2.12: Ideal PAE vs. Input Amplitude for Typical SCPA, Dual-Waveform, and Single-Waveform Duty Cycle Tuning Configurations.

Generating two waveforms with $\Delta\varphi = 60^\circ$ may be more or less preferable to the alternative of creating waveforms with duty cycles of $d_\alpha = 1/3$ and $d_\beta = 2/3$, depending on the system. A combination of different phases and amplitudes for multiple waveforms can also be used to suppress multiple harmonics [7].

It is important to note that the dual-waveform harmonic suppression technique is rather sensitive to imperfections in the duty cycle, phase difference, and amplitude of the waveforms. Figure 2.13 displays plots of the resulting second- and third-order harmonics for the multi-phase variant. Input v_α has an amplitude of $A = 1$, duty cycle $d = 1/2$, and phase $\varphi = 0^\circ$. A sweep is performed for each of the respective parameters of input waveform v_β .

As can be seen from the figures, any deviation from $A = 1$, $d = 1/2$, and $\Delta\varphi = 60^\circ$ can severely impact the degree to which the harmonics are suppressed. Thus, it is vital that the waveforms retain values as close to perfect as possible for these parameters.

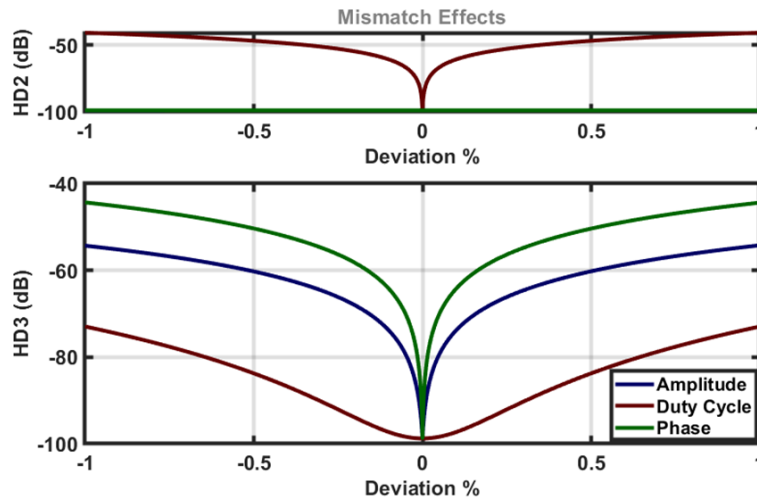


Figure 2.13: HD2 and HD3 vs. Deviations in Duty Cycle, Phase, and Amplitude.

2.5: Explicit Cancellation

So far, duty cycle and phase have been discussed as parameters used for harmonic suppression. Another solution is found by adding together waveforms of different frequency. In this *explicit cancellation* scheme seen in figure 2.14, waveforms v_α and v_β are combined, with v_β having three times the frequency and one-third the amplitude of v_α .

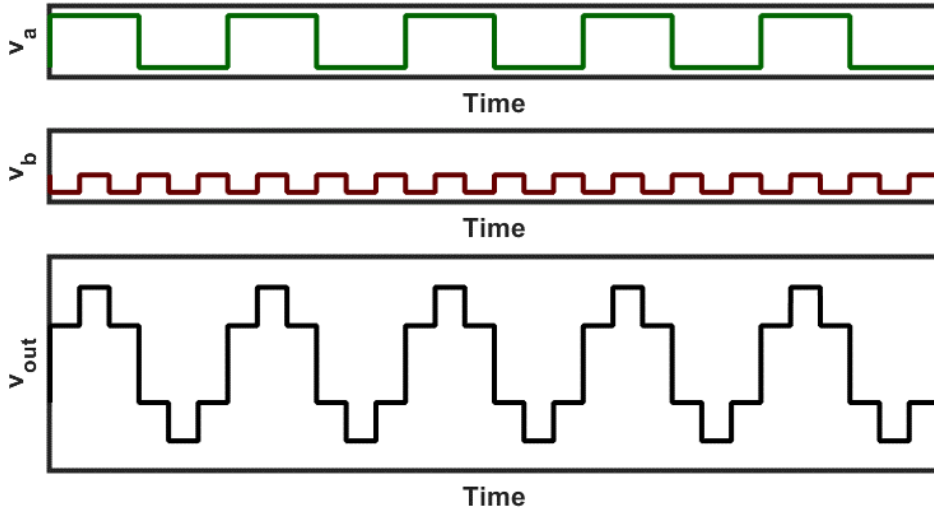


Figure 2.14: Component Waveforms and Output Waveform for Explicit Cancellation Configuration.

The waveforms have a phase difference of $\Delta\phi = 180^\circ$. The fundamental component of v_β is thus equal and opposite to the third harmonic of v_α , effectively cancelling it out. The amplitudes for each waveform are set by driving 3/4 of the array with v_α , while the remaining 1/4 is driven with v_β .

Subjecting this configuration to the same analysis as performed above reveals a severe reduction in PAE, even when compared to other multi-waveform schemes. As v_β contains no spectral components at the carrier frequency, the amplitude of the fundamental of the resulting waveform is limited to 3/4 of its maximum potential value in a typical SCPA configuration.

Moreover, as v_β is at three times the frequency of the carrier, each carrier period will contain many instances of charging and discharging of the array, increasing the equivalent input capacitance C_{in} , and thus the dynamic power consumption P_{SC} compared to the configurations examined prior. The equivalent input capacitance for the explicit cancellation scheme is calculated as:

$$C_{in} = \left(2N - \frac{n}{2}\right) \frac{n}{N^2} \frac{3C_{array}}{4} \quad (2.26)$$

With a resulting ideal power-added efficiency of:

$$PAE_{ideal} = \frac{1}{1 + \frac{\left(2N - \frac{n}{2}\right) \pi^2 R_L C_{array} f_{LO}}{3n/2}} \quad (2.27)$$

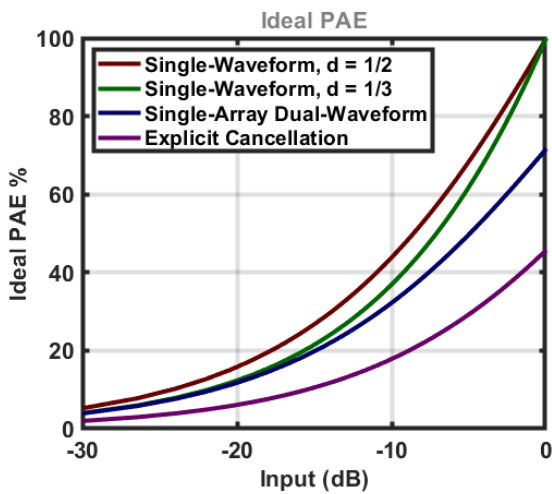


Figure 2.15: Ideal PAE vs. Input Amplitude for Typical SCPA, Dual-Waveform, Single-Waveform Duty Cycle Tuning, and Explicit Cancellation Configurations.

A plot of the above expression can be seen in figure 2.15, alongside plots for ideal power-added efficiency of the following configurations: standard SCPA, single-waveform duty-cycle tuning, and dual-waveform duty-cycle or phase tuning. These correspond to equations 3.8, 3.11, and 3.25, respectively.

After examining this plot, it is obvious that the explicit cancellation method is largely inferior in terms of efficiency to either of the two dual-waveform harmonic suppression methods discussed previously. Due to this fact, coupled with the complexity required to create and maintain the related input waveforms, this method was deemed unworthy of pursuing further for this project.

2.6: Dual-Array

Thus far, a method of waveform summation has been examined, in which a single switched-capacitor array is segmented into parts, with each driven by a different waveform. This is equivalent to multiple conventional sub-arrays being connected in parallel at their common top plate node. Although convenient, this configuration introduces a significant reduction in power-added efficiency. As discussed above, the uninhibited exchange of charge between sub-arrays results in increased dynamic power consumption with no added benefit to output power.

In order to isolate the SCPA arrays, inductance can be used to create an open during waveform edges. Each sub-array can be considered individually as a conventional array with its own resonant inductor. Waveform summation can be simply achieved by utilizing a differential configuration, wherein the load is connected to both arrays in series as seen in figure 2.16. In this topology, waveform v_β is inverted, and now has a phase difference of $\Delta\phi = 240^\circ$ to v_α . Output voltage v_{out} is generated across the load, and can be calculated as:

$$v_{out}(t) = \frac{n}{N} (v_\alpha(t) - v_\beta(t)) \quad (2.28)$$

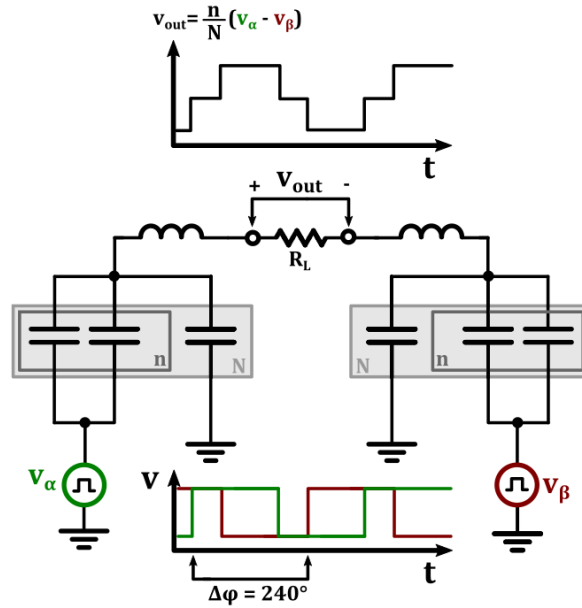


Figure 2.16: Schematic of Dual-Array Waveform Combination.

Considering two sub-arrays, each with capacitance equal to C_{array} , in parallel with respect to the supply, the resulting total input capacitance is calculated as:

$$C_{in} = 2 \frac{n(N-n)}{N^2} C_{array} \quad (2.29)$$

Seeing as the two arrays are independent, this expression is predictably double that of equation 2.6, and is equal to zero for maximum code $n = N$. In the case of a dual-waveform phase-tuning harmonic suppression configuration, the output power is calculated as:

$$P_{out} = \frac{8}{\pi^2} \frac{n^2 V_{DD}^2}{N^2 R_L} \cos^2\left(\frac{\pi}{6}\right) \quad (2.30)$$

Since the two arrays are connected in series with respect to the load, the maximum swing becomes $2 V_{DD}$, effectively quadrupling output power compared to the previously examined single-array configuration. Ideal power-added efficiency is calculated as:

$$PAE_{ideal} = \frac{1}{1 + \frac{(N-n)\pi^2 R_L C_{array} f_{LO}}{3n}} \quad (2.31)$$

It is important to note that the dual-array configuration also utilizes a separate inductor for each array, requiring double the total inductance compared to the single-array solution. If total inductance were to be kept equal between the two methods, array capacitance would need to be doubled for each of the two arrays in the case of the dual-

array configuration. Doing so would bring equation 2.31 in line with equation 2.11 of the single-array, single-waveform harmonic suppression solution.

Figure 2.17 presents a plot of the expression of equation 2.31, alongside plots of equations 3.8 and 3.25, corresponding to the single-array configurations of a typical SCPA and dual-waveform harmonic suppression, respectively. In order to provide a fair comparison, array capacitance was adjusted to equalize total inductance for all cases.

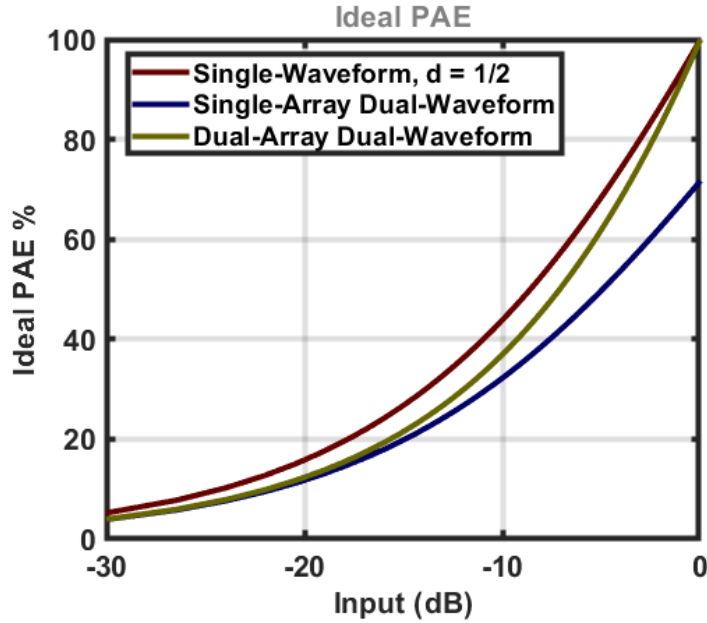


Figure 2.17: Ideal PAE vs. Input Amplitude for Typical SCPA, Dual-Array Dual-Waveform, and Single-Array Dual-Waveform Configurations.

As evident from this plot, a dual-array solution is preferable to a single-array dual-waveform solution as far as efficiency is concerned. Ideal PAE reaches 100% at maximum input amplitude since the two arrays are isolated and do not load each other during input waveform edges.

2.7: Transformer

Waveform combination can also be achieved by using an output transformer to transfer power from the arrays to the load. In this scenario, illustrated in figure 2.18, the primary coil L_p of the transformer is connected directly to the array top-plates. L_p is magnetically coupled to L_s , the secondary coil, through magnetic coupling constant k . The inductance of the transformer is responsible for resonating with the array capacitance at the fundamental frequency, replacing the inductors used for the aforementioned series configuration.

Utilizing an output transformer introduces a number of advantages. A differential pair of SCPA arrays may be connected to a single-ended load, such as an RF antenna. Output power can be controlled by tuning the transformer turn ratio N so as to impose a higher or lower amplitude upon a fixed load, according to the application specification. Moreover, connecting multiple transformers allows for the summation of any number of waveforms, if required [8].

On-chip transformers do, however, come at the cost of a lower quality factor Q than their simpler inductor alternatives [?]. Since a higher ohmic resistance is present for the same amount of inductance, more losses and thus lower efficiency is expected for such a configuration. Regardless, the use of a transformer as a balun is a necessity given the single-ended nature of the output load.

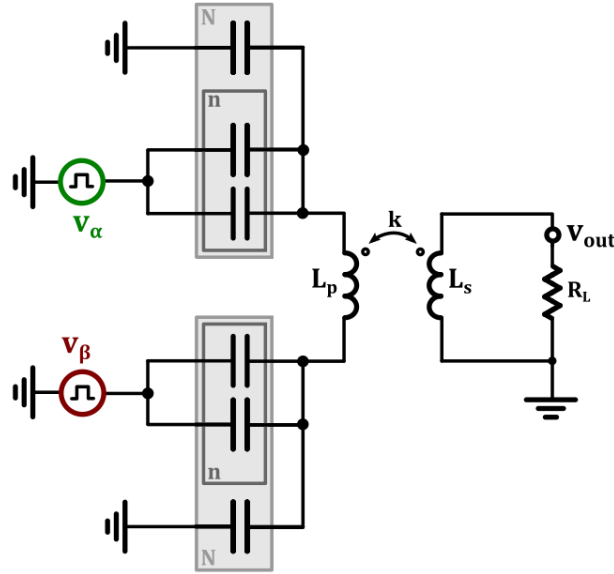


Figure 2.18: Schematic of Dual-Array Waveform Combination Using a Transformer.

2.7.1: Mirrored Waveforms

Mismatch between the amplitudes of the combined waveforms can also be catastrophic for harmonic suppression. This mismatch can occur as a result of asymmetry in the output transformer. Since a switched-capacitor array is connected to each of the primary terminals of the transformer, any disparity between transformer voltage gain parameters will lead to uneven voltage summation, and thus imperfect harmonic suppression as shown in figure 2.13.

Although no dedicated tuner is designed for the purposes of maintaining amplitude equality between the two waveforms, a decent solution is found in creating two additional, *mirrored* waveforms. Waveforms v'_α and v'_β are, in concept, inverted versions of v_α and v_β , respectively, with $v'_\alpha = -v_\alpha$ and $v'_\beta = -v_\beta$. These are used as inputs to a duplicate set of SCPA arrays, and are combined with the original two waveforms through a duplicate transformer.

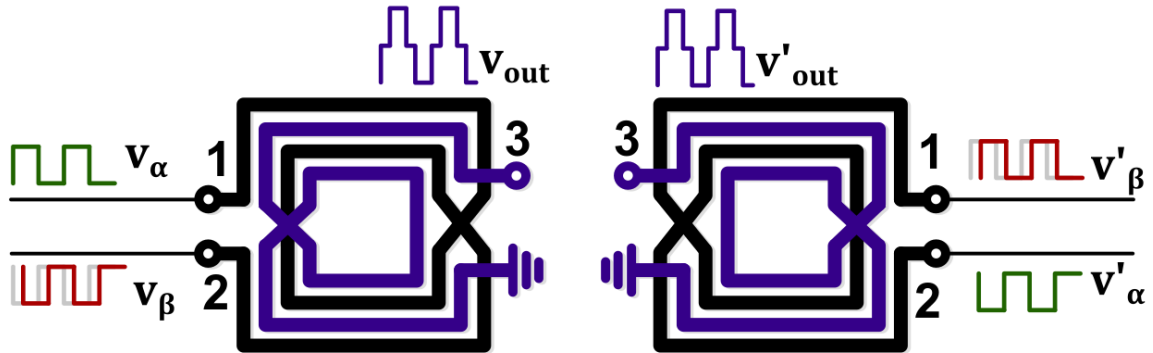


Figure 2.19: Illustration of "Mirrored Waveforms" Configuration.

However, each is connected to the transformer terminal opposite of that of its counterpart, thus minimizing any amplitude disparity resulting from asymmetry when all four are added together. Inputs v_α and v_β are differentially connected, with v_α as positive and v_β as negative, whereas v'_β is connected to the positive terminal of the duplicate transformer, with v'_α as negative. Figure 2.19 displays an illustration of this configuration. Assuming uneven voltage gains of $S_{13} \neq S_{23}$, the resulting waveform can be calculated as:

$$v_{out} + v'_{out} = S_{13}v_\alpha - S_{23}v_\beta + S_{13}v'_\beta - S_{23}v'_\alpha = (S_{13} + S_{23})(v_\alpha - v_\beta) \quad (2.32)$$

For comparison, using only v_α and v_β in a single-transformer configuration would result in uneven addition of $v_{out} = S_{13}v_\alpha - S_{23}v_\beta$, with a predictable impact on harmonic suppression.

The ideal phase values for all four waveforms are:

$$\varphi_\alpha = 0^\circ, \quad \varphi_\beta = 240^\circ, \quad \varphi'_\alpha = 180^\circ, \quad \varphi'_\beta = 60^\circ \quad (2.33)$$

2.8: Design Proposal

Considering the discussion and analysis presented throughout this chapter, a design proposal can be made. The goal is to create a switched-capacitor power amplifier which achieves suppression of the second and third harmonic, while maintaining high power-added efficiency at power back-off.

Harmonic suppression is realized by employing the dual-waveform method described previously, in which two waveforms of different characteristics are combined in order to create an output signal devoid of second- and third-order harmonic distortion. The $d = 1/2$, $\Delta\varphi = 60^\circ$ variant is preferred over its alternative, due to the relative ease of implementation. Considering a differential square-wave local oscillator input waveform of $d = 1/2$, the design challenge is relegated to maintaining this duty cycle value, while applying a phase shift to one of the two LO waveforms. The alternative method of $d_1 = 1/3$, $d_2 = 2/3$ may have been preferable in the case of a sinusoidal LO input.

In order to apply the phase shift, a precision phase tuner must be designed. As can be seen from figure 2.13, even slight deviations from $\Delta\varphi = 60^\circ$ can prove disastrous for the quality of $HD3$ suppression. Thus, the phase tuner must be able to be calibrated to a value as close to 60° as possible. Similarly, a duty cycle tuner must also be employed in order to tune the waveforms to $d = 1/2$.

Although the calibration of these tuners can be performed manually in order to counteract process variation, maintaining perfect waveforms is also important in the presence of fluctuations in temperature, supply voltage, and LO frequency. Thus, a feedback loop [9] is implemented, which allows the tuners to be calibrated in real time. By monitoring the waveforms and performing comparisons against references, the system will be able to direct their phase difference and duty cycles towards the desired values prior to transmission.

For additional resilience against amplitude mismatch, the “mirrored waveforms” technique will be implemented. Although v'_α and v'_β can be generated by using an inverter to apply a phase shift of 180° to v_α and v_β , generating all four waveforms independently by using separate duty and phase tuners will provide the most control over their parameters. v_β and v'_β are generated by applying a 60° phase shift to the negative and positive waveform of the differential LO, respectively. The four arrays are matched to the load through the two transformers.

Figure 2.20 presents a block diagram of the proposed system.

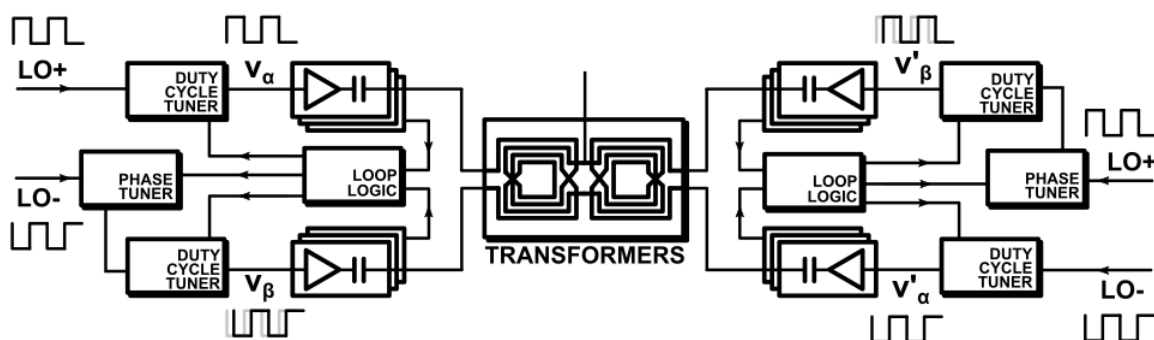


Figure 2.20: Block Diagram of Proposed System.

3: DESIGN

3.1: Matching

Proper impedance matching of the SCPA array to the load presents a challenge. An equivalent circuit model of a transformer is presented in the schematic of figure 3.1. The model consists of two inductors L_1 and L_2 , one in series with the primary port and one shunt. The shunt inductor is connected in parallel to an ideal transformer, whose secondary port comprises the secondary port of the model itself. The equivalent inductor inductances are related to k , the magnetic coupling coefficient. The turn ratio T of the ideal transformer is calculated as $T = k\sqrt{L_p/L_s}$, wherein L_p and L_s refer to the primary and secondary inductances of the transformer [10].

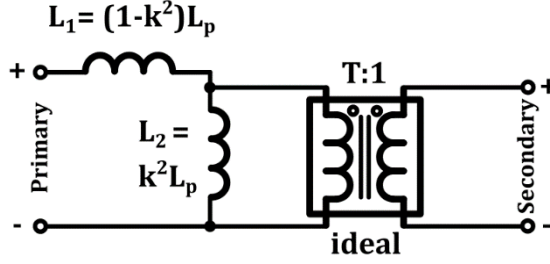


Figure 3.1: Equivalent Circuit Model of Transformer. Adapted from [7].

Figure 3.2 displays a simplified schematic of a switched-capacitor array connected to a load through a transformer. A capacitor is also added in parallel to the load, so as to resonate out the shunt inductor L_2 present in the equivalent transformer model. This creates an open at the fundamental frequency, only retaining a single path from the array towards the load through series inductor L_1 . An appropriate value for array capacitance is then selected, which resonates with the series inductor, thus creating a short at the fundamental frequency as discussed earlier. For given values of k, L_p, L_s , the capacitances can be calculated:

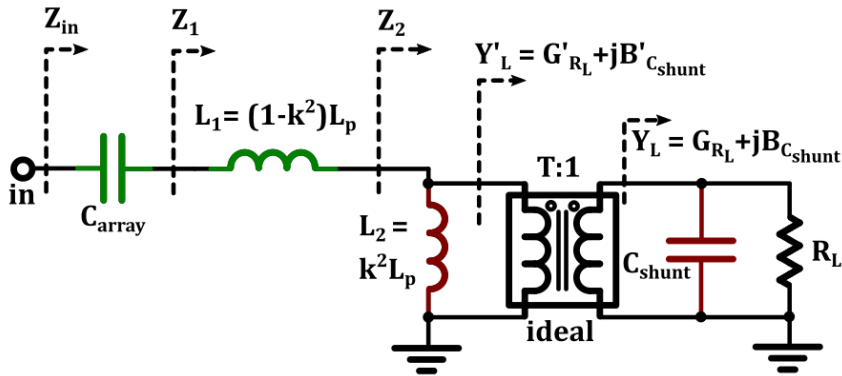


Figure 3.2: Illustration of Impedance Matching Using a Transformer and Shunt Capacitor.

Admittance $Y_L = G_{R_L} + jB_{C_{shunt}}$ is created by the combination of R_L and C_{shunt} . Through the transformer, this admittance is seen by the primary as $Y'_L = Y_L/T^2$, with susceptance $B'_{shunt} = j\omega C_{shunt}/T^2$, implying an equivalent capacitance of $C'_{shunt} = C_{shunt}/T^2$. This capacitance needs to resonate with $L_2 = k^2L_p$ in order to create a purely ohmic impedance at Z_2 for ω_{LO} .

For the shunt capacitance connected to the load:

$$C'_{shunt} = \frac{C_{shunt}}{T^2} = \frac{1}{\omega_{LO}^2 k^2 L_p} \rightarrow C_{shunt} = k^2 \frac{L_p}{L_s} \frac{1}{\omega_{LO}^2 k^2 L_p} \rightarrow C_{shunt} = \frac{1}{\omega_{LO}^2 L_s} \quad (3.1)$$

Shunt capacitor C_{shunt} thus directly resonates with L_s , independently of transformer parameters. Z_2 is now purely ohmic and equal to $Z_2 = T^2 R_L$ at ω_{LO} .

Considering the addition of series inductance $L_1 = (1 - k^2) L_p$, impedance Z_1 at ω_{LO} becomes:

$$Z_1 = Z_2 + j\omega_{LO}L_1 = T^2R_L + j\omega_{LO}(1 - k^2)L_p$$

Capacitor C_{array} must now resonate with L_1 in order to create an ohmic Z_{in} at ω_{LO} :

$$C_{array} = \frac{1}{\omega_{LO}^2 (1 - k^2)L_p} \quad (3.2)$$

Input impedance Z_{in} at ω_{LO} will now be equal to:

$$Z_{in} = T^2R_L = k^2 \frac{L_p}{L_s} R_L \quad (3.3)$$

Minimizing array capacitance is an important factor in increasing power back-off efficiency. As can be seen from equation 2.31, low values of C_{array} result in higher values of ideal PAE for $n < N$. Since charge is deposited into the array each cycle, a lower array capacitance would bring about a reduction in dynamic power consumption.

This can be achieved in a number of ways. It is obvious from equation 3.2 that higher values of L_p , and thus L_1 , will resonate with smaller values of C_{array} . Another approach would be to minimize the magnetic coupling coefficient k ; as k is reduced, the series inductance L_1 of the transformer model comprises a larger portion of the primary inductance L_p .

3.1.1: “No Shunt” Matching

An alternate method for impedance matching of the array using a transformer would be to completely forgo the shunt capacitor and instead use the array capacitance as the sole capacitive element present in the matching network, as seen in figure 3.3. C_{array} now resonates with the entire reactance presented by L_1 , L_2 , and R_L . The appropriate value for array capacitance in this configuration can be calculated as:

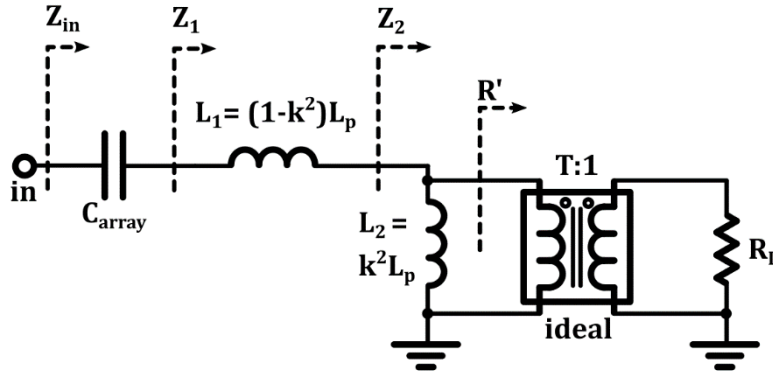


Figure 3.3: Illustration of Impedance Matching Using a Transformer and No Shunt Capacitor.

$$C_{array} = \frac{1}{\omega_{LO}^2 L_p} \frac{R_L^2 + \omega^2 L_s^2}{R_L^2 + (1 - k^2)\omega^2 L_s^2} \quad (3.4)$$

This can be calculated as follows:

Load resistance R_L is transformed through the ideal transformer into $R' = T^2R_L$. Impedance Z_2 is calculated as:

$$Z_2 = j\omega L_2 \parallel R' = \frac{\omega^2 L_2^2 R' + j\omega L_2 R'^2}{R'^2 + \omega^2 L_2^2}$$

With the addition of series inductor L_1 , impedance Z_1 is calculated as:

$$Z_1 = j\omega L_1 + Z_2 = \frac{1}{R'^2 + \omega^2 L_2^2} [\omega^2 L_2^2 R' + j\omega(L_2 R'^2 + L_1 R'^2 + \omega^2 L_1 L_2^2)]$$

Array capacitor C_{array} needs to cancel out the reactive part of Z_1 in order to create a purely ohmic impedance at Z_{in} for $\omega = \omega_{LO}$.

$$Z_{in} = Z_1 + \frac{1}{j\omega C_{array}}, \quad X_{in} = X_1 - \frac{1}{\omega C_{array}} = 0 \rightarrow C_{array} = \frac{1}{\omega X_1}$$

$$C_{array} = \frac{1}{\omega} \frac{R'^2 + \omega^2 L_2^2}{\omega(L_2 R'^2 + L_1 R'^2 + \omega^2 L_1 L_2^2)}$$

Substituting the following:

$$L_1 = (1 - k^2)L_p, \quad L_2 = k^2 L_p, \quad R' = T^2 R_L, \quad T = k \sqrt{L_p/L_s}, \quad \omega = \omega_{LO}$$

The equation for C_{array} can be formulated:

$$C_{array} = \frac{1}{\omega_{LO}^2 L_p} \frac{R_L^2 + \omega_{LO}^2 L_s^2}{R_L^2 + (1 - k^2)\omega_{LO}^2 L_s^2}$$

Input impedance Z_{in} at ω_{LO} is now equal to:

$$Z_{in} = \frac{\omega_{LO}^2 L_2^2 R'}{R'^2 + \omega_{LO}^2 L_2^2} = k^2 \frac{\omega_{LO}^2 L_p L_s}{R_L^2 + \omega_{LO}^2 L_s^2} R_L \quad (3.5)$$

3.1.2: Comparison of Matching Techniques

A comparison can be made between the “shunt” and “no shunt” matching methods mentioned above, referring to equations 4.2 and 4.4, respectively. When comparing these two equations, it can be seen that the “no shunt” method produces smaller values for array capacitance compared to the alternative, albeit with added design complexity.

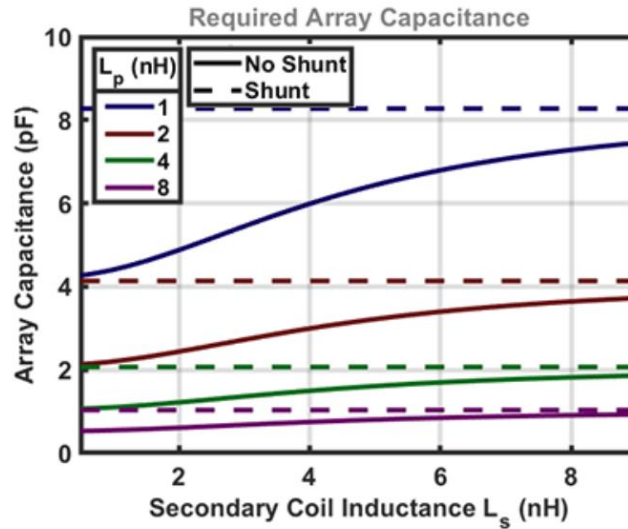


Figure 3.4: Plot of Required Array Capacitance vs. L_p and L_s for “Shunt” and “No Shunt” Matching Configurations.

Figure 3.4 displays the resulting C_{array} values for a magnetic coupling coefficient of $k = 0.7$, matched for a frequency of $f_{LO} = 2.45$ GHz. The independent parameters in this example are the primary and secondary coil inductances L_p and L_s . This plot reveals a significant decrease in C_{array} in the case of the “no shunt” method, meaning not only a smaller area, but a potential advantage in terms of power back-off efficiency, as per equation 2.31. It also becomes clear that lower values of L_s reduce C_{array} further, reaching as low as half the capacitance values required for shunt-capacitor matching.

Although not pictured in the figure, C_{array} also becomes less dependent on k as L_s is decreased in “no shunt” configuration. This could also mean more resilience to process variations of the transformer, as well as the capacitors themselves.

In order to further compare the two methods, their power output can also be computed. Adapting equation 2.7:

$$P_{out} = \frac{2}{\pi^2} \frac{n^2 V_{DD}^2}{N^2 R_{in}} \quad (3.6)$$

In this case R_{in} refers to the ohmic input impedance of the array, as in the real part of Z_{in} , seen in the schematics of figures 3.2 and 3.3. In previous examples where a simple series resonator was used, the input impedance was equal to R_L . However, this is not the case when a transformer is used. As calculated in equations 4.3 and 4.5 for the “shunt” and “no shunt” methods respectively, R_{in} can vary based on design parameters:

$$R_{in,s} = k^2 \frac{L_p}{L_s} R_L, \quad R_{in,ns} = k^2 \frac{\omega_{LO}^2 L_p L_s}{R_L^2 + \omega_{LO}^2 L_s^2} R_L \quad (3.7)$$

Figure 3.5 displays a plot for R_{in} under both matching configurations for $k = 0.7$, $f_{LO} = 2.45$ GHz. As L_s decreases, the input resistance $R_{in,s}$ for the “shunt” topology will increase. This is not the case for $R_{in,ns}$ of the “no shunt” topology, wherein the resistance will start to decrease after reaching a peak at $L_s = R_L/\omega_{LO}$. Very small values of $R_{in,ns}$ can thus be achieved by minimizing L_s using this method. Considering the relation of equation 3.6. It can be seen that the “no shunt” matching configuration could provide increased output power for decreasing values of L_s , thus further reducing the area cost for high-power designs.

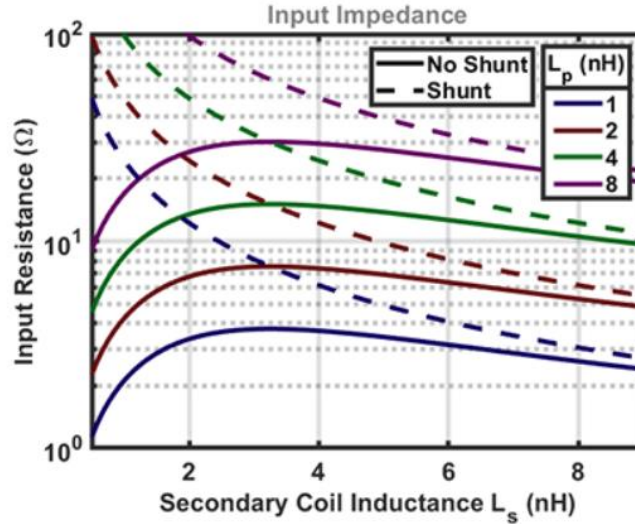


Figure 3.5: Plot of Input Resistance vs. L_p and L_s for “Shunt” and “No Shunt” Matching Configurations.

3.1.3: “No Shunt” Regions

It can also be observed that a switched-capacitor PA matched in “shunt” configuration behaves similar to a voltage source, since its output power is reduced as T is increased. For the voltage V_s across the secondary coil, where the load is located:

$$V_s = \frac{V_p}{T}$$

Where V_p represents the voltage across the primary coil, provided by the voltage source. The RMS power provided by that voltage source to a load R_L through a transformer can then be calculated as:

$$P = \frac{V_s^2}{R_L} = \frac{V_p^2}{R_L T^2}$$

Since, from equation 3.7, $R_{in,s} = T^2 R_L$ for a “shunt” matched SCPA, this configuration can be considered as operating in “voltage mode”.

A similar behavior is also observed for the “no shunt” technique, but only for low values of T . Indeed, as $L_s \rightarrow \infty$, the two methods become equivalent; L_s dominates the expressions for C_{array} and R_{in} of the “no shunt”

configuration, bringing them in line with their “shunt” counterparts. The expression for C_{shunt} (equation 3.1) also reaches zero. However, for higher values of T , as $L_s \rightarrow 0$, the following calculations can be made for a “no shunt” matched PA:

$$R_{in,ns} \rightarrow \frac{k^2 \omega_{LO}^2 L_p L_s}{R_L}, \quad P_{out} \rightarrow \frac{V_{DD}^2}{2 \frac{k^2 \omega_{LO}^2 L_p L_s}{R_L}}, \quad P_{out} \rightarrow \frac{V_{DD}^2}{2 k^4 \omega_{LO}^2 L_p^2} T^2 R_L \quad (3.8)$$

This relation between P_{out} , T , and R_L is similar to that of $P = T^2 I_p R_L$, describing the power provided to a load by a current source through a transformer. Therefore, a “no shunt” matched SCPA connected to a high- T transformer can be hesitantly considered as operating in “quasi-current mode”. A plateau exists for $L_s = R_L / \omega_{LO}$ between the “quasi-voltage mode” and “quasi-current mode” regions. For this example, it is found for $L_s = 3.25 \text{ nH}$.

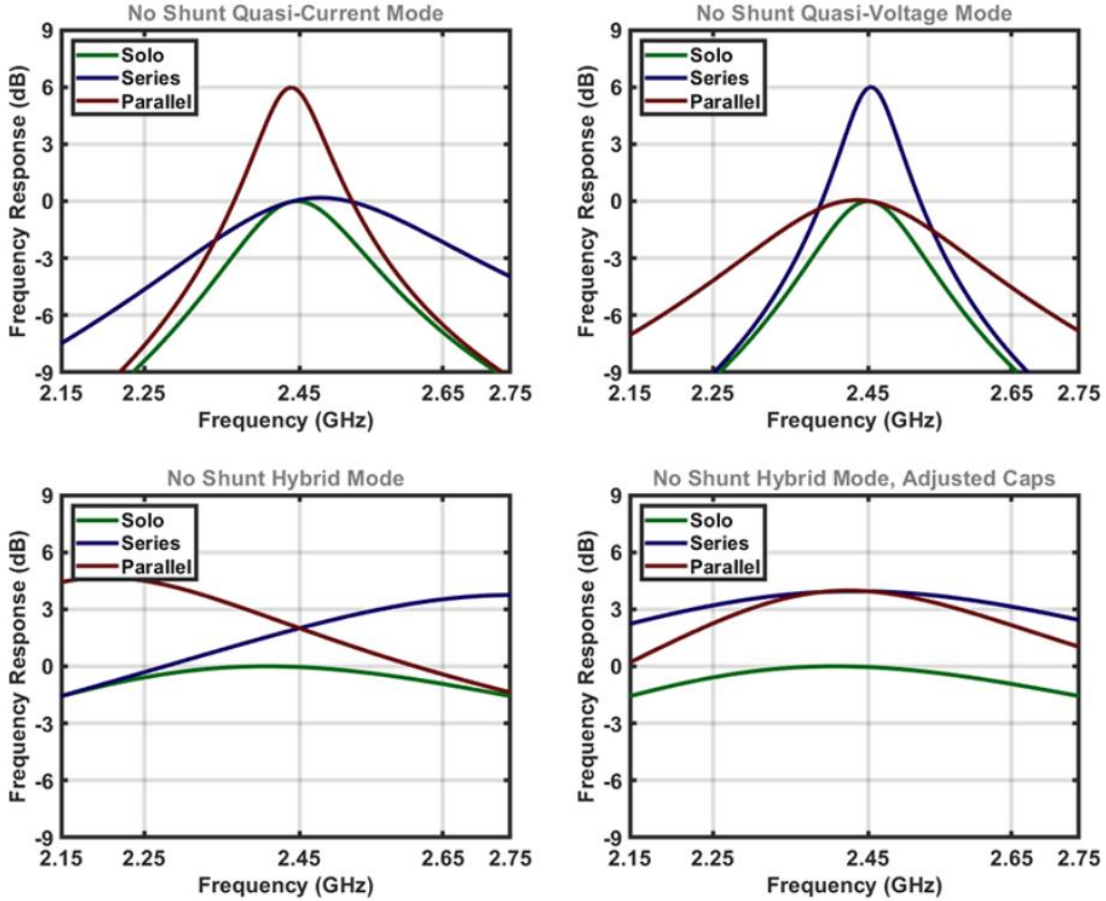


Figure 3.6: Frequency Responses of “No Shunt”-Matched Array Combination Configurations, for Different L_s .

In order to confirm the existence of these two distinct regions, a simulation was performed. In this example, an SCPA array is set-up in “no shunt” configuration, and connected to a load through a transformer of a specific L_s . The same set-up is then duplicated, and the secondary coils of their transformers are connected together; in parallel and then in series. The frequency response of all three cases is then simulated for different values of L_s and displayed in the plots of figure 3.7.

For low values of L_s , power combination takes place when the transformers are connected in parallel, whereas no extra power is gained when connecting them in series, verifying “quasi-current mode” operation. Output power is, however, not exactly quadrupled in the case of $L_s = 500 \text{ pH}$; as expected, a gain of $+6.02 \text{ dB}$ is only achieved for extremely high values of $L_s \rightarrow 0$.

The inverse is true for high L_s values, operating in “quasi-voltage mode”. Whereas parallel combination does not increase output power, series connection does. The PA behaves similar to a voltage source for $L_s = 50 \text{ nH}$. Again, a true gain of $+6.02 \text{ dB}$ is only found for $L_s \rightarrow \infty$.

Since, for this simulation, $\omega_{LO} = 2.45 \text{ GHz}$, $R_L = 50 \Omega$, a value for $L_p = 3.25 \text{ nH}$ represents the “hybrid-mode” plateau of $L_s = R_L/\omega_{LO}$ connecting the two modes. In this intermediate state, both series and parallel transformer connections provide the same gain of $\sim 2 \text{ dB}$ at the frequency of interest.

Adjusting array capacitance in both cases improves resonance, bringing the gain to $\sim 4 \text{ dB}$. In this scenario, a multiplier of 0.84 needed to be applied to the parallel-connected arrays, whereas a multiplier of 1.24 was necessary for the series-connected system. It is also important to note that in both “hybrid-mode” cases, especially in series connection, the resulting matching bandwidth is significantly higher than any other mode of operation.

3.1.4: Efficiency Comparison & Simulation

Since the equations for C_{array} and P_{out} are known for both matching topologies, a comparison of their ideal efficiencies can be made. Adapting equation 2.8 for $n = N/2$ results in the formula for PAE at half input amplitude, or at a power back-off of -6 dB :

$$PAE_{ideal|-6 \text{ dB}} = \frac{1}{1 + \frac{\pi^2 R_{in} C_{array} f_{LO}}{2}} \quad (3.9)$$

Parameters R_{in} and C_{array} are substituted from equations 4.2, 4.4, and 4.6. Figure 3.8 displays a plot for ideal back-off PAE, using the same conditions as above; $f_{LO} = 2.45 \text{ GHz}$ and $k = 0.7$.

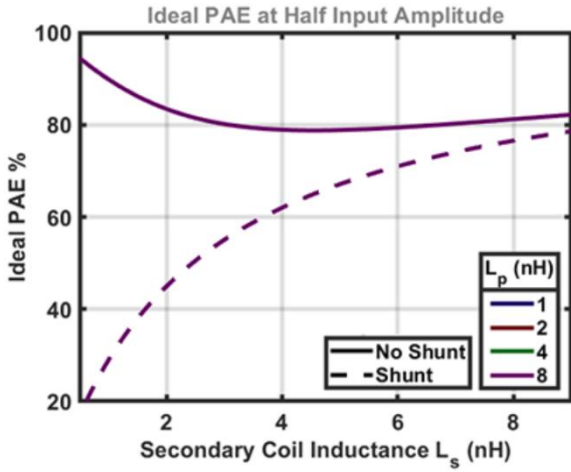


Figure 3.7: Plot of Ideal PAE at Half Input Amplitude vs. L_p and L_s for “Shunt” and “No Shunt” Matching Configurations.

From the plot of figure 3.8 it can be seen that the “no shunt” matching method achieves higher theoretical power back-off efficiency than its alternative regardless of the value of L_s , with that difference becoming a lot more considerable for lower values of L_s . This comes as no surprise considering the overall lower values for both R_{in} and C_{array} in the “no shunt” topology, and their relation to PAE from equation 3.9. Interestingly, changing L_p has no effect on the ideal calculated efficiency in either method.

The above calculations do not, however, take the ohmic impedance R_m of the inductors [11] into account. Considering the low R_{in} values achievable in “quasi-current mode” and the typically low quality-factors of on-chip monolithic transformers, ohmic losses can have significant effects on both output power and efficiency compared to theoretical calculations. For a given quality-factor Q of inductor L_m :

$$R_m = \frac{\omega_{LO} L_m}{Q} \quad (3.10)$$

In order to provide a more realistic comparison, a simulation was performed. In this setup, two SCPA arrays built with lossless switches are connected to a 50Ω load through lossy transformers with $Q = 10$ for both coils, which present ohmic resistances of R_p and R_s . Each array is matched using either “shunt” or “no shunt” matching as described previously. Parameters L_p and L_s are swept, and $k = 0.7$, $f_{LO} = 2.45 \text{ GHz}$, $V_{DD} = 0.8 \text{ V}$ are kept constant. Figures 3.8 and 3.9 display the resulting plots for output power and efficiency.

Figure 3.9 presents the plot of peak output power versus L_p and L_s . For a given L_p curve, this plot is similar to that of figure 3.6 except inverse as output power is increased for reduced input resistance. The “no shunt, quasi-current mode” region is visible for low L_s for “no shunt” matching, whereas the expected increase in output power for high L_s in “quasi-voltage mode” is suppressed here by the increase in inductor resistance R_s .

For the same reason, the “shunt” matched system displays a smoother power reduction as L_s is reduced compared to the ideal calculations of figure 3.6. Higher values of L_p predictably reduce output power as the corresponding ohmic loss R_p is increased.

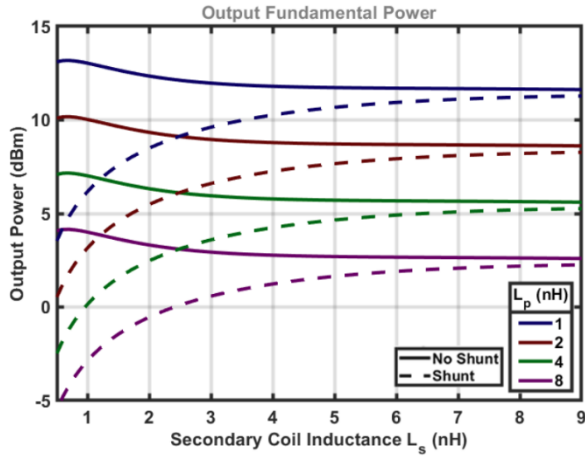


Figure 3.9: Simulated Plots of Peak Output Power vs. L_p and L_s for "Shunt" and "No Shunt" Matching Configurations.

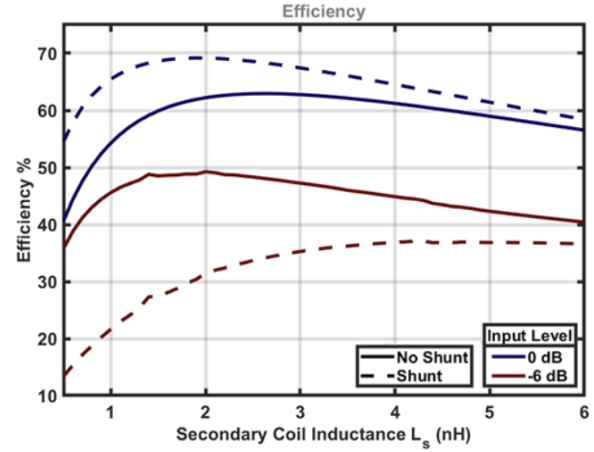


Figure 3.8: Simulated Plots of Efficiency at Peak Input Amplitude vs. L_p and L_s for "Shunt" and "No Shunt" Matching Configurations.

Peak output PAE, considered to be constant and equal to 100% in the ideal case, is now significantly reduced due to ohmic losses, and now displays a relation to L_s . Since C_{array} does not provide any dissipation at full input power, the plot of figure 3.10 is most comparable to that of figure 3.5, wherein the relation between input resistance R_{in} and L_s can be seen. In both matching configurations, lower R_{in} values correspond with low efficiency, as R_p and R_s overpower R_{in} . Interestingly, changing L_p does not influence efficiency, as $R_{in,s}$, $R_{in,ns}$, and R_p are all linear with L_p as can be seen from equations 4.6 and 4.10.

Figure 3.12 displays a simulated plot for efficiency at half input amplitude, or -6 dB back-off. Although efficiency suffers for extremely low $R_{in,ns}$ values deep into "quasi-current mode", the reduction in C_{array} provided by "no shunt" matching can compensate for this loss at low L_s values compared to the "shunt" matched alternative. Although peak efficiency is higher using "shunt" matching, "no shunt" provides an advantage in back-off efficiency.

A very clear trade-off exists between peak and back-off efficiency in the "shunt" case, as a small C_{array} requires large inductors with high values for R_m [11]. In contrast, "no shunt" matching can achieve maximum power back-off efficiency for low values of L_s in early "quasi-current mode", wherein peak input efficiency is still adequate. In every case, L_p serves only as a "knob" for adjusting output power to match design specifications with no effect on efficiency, provided that inductor quality-factor remains constant.

Moreover, the "no shunt" scenario allows access to higher output power levels at its region of maximum efficiency, as output power is significantly less dependent on L_s than its alternative. The nature of the "shunt" matching method means that high power levels are only available for extremely low values of L_p , which can lead to a very restricting implementation. If L_s is instead used to increase power, a significant drop in peak efficiency will come as a result.

In the case of "no shunt" matching, L_s can be reduced further in order to maximize efficiency at different levels of power back-off. However, the drawbacks seem to outweigh the benefits as small improvements in efficiency at extreme back-off beget a significant efficiency drop at higher power levels.

For a fair comparison of efficiency between the two methods, the same output power level needs to be achieved in both configurations. From figure 3.9, it can be seen that both methods are able to reach $P_{out} = 8$ dBm in multiple different combinations of L_p and L_s . For this test, L_s values that maximize efficiency at half input amplitude will be selected for both, and corresponding L_p values will be used to tune the power to 8 dBm. These values are:

"Shunt": $L_p = 1.90$ nH, $L_s = 5.50$ nH. "No shunt": $L_p = 2.83$ nH, $L_s = 1.70$ nH.

A test simulation under the same conditions was performed using these parameters. The resulting plots for fundamental output power, efficiency, and normalized frequency response are displayed in figure 3.14. In this case, peak efficiency is almost equal for both methods, however the "no shunt" configuration is still advantageous at every AM level.

Comparing the frequency responses, the "shunt" technique provides a much larger matching bandwidth, as well as increased filtering for the third harmonic and upwards. Considering the harmonic suppression capabilities for HD_2 and HD_3 of the system designed for this thesis, the frequency response of the "shunt" method seems to be preferable, as far-out harmonics not handled by waveform tuning will be adequately filtered.

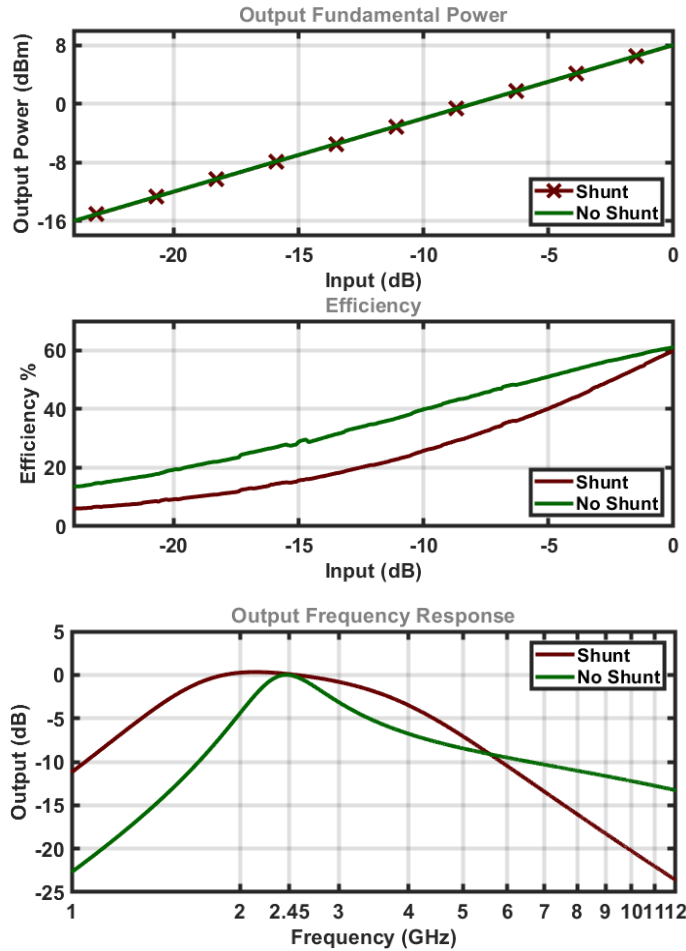


Figure 3.10: Simulated Plots for "Shunt" and "No Shunt" Matching Configurations Optimized for Efficiency at Half Input Amplitude. Top: Output Power vs. Input Amplitude. Middle: Efficiency vs. Input Amplitude. Bottom: Normalized Frequency Response.

However, the increase in power back-off efficiency provided by smaller array capacitances, as well as the increased design flexibility it offers when operating in "quasi-current mode", make "no shunt" matching the method of choice for this project. Wide-band transmitters, as well as PA architectures which do not utilize switched-capacitor amplitude modulation, can instead greatly benefit from the increased peak efficiency and sharper roll-off of conventional "shunt" matching.

3.2: Tuner Design

According to the design proposal described previously, tuner units for precise calibration of the RF waveforms must be designed. In order to achieve proper suppression of the second and third harmonics of the output signal, the four waveforms, v_α , v'_α , v_β , and v'_β , must be tuned to specific phase differences as set by equation 2.33, as well as duty cycles of $d = 50\%$.

As showcased in the plots of figure 2.13, minimizing deviation from these ideal waveform characteristics is a priority. Thus, the tuners need to be able to change the duty cycle and phase of the waveforms in very small increments, so as to counteract even the slightest of variations which can become catastrophic with regard to spectral purity of the output signal.

Moreover, the tuners must be controlled digitally. This would allow for easy tuning and storage of optimal tuner states in the case of manual calibration, as well as facilitating the implementation of a feedback loop. This feedback loop, as mentioned in the design proposal, would be able to access the tuners and decide on the correct settings based on the resulting waveforms using digital logic.

Linearity and monotonicity are less of a priority. The tuners, effectively digital-to-time converters (DTC's), would be able to transpose one or both edges of a waveform – referring to the duty cycle tuner and phase tuner, respectively – in the time domain based on a digital code. Whether or not a change in code will result in a set and specific shift in time is of little importance to the loop logic, which will nevertheless attempt to increase or decrease the code

until the optimal value is reached. Monotonicity is of higher significance in this case, although it can be tolerated up to a point, as will be discussed later.

Finally, the tuners should have sufficient tuning range. Considering worst-case process, temperature, and voltage (PVT) variations, there must always be a tuner state which will allow for the desired duty cycle and phase characteristics of all waveforms.

3.2.1: Current Starve

Considering the first two specifications, those of accuracy and digital control, a digitally controlled delay element [12] (DCDE) utilizing current-starved inverters [13] (CSI) is utilized as the building block of the tuners. These delay elements are able to transpose either edge of a waveform by manipulating the drive strength of the pull-up and pull-down component of an inverter, thus influencing the resulting propagation delays.

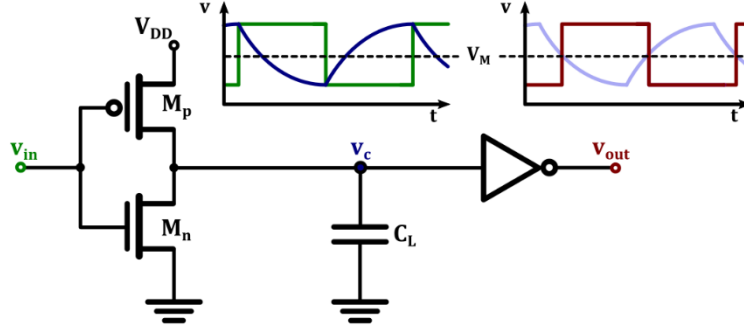


Figure 3.11: Schematic of Inverter Pair with Explicit Capacitive Load, with Node Voltage Waveform Illustrations.

Figure 3.15 displays a schematic of an inverter consisting of transistors M_p and M_n , loaded with a capacitor C_L , followed by a second, ideal inverter. The first inverter is driven by a square wave v_{in} . The load capacitor is charged and discharged through the on-resistances R_p and R_n , of transistors M_p and M_n , respectively. The output v_{out} of the ideal inverter will be a delayed duplicate of v_{in} . The resulting delay depends, in approximation, on the values of R_p , R_n , and C_L . A non-ideal second inverter would contribute to C_L , as well as present delay of its own, however this will not be considered for this example.

When M_n is enabled, as seen in figure 3.16, the capacitor voltage can be calculated as:

$$V_C = V_{DD} \left(1 - e^{-\frac{t}{R_n C_L}} \right) \quad (3.11)$$

Assuming an ideal switching threshold of $V_{DD}/2$ for the second inverter, the high-to-low propagation delay t_{pHL} can be calculated as:

$$t_{pHL} = R_n C_L \ln(2) \quad (3.12)$$

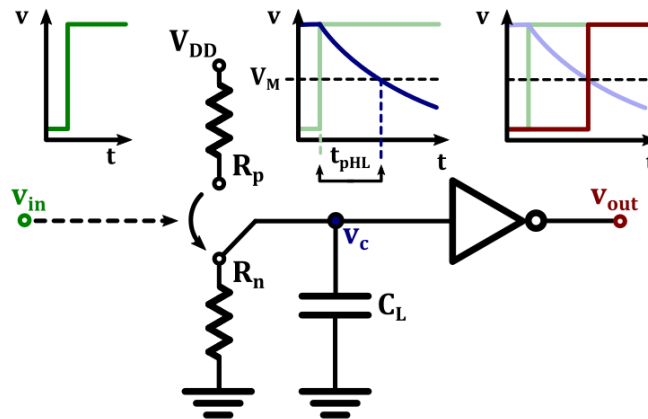


Figure 3.12: Equivalent Circuit to 4.15 with Illustrations for Propagation Delay.

Adding an extra NMOS M_{ne} in parallel to the first, as seen in figure 3.17, will reduce pull-down on-resistance, thus reducing propagation delay for the falling edge. This predictably results in a larger duty cycle for the waveform in the output of the second inverter.

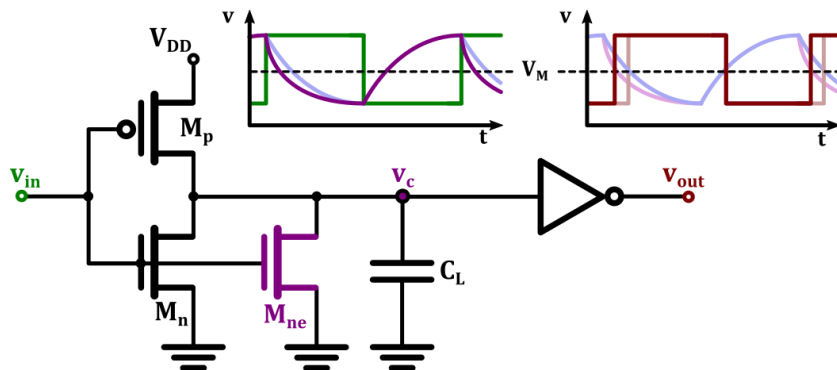


Figure 3.13: Continuation of Figure 3.15 with Additional NMOS.

Figure 3.18 displays an illustration of a digitally-controlled current-starved inverter, wherein N multiple equally-sized NMOS transistors $M_{ne,i}$ are connected between M_n and ground. The gates of $M_{ne,i}$ are connected to either V_{DD} or ground, depending on thermometer code $D < 1:N >$. As more bits have a positive value, more of the extra transistors are enabled, resulting in an increase of the output duty cycle d_{out} . Granted, the gate of “baseline” transistor $M_{ne,0}$ is always connected to V_{DD} in order for the inverter to be operational regardless of code D_{ne} .

As an example, time-domain plots for v_{in} , v_c , and v_{out} are displayed in figure 3.19, as generated from a simulation of a system similar to that of figure 3.18. Code D is swept, enabling more extra NMOS units to increase output duty cycle.

A plot of output duty cycle versus code D is presented in figure 3.20. A non-linear relation between D and resulting duty cycle can be observed. This comes as no surprise, given that the reduction in pull-down on-resistance is a result of the addition of equal resistances in parallel. As more transistors are enabled, the on-resistance of M_n dominates the pull-down, with each extra NMOS having very little effect. These diminishing returns can allow for very small increments to be added to or removed from the duty cycle, provided that a large number of units are already enabled.

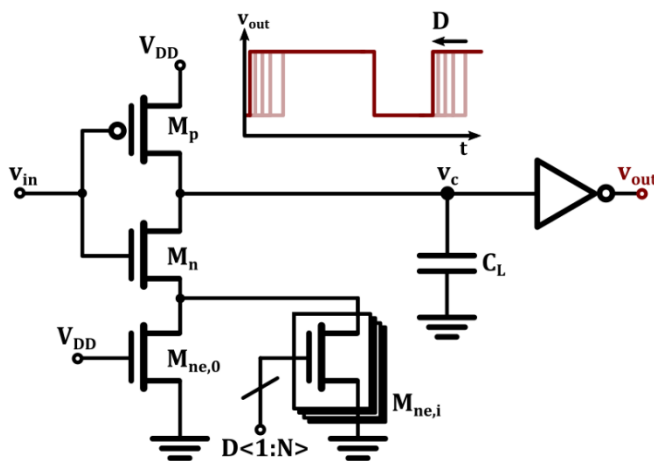


Figure 3.15: Schematic of NMOS-Based Current-Starve Inverter Duty Cycle Tuner, with Output Waveform Illustrations.

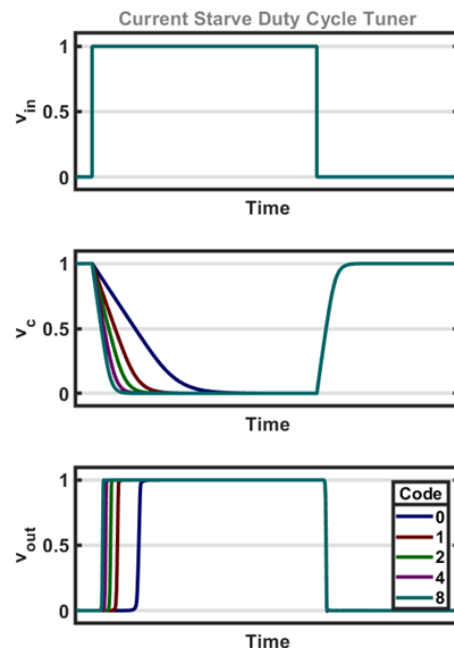


Figure 3.14: Simulated Time-Domain Plots of Node Voltages, for Different D .

This can be illustrated from the right-hand plot of figure 3.20, wherein codes 10 through 18 are considered. For this specific configuration, duty cycle increments of less than 0.07% are achieved, resulting in very precise, monotonous tuning. Although the curve remains non-linear, this is not a significant issue at this point.

By properly sizing the baseline transistor $M_{ne,0}$, a decent tuning curve can be implemented with a minimal number N of extra controlled NMOS units. The sizing of the PMOS transistor M_p affects pull-up on-resistance, and can be adjusted so as to center the curve towards the desired duty cycle value of 50%.

The precision granted by the technique described above does, however, come at the cost of range. For the example of figure 3.20, the tuning range is confined between duty cycle values of 49.85% and 50.07% if a baseline transistor of normalized width $S_{ne,0} = 11$ is used. Considering the drastic changes in duty cycle that can be brought forth due to process variation, a range this limited will not be sufficient for calibration towards $d = 50\%$.

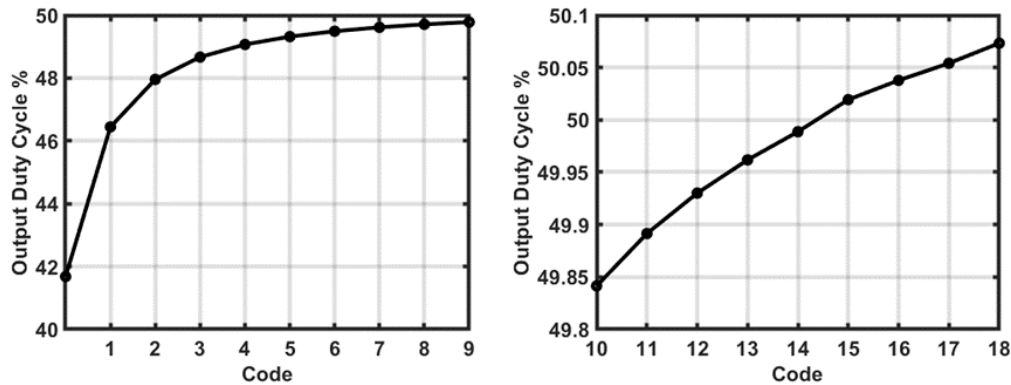


Figure 3.16: Simulated Plots for Duty Cycle vs. D . Right: Zoomed In.

3.2.2: “Coarse/Fine” Configuration

In order to increase range, a separate preceding coarse duty cycle tuning stage is considered. Examining the plot of figure 3.20 reveals that larger increments are possible for smaller values of D . By designing the coarse duty cycle tuner with a smaller number of baseline unit transistors, so as to operate, for instance, between codes 1 and 5 on the example of figure 3.20, the overall tuning range can be significantly increased. A such configuration is illustrated in figure 3.21.

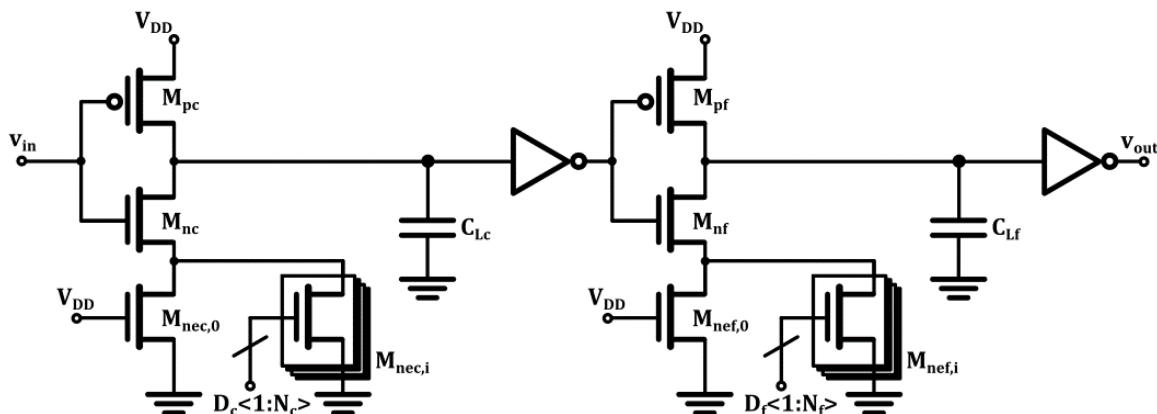


Figure 3.17: Schematic of “Coarse/Fine” Duty Cycle Tuner Configuration.

Figure 3.22 displays qualitative plots for a coarse-fine configuration. A “compound” tuning code, which will cycle through the entire fine-tuning range for each coarse step, is devised. Both component tuning curves – coarse and fine – are non-linear, as demonstrated in the simulation results of figure 3.20. The result is likely a non-monotonous tuning curve; since the step size of the coarse tuner is inconsistent, abrupt changes occur during coarse transitions. If the fine-tuning range is scaled down to the smallest coarse step, the result would be sharp, albeit monotonous, transitions at larger coarse steps, jeopardizing tuning resolution.

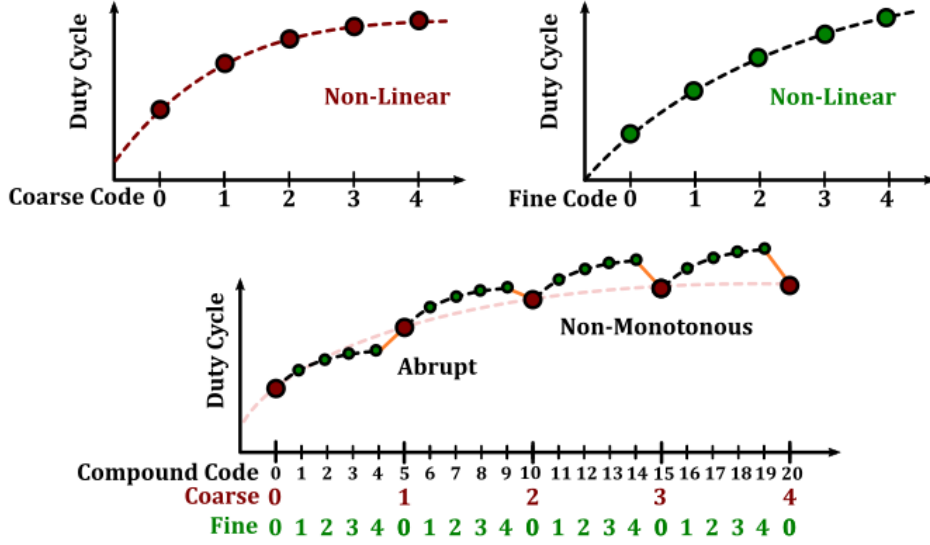


Figure 3.18: Qualitative Plots of "Coarse/Fine" Tuning Curve, with Non-Linear Coarse Stage.

In order to preserve the quality of the calibration, the coarse tuner must have a linear curve. In order to achieve this, attention must be paid to the sizing of the tuning NMOS units $M_{nec,i}$. Figure 3.23 displays an illustration of a coarse duty cycle tuner with N discrete stages. The $N + 1$ transistors are of equal length, and their on-resistance can be approximated as:

$$R_i = \frac{R'}{S_i} \quad (3.13)$$

Where R' is the on-resistance of a minimum width and length transistor, and size S_i is the width-to-length ratio of transistor i normalized to this minimum size. It is calculated as:

$$S_i = \frac{W_i/L_i}{W_{min}/L_{min}}$$

From equation 3.12, the propagation delay for the falling edge can be calculated as:

$$t_{pHL,i} = \left(\frac{R'}{S_{nc}} + \frac{R'}{\sum_{j=0}^i S_j} \right) C_L \ln(2) \quad (3.14)$$

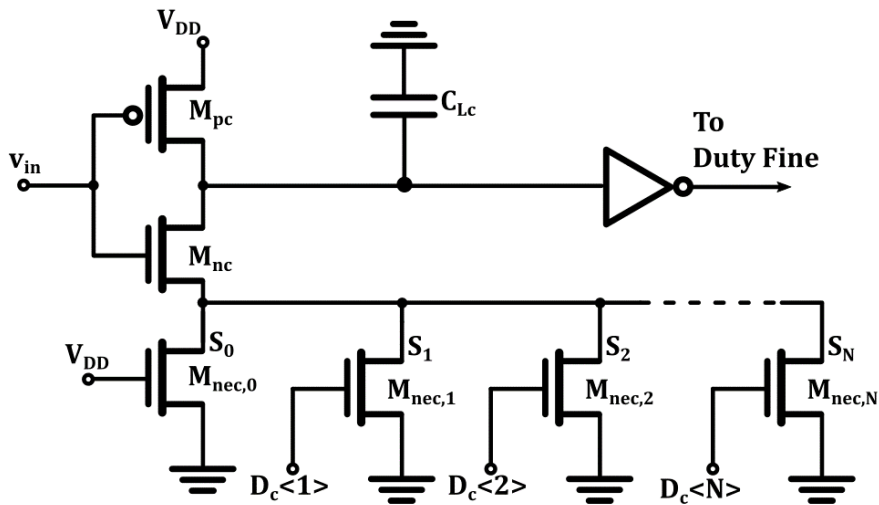


Figure 3.19: Schematic of Coarse Duty Cycle Tuner, with Separately Sized NMOS Units.

S_{nc} refers to the size of transistor M_{nc} . $t_{pHL,i}$ refers to the high-to-low delay when all tuning transistors up to i are enabled. Since thermometer code is used to control them, the contribution Δt_i of each unit depends on those enabled before it:

$$\Delta t_i = \left(\frac{1}{\sum_{j=0}^{i-1} S_j} - \frac{1}{\sum_{j=0}^i S_j} \right) R' C_L \ln(2) \quad (3.15)$$

In order to create equally sized steps, i.e., equalize Δt for every stage, the following calculations are done:

Set $n_i = \sum_{j=0}^{i-1} S_j$: Total width prior to addition of stage i .

$$\Delta t_i = \left(\frac{1}{n_i} - \frac{1}{n_i + S_i} \right) R' C_L \ln(2), \quad \Delta t_{i+1} = \left(\frac{1}{n_i + S_i} - \frac{1}{n_i + S_i + S_{i+1}} \right) R' C_L \ln(2)$$

For $\Delta t_i = \Delta t_{i+1}$:

$$S_{i+1} = W_i \frac{n_i + S_i}{n_i - S_i}, \quad n_{i+1} = n_i + S_i \quad (3.16)$$

Based on this formula, a sequence of transistor sizes can be calculated for a given pair of S_0, S_1 . By using variables R_{nc} and C_L to set up a proper time interval for the first step, the subsequent stages can be sized accordingly for a linear tuning curve.

Figure 3.24 displays plots generated using the formula of equation 3.16 for $S_1 = 1$, i.e., a minimum sized first stage $M_{nc,1}$. Different curves represent different sizes S_0 of the “baseline” transistor $M_{nc,0}$. The Y-axis displays the proper size S_i of its corresponding stage i on the X-axis, for a linear tuning curve.

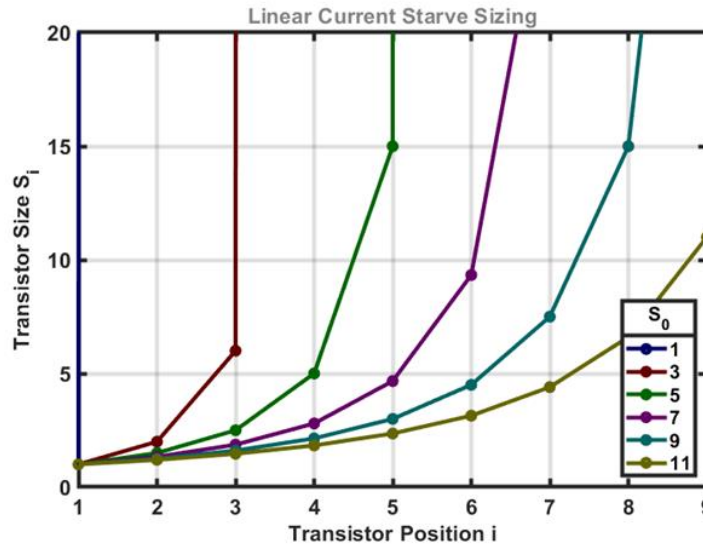


Figure 3.20: Stage Sizes for Linear Tuning vs. Baseline Size.

A few interesting observations can be made. Higher values of S_0 , as used for the fine tuner, require less scaling of subsequent stages to maintain linearity. It should be noted, however, that the steps themselves are rather small, as evidenced by equation 3.15. Thus, higher values of C_L and R' must be used for a coarse tuner.

In the case of C_L this, among other issues, is a detriment in terms of area and power consumption. In the case of R' , since it represents the on-resistance of a minimum-size transistor, this would require a significant increase in transistor length, resulting in impractical or cumbersome design.

On the other hand, small values of S_0 require for subsequent stages to be increasingly larger in order to maintain tuner linearity. The desired widths often reach infinity rather early in the sequence, limiting the number of stages possible. For $S_0 = 1$, a minimum sized baseline transistor, the first step is so large that it becomes impossible to replicate, no matter the size of the subsequent stage. This can be further illustrated in figure 3.20.

An intermediate size for S_0 , such as $S_0 = 5$, may present the perfect balance for a linear coarse tuner with an adequate number of stages, and a decent step size which does not require extreme values for C_L , R' , or S_i . As an

example, figure 3.25 displays a simulated duty cycle plot for $S_0 = 5$ and $S_1 = 1$, wherein the size of all subsequent stages was set by using equation 3.16. The resulting widths are:

$$S_0 = 5, \quad S_1 = 1, \quad S_2 = 1.5, \quad S_3 = 2.5, \quad S_4 = 5, \quad S_5 = 15$$

For comparison, a plot is also presented for the case where all stages for $i \neq 0$ have a size of $S_i = 1$, i.e., prior to scaling.

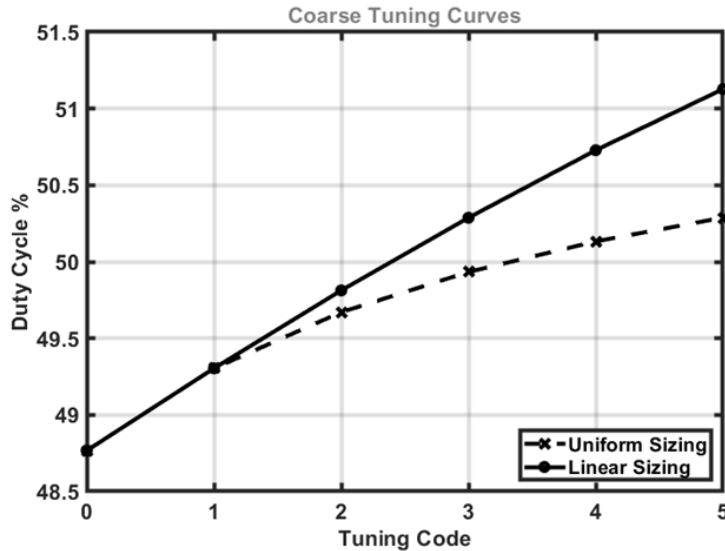


Figure 3.21: Duty Cycle vs. Coarse Code for Linear and Uniform Unit Sizing.

Whereas the unscaled configuration displays the familiar non-linearity also present in figure 3.20, the curve is highly more linear after each stage had been scaled. Although the relation is not perfectly linear, any issue that may present itself during coarse transitions in the overall tuning curve would be well within limits of tolerance.

Qualitative plots of a coarse-fine duty cycle tuning configuration are displayed in figure 3.26. In contrast with figure 3.22, the coarse tuner is now linear, resulting in smoother and monotonous coarse transitions, as well as increased range. In order to avoid redundancy while maintaining monotonicity and resolution, the coarse step size must be adjusted to fit the entire range of the fine tuner, and be larger than it by an increment at most equal to the largest step found in the fine tuner.

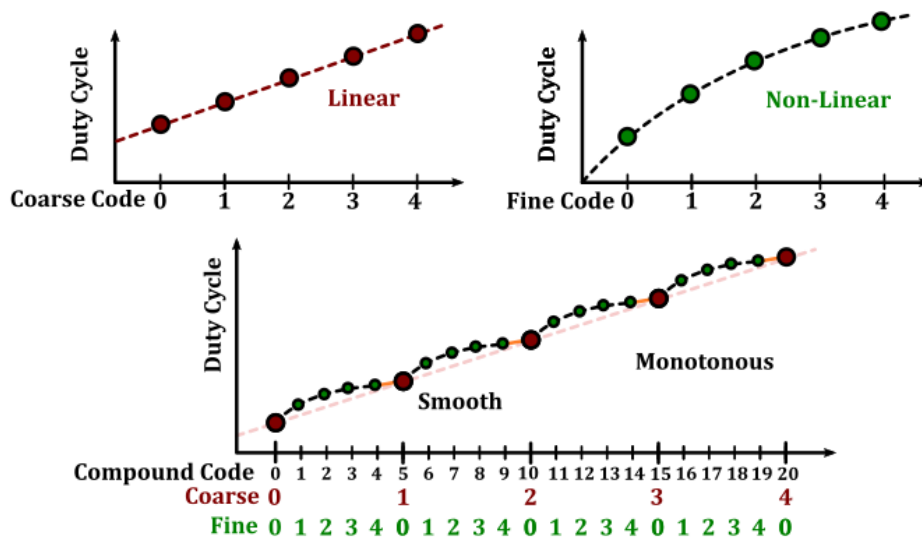


Figure 3.22: Qualitative Plots of "Coarse/Fine" Tuning Curve, with Linear Coarse Stage.

This may not always be possible in practice, and certain abrupt or non-monotonous transitions may appear, especially under conditions of extreme variation. Regarding the implementation of tuners in this project, it is important that these "glitches" do not present a threat to spectral purity. In other words, there must always be an

attainable duty cycle value for which the resulting output signal harmonics are within specification, in spite of any glitches in the curve.

Considering the above, minimizing the number of coarse transitions, and thus maximizing the range of the fine-tuner, becomes a desirable design goal. For this reason, transistor scaling similar to the method used for the coarse tuner can be used on the fine tuner as well. Although linearity of the fine tuner is not a priority, this helps reduce the number of coarse stages needed to reach the desired duty cycle range of the overall tuner.

As an alternative to the NMOS-based duty cycle tuner described above, a PMOS-based current-starve tuner can easily be designed. Such a tuner would influence the rising edge of the waveform by reducing pull-up on-resistance, thus speeding up the charging of capacitor C_L and reducing t_{pLH} . When followed by a second inverter, this results in a reduction of output duty cycle.

An illustration of a PMOS-based tuner is shown in figure 3.27, an equivalent to figure 3.18. The system operates like, and is designed based on the same principles as its NMOS counterpart. The difference is its effect on output duty cycle, and the design of $D < 1:N >$, now inverted thermometer code, which will send a bit equal to 0 in order to enable a unit. A combination of PMOS and NMOS tuners is used in this project. As will be discussed later in this thesis, this allows for better control of the four component waveforms used to create the output signal.

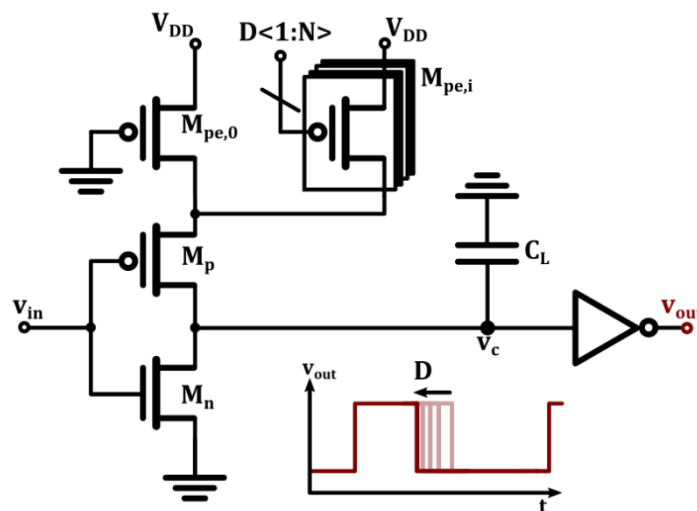


Figure 3.23: Schematic of PMOS-Based Current-Starve Inverter Duty Cycle Tuner, with Output Waveform Illustrations.

3.2.3: Phase Tuner

Following the same principles, a current-starve phase tuner can also be designed. By having extra transistors on both the pull-up and pull-down networks, and assuming that they are sized properly so as to provide equal amounts of on-resistance per unit, both edges of the waveform will shift in time, while maintaining the same duty cycle.

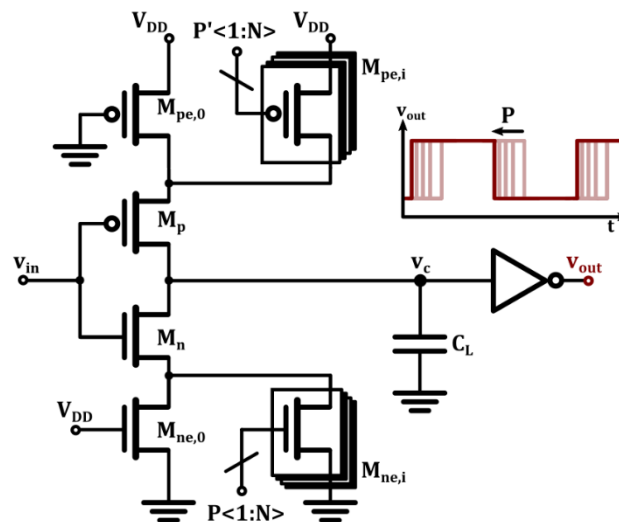


Figure 3.24: Schematic of Current-Starve Inverter Phase Tuner, with Output Waveform Illustrations.

Figure 3.28 presents an illustration of such a design. In this configuration, controllable NMOS units are added to the pull-down, whereas equivalent PMOS are added to the pull-up. They are controlled by the same code $P < 1:N >$, which is inverted into $P' < 1:N >$ for the PMOS units.

As P_e grows larger, and more units are enabled on both networks, both high-to-low and low-to-high propagation delays t_{pHL} , t_{pLH} are made smaller, moving the entire waveform “backwards in time”. This is equivalent to a negative phase shift on the waveform, and is used in this project to fine-tune the phase differences between the four component waveforms in order to perform harmonic suppression of the output signal.

As an example, analogous to that of figures 3.19 and 3.20, figure 3.29 presents simulated time-domain plots for v_{in} , v_c , and v_{out} for the system of figure 3.28, sweeping code P_e . A plot of phase shift versus P_e , for which the phase difference between v_{out} and v_{in} is measured, is shown in figure 3.29, displaying similar characteristics to the duty cycle tuner of figure 3.20.

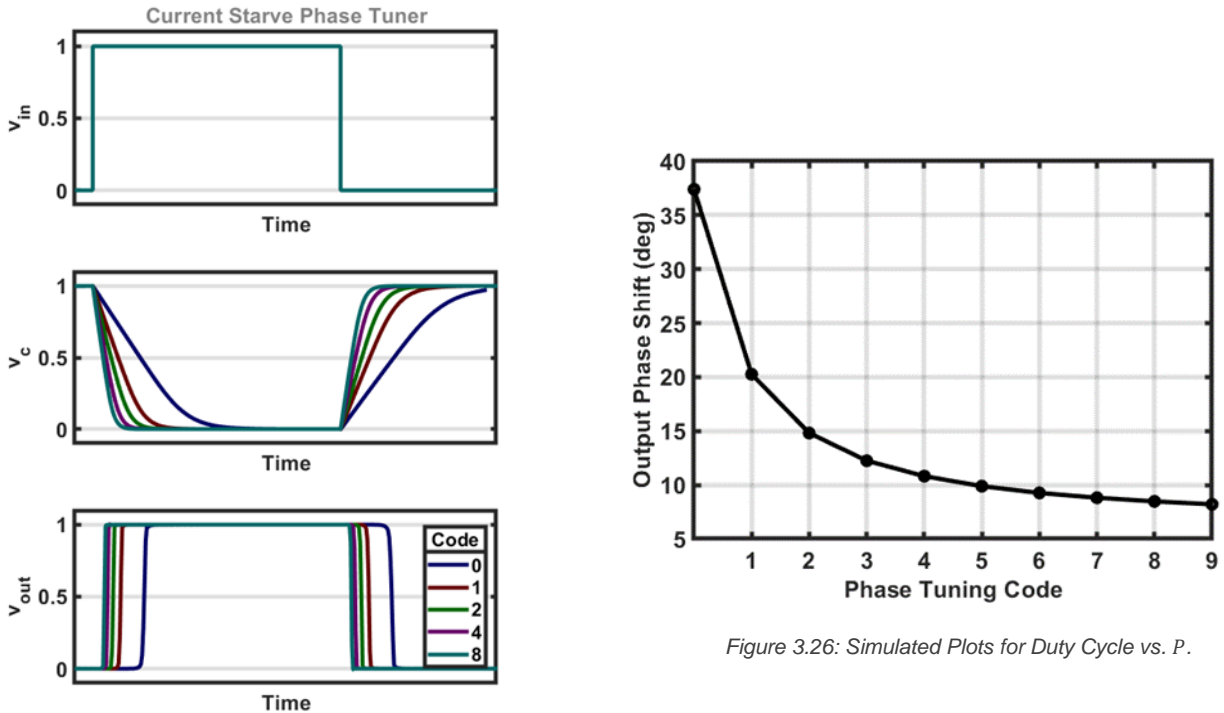


Figure 3.25: Simulated Time-Domain Plots of Node Voltages, for Different P .

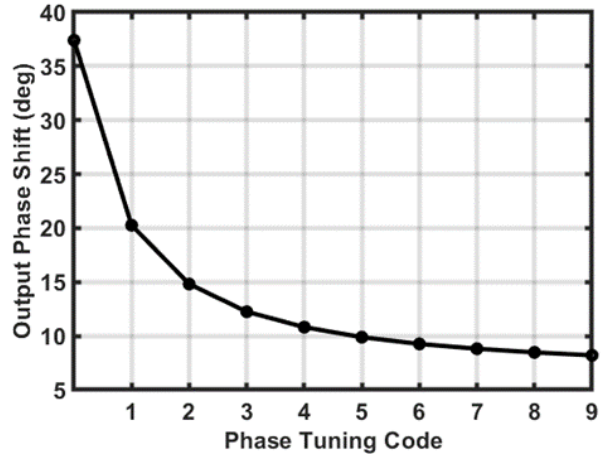


Figure 3.26: Simulated Plots for Duty Cycle vs. P .

The same design guidelines and restrictions as discussed for the duty cycle tuners also apply in this case. A coarse-fine configuration is implemented, wherein a linear tuning curve is desired for the coarse stage. Granted, since the effects of phase deviation in harmonic suppression differ from those of duty cycle deviation, different specifications must be developed with regard to range and resolution.

An additional design restriction is added in the form of relative sizing between the PMOS and NMOS units. In order to maintain the waveform duty cycle, and since the same code is used for both tuning networks, both edges must be shifted by the same amount per unit. Since this cannot always be the case in the presence of process variation, this phase tuner is best used in conjunction with a duty cycle tuner which will fine-tune any deviations.

Despite the ability of the duty cycle tuner to rectify errors caused by a mismatch in unit on-resistance between the PMOS and NMOS networks, such instances of coupling between phase and duty cycle are highly undesirable. When these are present, they provide an iterative element to the tuning process, significantly increasing calibration time.

Minimizing load capacitance C_L , at least on the phase tuner, is important for suppression of “phase/duty coupling”. Even small differences between R'_p and R'_n , the on-resistances of unit-size PMOS and NMOS transistors as used for equation 3.14, are amplified by C_L . This results in exaggerated differences between propagation delays t_{pHL} and t_{pLH} , which affects the resulting duty cycle.

Extrapolating from equation 3.14, for the case where $W_n = W_p = W_{np}$, i.e., both M_n and M_p are scaled equally compared to their respective unit sizes:

$$t_{pHL} = \left(\frac{R'_n}{W_{np}} + \frac{R'_n}{\sum W_i} \right) C_L \ln(2), \quad t_{pLH} = \left(\frac{R'_p}{W_{np}} + \frac{R'_p}{\sum W_i} \right) C_L \ln(2) \quad (3.17)$$

$$t_{pHL} - t_{pLH} = (R'_n - R'_p) C_L \ln(2) \quad (3.18)$$

It becomes, therefore, important to not rely on large values of C_L for this design. In fact, an explicit load capacitor is not necessary and can be omitted, since the subsequent inverter stage can provide appropriate capacitive loading when properly sized. Instead, focus should be put to transistor sizing, using design equations such as 4.15 and 4.16 to create the desired step sizes for minimal C_L .

In a scenario where more range is required for the phase tuner, as has been the case for this project, a simple tapped-inverter delay line can be added as a coarse stage, relegating the subsequent two current-starve tuners into "fine" and "super-fine" stages.

Serially connecting several equally-sized inverter pairs, each with a propagation delay of t_p tuned to fit the entire tuning range of the subsequent stages, and connecting each of their outputs to a separate, digitally controlled CMOS transmission gate creates the coarse phase tuner stage of figure 3.31. The tapped-inverter curve is inherently linear, as multiples of t_p , the coarse inverter propagation delay, are added to the total shift.

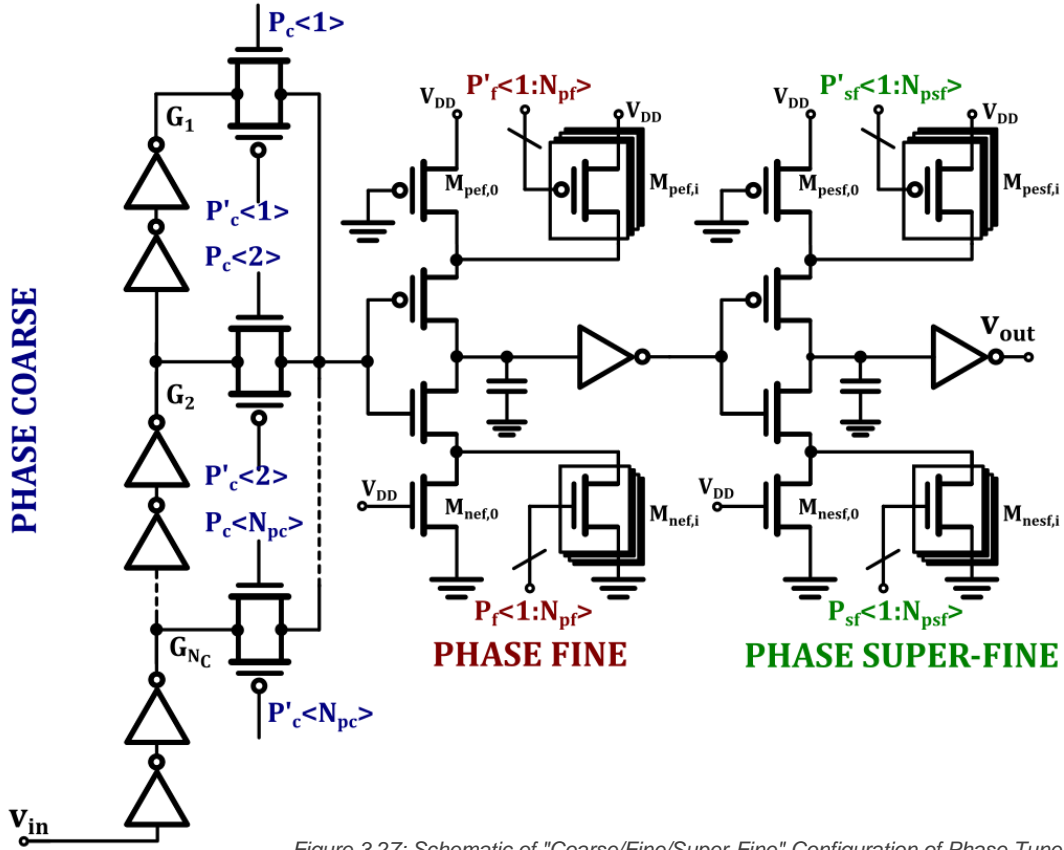


Figure 3.27: Schematic of "Coarse/Fine/Super-Fine" Configuration of Phase Tuner.

The code used to control the tapped-inverter coarse tuning stage has the form of spatial unary code [14], wherein there can only exist one bit with a value of 1, representing a single transmission gate being enabled. The waveform at the output of the corresponding inverter is used as the input to the fine stage. Figure 3.31 also illustrates the fine and super-fine phase tuning stages, as described previously.

Compound code $P < 1:N >$ is designed to operate the phase tuner, with $N = N_c + N_f + N_{sf}$, the total number of controllable of NMOS units used to calibrate the phase shift. Its inverse, $P' < 1:N >$ is used for the complimentary PMOS units.

3.2.4: Tuner Diagrams

The NMOS-based two-stage duty cycle tuner discussed previously is displayed in the illustration of figure 3.32. The compound code controlling this system is, $D < 1:N >$, with $N = N_c + N_f$.

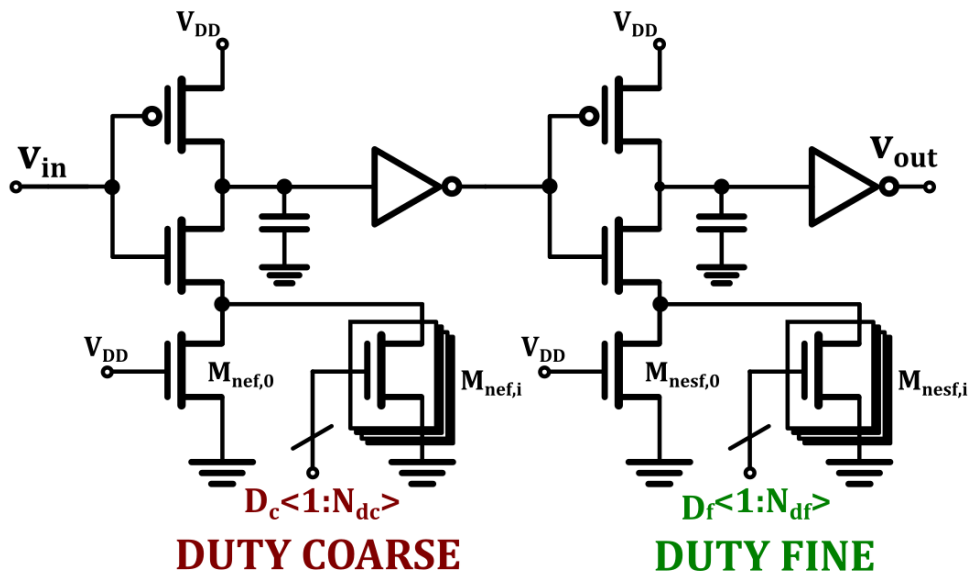


Figure 3.28: Schematic of "Coarse/Fine" Configuration of Duty Cycle Tuner.

Figure 3.33 presents block diagrams of the multi-phase duty cycle tuner and phase tuners described above, in which the segmentation of compound codes P and D is also visible.

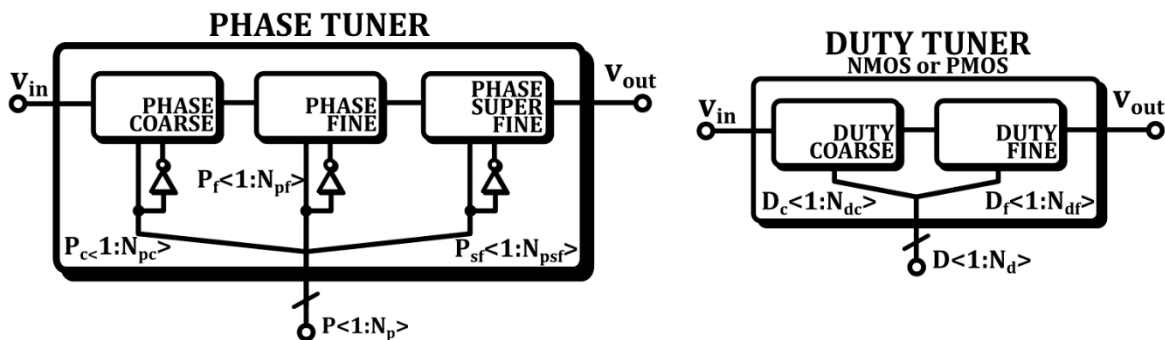


Figure 3.29: Block Diagrams of Phase Tuner and Duty Cycle Tuner.

3.3: Tuning Loop

At this point in the design procedure, digitally controlled tuners exist for duty cycle and phase shift. By combining the two, exact values can be set for the duty cycle and phase difference of two waveforms. Although this calibration can be performed manually in order to counteract process variation, voltage and temperature variations cannot be predicted or accounted for at this stage.

Moreover, another issue arises when multiple different frequencies are expected for the waveforms, as is the case for a multi-channel RF transmitter. Since the tuners manipulate inverter propagation delay times, which are translated into duty cycle or phase differences relative to the frequency of the waveform, it is expected that a given tuner state will provide different results for different channel frequencies. Even for narrow-band signals, these differences can be enough to threaten the spectral purity of the output.

As an example, a phase tuner state, i.e., a specific value for code P , can be considered, for which a phase difference of $\Delta\varphi = 60^\circ$ between two waveforms is achieved for an LO frequency of $f_{LO} = 2.4 \text{ GHz}$. In order to achieve this, a total time shift of $t_p = 69.44 \text{ ps}$ is applied through calibration of the delays of the various components of the phase tuner system.

If the same code P is used for $f_{LO} = 2.5 \text{ GHz}$, a different channel frequency, this propagation delay of $t_p = 69.44 \text{ ps}$ would be equivalent to a phase shift of $\Delta\phi = 62.5^\circ$. Considering the sensitivity of the harmonic suppression technique, as demonstrated earlier in figure 2.13, this 4.17% deviation from the ideal value of $\Delta\phi$ would have dire consequences for the third-order harmonic distortion of the output signal.

Therefore, a real-time calibration system must be implemented. This system must be able to provide optimal values for D and P , referring to the digital codes controlling the duty cycle and phase tuner, respectively, under every condition. Rather than attempt to measure and compensate for each variation, a simpler solution is found in the form of a feedback loop.

As a concept, this feedback loop should be able to monitor the resulting waveforms, as generated by the tuners, compare their duty cycles and phase differences to ideal references, and alter the relevant digital control codes until no discrepancy exists between outputs and references. A simple system-level block diagram of this concept is illustrated in figure 3.34.

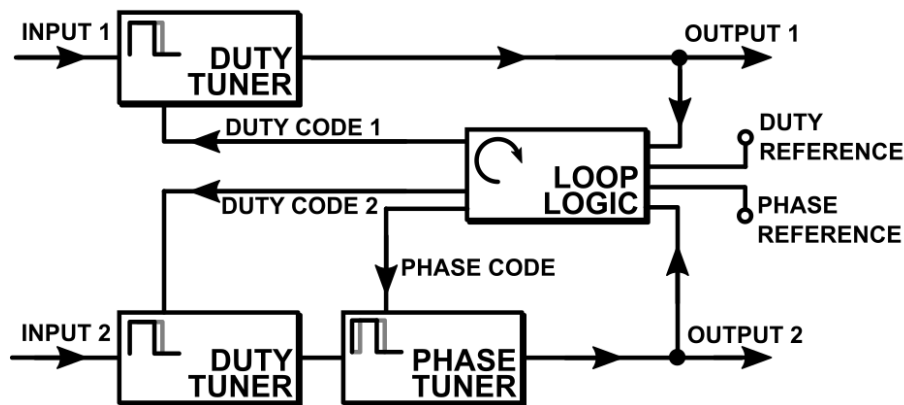


Figure 3.30: Block Diagram of Tuning Feedback Loop and Tuners.

While there is not much to discuss regarding the block diagram of figure 3.34, it is important to point out that each of the two waveforms is “probed” once by the loop logic, as opposed to having a separate probe point in between the two tuners for the second waveform. Although this may seem like an oversight, it is done by design in order to counteract any corruption in duty cycle caused by the phase tuner.

Even in the absence of “phase/duty coupling” as established earlier, a drift in duty cycle may be caused by imbalances in the voltage transfer characteristics of subsequent stages, especially in more extreme cases of process variation. This can be compensated for by probing as close to the output as possible, so as to get a more “complete picture” of the waveform, which will include any alterations made during propagation of the waveform through the inverter chain.

In practice this means that any subsequent circuitry, such as the power amplifier switches and buffer stages leading up to them in this case, will need to precede the probing point, as these can contribute to corruption of the waveform. This also makes the order in which the duty cycle tuner and phase tuner are connected irrelevant. These two considerations can be further illustrated by comparing the block diagrams of figures 2.20 and 3.34.

The loop circuitry must be able to extract the two relevant characteristics, namely duty cycle and phase, from the probed waveforms in order to compare them to ideal references. Thus, theoretical “duty cycle detector” and “phase detector” components are employed for this point in the design procedure. Although the implementation of these components will be discussed later, it is important to mention that their outputs should be given in a format (voltage, current, digital code, etc.) comparable to that of the references.

The loop logic itself must be able to compare the detector outputs to the references and make any necessary adjustments to the digital code in order to equalize them. The flowchart of figure 3.35 describes the logic process. The probe waveform duty cycle is read from the “duty cycle detector” component, and compared to the reference. If the duty cycle is higher than the reference, code D controlling the tuner is decreased, thus decreasing the duty cycle. The opposite happens if the waveform duty cycle happens to be lower than the reference.

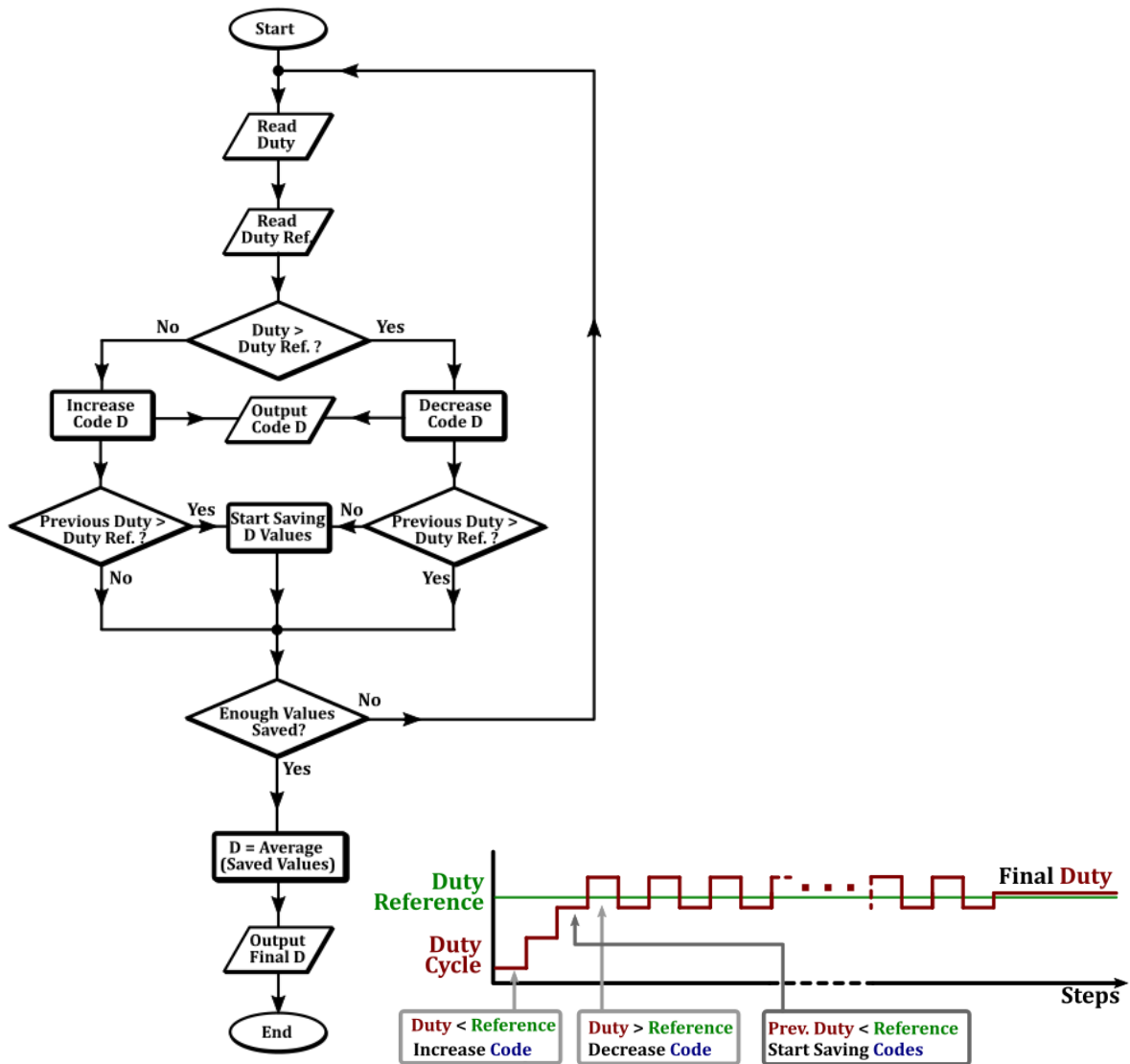


Figure 3.32: Flowchart of Basic Loop Logic.

Figure 3.31: Illustration of Ideal Loop Operation.

In the case when, for two consecutive samples, the duty cycle changes from being higher than the reference to being lower, or the inverse, this signifies that the optimal value for duty cycle is found between the two codes used for these two samples. The first time this happens during the tuning process, the values for code D start being recorded.

What is expected to follow is an oscillation between the two values for the remainder of the process, as can be seen from the qualitative plot of figure 3.36. After a set number of samples, the average of all recorded D values is calculated, serving as the final value for code D .

The purpose of this “wait-and-average” function is to protect against non-idealities in the tuning process. Figure 3.37 presents another qualitative plot, wherein a “glitch” appears. The duty cycle is increased, likely because of a coarse transition on the phase tuner. In the absence of a “wait” function, the final value for D would be significantly higher than the optimal one. In this case, however, the error is recovered as the code is gradually reduced and is given enough time to eventually settle around its optimal value.

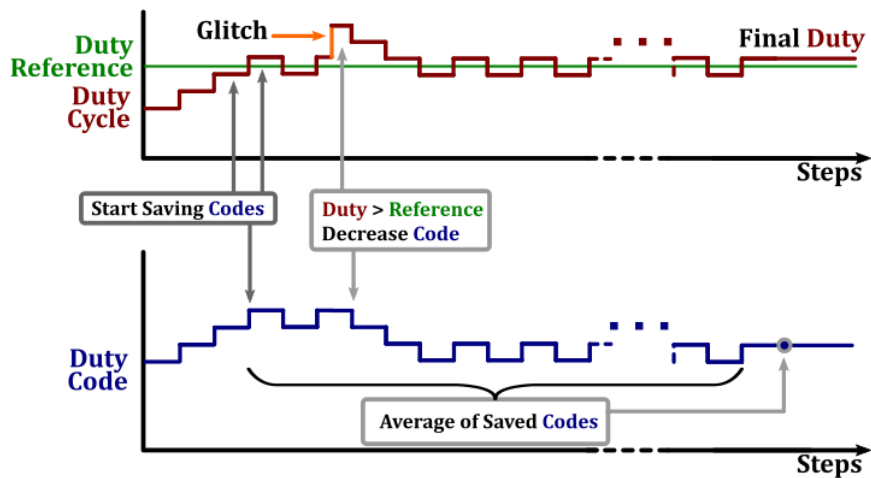


Figure 3.33: Illustration of Loop Operation with Glitch.

The actual averaging function of the logic is also beneficial in the case of large response delays. As a large number of inverter stages can be found between the tuner and probe, the effects of a change in code D are measured after a possibly significant amount of time. Moreover, the “duty cycle detector” component as well as the logic itself offer delays of their own, depending on their implementation.

Figure 3.38 displays another qualitative plot similar to those presented previously. In this scenario, t_{delay} is added as discussed above. This results in a significant increase in the amplitude of the oscillation, as the logic cannot reverse the tuning direction until several steps after an optimal value for duty cycle has been reached. This optimal value is, however, recovered through averaging. Thus, the averaging function can allow for faster operation of the

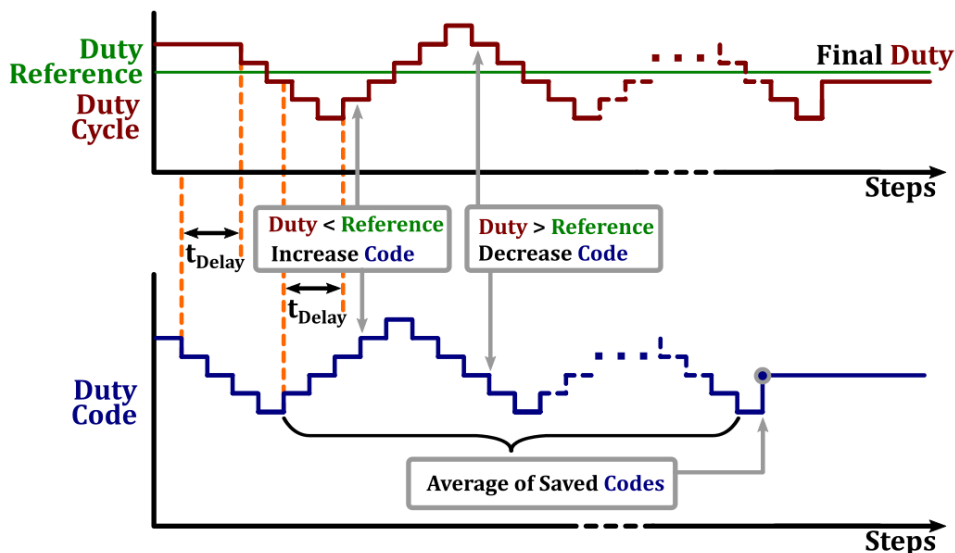


Figure 3.34: Illustration of Loop Operation with Delay.

loop logic, as well as increased distance between tuner and probe, which in turn can increase accuracy as discussed previously.

Although the above refers to an NMOS-based duty cycle tuner, the process is no different than that of a PMOS-based duty cycle tuner or a CMOS phase tuner, except for a change in direction; in those cases, a decrease in code begets an increase in the relevant quantity.

4: IMPLEMENTATION

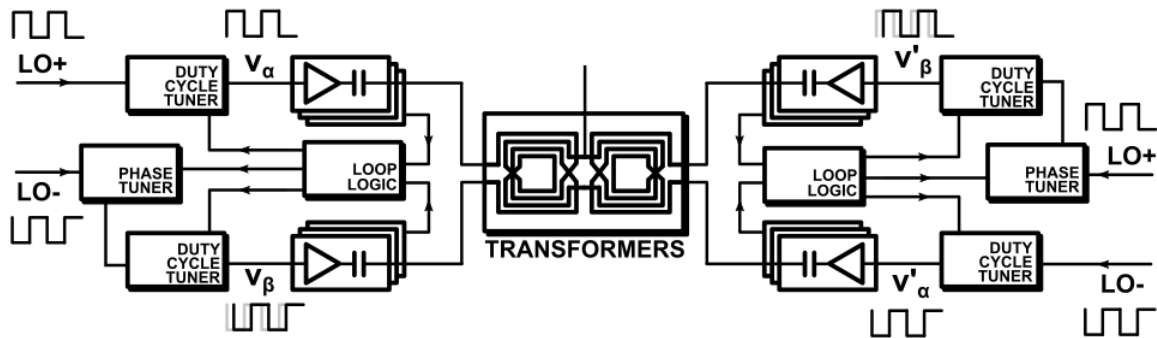


Figure 4.1: Block Diagram of Proposed System. Also Seen in fig. 2.20.

Thus far, the basic operating principles and design considerations for the building blocks of the proposed design have been discussed. The block diagram of figure 2.20, presented again here as figure 4.1, contains four switched-capacitor power amplifier arrays, each driven by waveforms tuned for duty cycle and phase via tuner blocks, in order to achieve suppression of third-order and even-order harmonics.

Two feedback loops monitor the duty cycle and phase difference of the waveforms, as seen at the PA outputs, and perform appropriate adjustments prior to transmission. Each pair of PA outputs is connected to an on-chip transformer in differential configuration, with the two transformer secondary coils joining together towards the single-ended output.

Schematic-level design and simulation of the system was done in Cadence Virtuoso, using 22nm technology, with several logic blocks written as Verilog code. Design specifications included a peak output power of 12 dBm, operating frequencies between 2.4 GHz and 2.5 GHz, and a maximum harmonic power of -41 dBm, at a supply voltage of $V_{DD} = 0.8 V$.

4.1: Transformer

Implementation of the transformers has proven to be the most time-consuming, and ultimately least rewarding, part of the implementation process. Interest was initially focused towards implementation of a single 3-coil transformer, or “4-way power combiner” [15], [16], which would be able to accommodate all four PA arrays, as seen in the concept illustration of figure 4.2. It was considered that such a design would lead to a more area-efficient design than the alternative, conventional two-transformer configuration that was eventually used in the final design.

Over 50 instances of “4-way” transformers were designed, rendered through EM simulation software, and tested in the context of the system. Much experimentation took place regarding the diameter, turn ratio, metal layers, trace width and spacing, and general architectures of these transformers, with some models employing parallel windings [17], interlaced schemes [18], or additional inductance, in series with the secondary coil, which would be employed as part of an explicit low-pass filter.

However, the requirement for high quality-factors at the operating frequency on all three coils, perfect symmetry between the two primaries, and a decently high self-resonant frequency posed a challenge given the complexity of such a design. Moreover, since the “no shunt” matching technique, as established in chapter 4.1.1, was used, a preference existed for “lopsided” designs employing primary coils of moderate inductance, coupled with significantly smaller secondaries.

As a result, the designed prototypes would suffer from either poor quality-factors, mostly regarding the small secondary coils, large coupling capacitance, or poor symmetry owing to the complexity of the architecture used. In some cases, the matching array capacitance was too large as the “no shunt” matching method was not fully taken advantage of. Therefore, harsh trade-offs existed between efficiency, both at peak power and power back-off, and harmonic suppression capabilities.

As experimentation progressed, more decent models were designed. However, these were lacking in terms of area and performance, compared to the two-transformer alternative. In many cases, the resulting output power would exceed the specifications, requiring additional primary inductance which proved troublesome to implement.

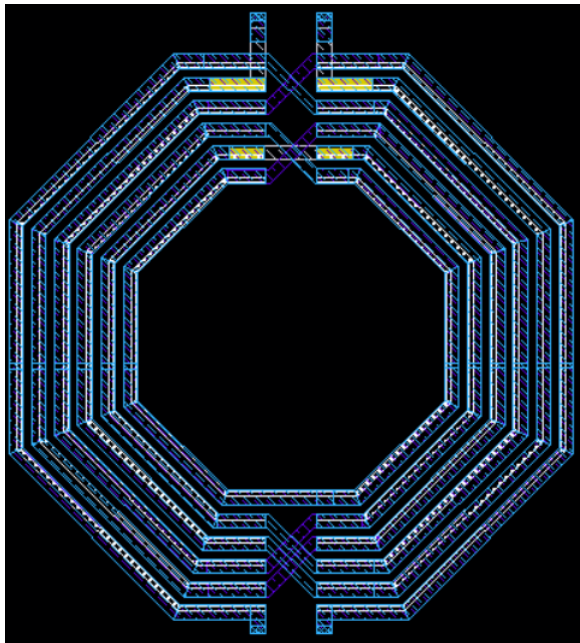
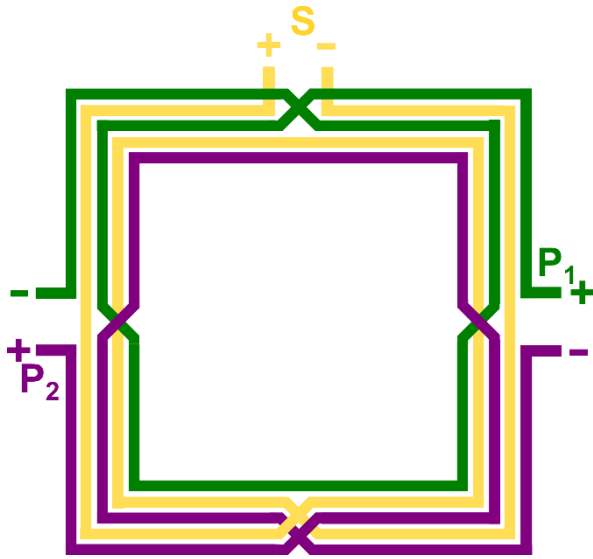


Figure 4.4: Layout of Implemented Transformer.

As no “4-way combiner” models existed in the design library, each prototype layout was manually designed and simulated. This fact, coupled with often long electromagnetic simulation processes, rendered this “trial-and-error” experimentation procedure highly time-consuming. Although significant progress had been made, the design of a “4-way” transformer and its implementation into this project was not pursued further.

Instead, a very conventional 2-coil transformer was implemented. Using the provided design library, a layout was automatically generated after setting values for L_p and L_s , and requesting an optimal quality-factor at the operating frequency. After a minimal amount of editing, the resulting transformer, pictured in figure 4.3, displayed the following characteristics for $f = 2.4 \text{ GHz}$:

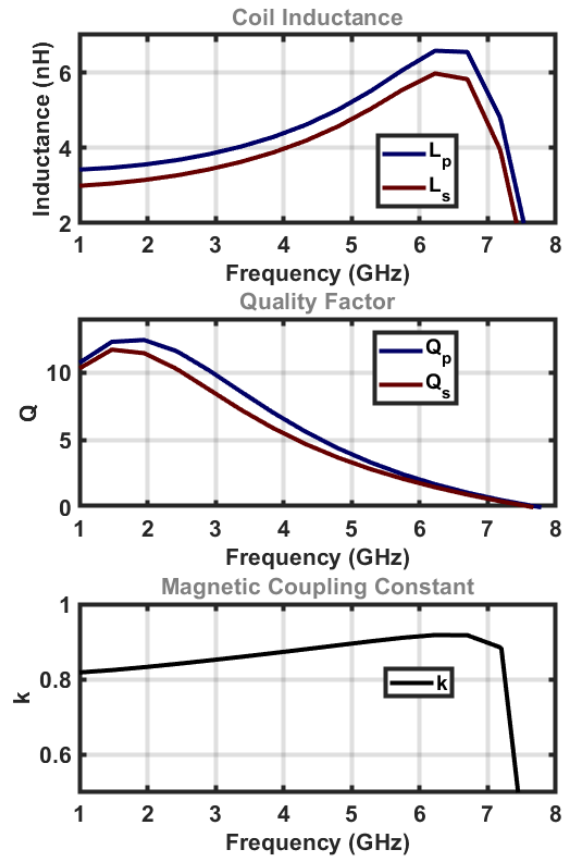


Figure 4.3: Simulated Plots of Inductances, Quality Factors, and Coupling Constant of Transformer.

$$Q_{\text{primary}} = 11.62, \quad Q_{\text{secondary}} = 10.35, \quad L_{\text{primary}} = 3.67 \text{ nH}, \quad L_{\text{secondary}} = 3.26 \text{ nH}, \quad k = 0.84$$

Figure 4.4 displays simulated plots of the above parameters versus frequency. Comparing these characteristics with the calculations and findings of chapter 4.1.3, it can be said that an SCPA array matched in “no shunt” configuration to a 50Ω load through this transformer would be operating in “hybrid mode”, close to the theoretical plateau of $L_s = R_L / \omega_{LO}$ as established earlier.

In this region, array input resistance R_{in} is the highest (figure 3.5), meaning less of an effect of inductor ohmic resistance on output power and efficiency. Although “hybrid mode” provides maximum peak efficiency according to the simulations of figure 3.10, where quality-factor is also considered, a lower L_s value would be preferable for back-off efficiency, as seen in figure 3.12. Operation in early “quasi-current mode” would also provide improved filtering for far-out harmonics (figure 3.7), an issue beyond the scope of this thesis.

Connecting the two transformers together can be done either in series or in parallel in “hybrid mode”, as demonstrated in figure 3.7. A parallel combination was chosen for this project due to the slightly increased filtering for second- and third- order harmonics compared to a series connection. Moreover, adjusting array capacitance for this case, i.e., correcting the frequency response from the bottom-left-hand plot of figure 3.7 to the bottom-right-hand one, requires a reduction in C_{array} rather than the increase required for a series connection.

Although a rather large value of $C_{array} = 4.6 \text{ pF}$ for the total capacitance of each of the four PA arrays was chosen for this project, the resulting harmonic performance and efficiency, as will be presented later, were adequate enough that no further experimentation on transformer design was deemed necessary.

4.2: PA Arrays

Implementation of the PA arrays was rather straightforward. The largest inverters at the array capacitor bottom-plates required the most attention, as improper sizing can have a drastic effect on efficiency. Although the expression for ideal PAE from equation 2.31 has been used thus far, only losses from charging the switched-capacitor array have been considered. However, the inverters themselves also contribute to overall power loss. Adapting from [11], for efficiency η at peak power:

$$\eta = \frac{P_{out}}{P_{DC}}, \quad P_{out} = \frac{V_{out}^2}{2R_{in}}, \quad V_{out} \approx \frac{2}{\pi} V_{DD} \frac{R_{in}}{\frac{R_{on}}{N} + R_{in}}, \quad P_{DC} \approx P_{out} + NC_{sw}V_{DD}^2f$$

Compared to earlier, these calculations also include switch on-resistance R_{on} and inverter input capacitance C_{sw} . For a given transistor length, the relation of these values to transistor width can be approximated as:

$$R_{on} = \frac{R'_{on}}{W}, \quad C_{sw} = WC'_{sw} \quad (4.1)$$

With R'_{on} and C'_{sw} referring to their respective quantities for minimum width transistors. W is transistor width normalized to minimum transistor width. Thus, efficiency can be calculated as:

$$\eta = \frac{1}{1 + \frac{\pi^2 f}{4R_{in}^2} NW C'_{sw} \left(\frac{R'_{on}}{NW} + R_{in} \right)^2} \quad (4.2)$$

Increasing W offers both positive and negative contributions to efficiency. An optimal value can, therefore, be found at:

$$W_{opt} = \frac{R'_{on}}{N R_{in}} \quad (4.3)$$

This value for W results in maximum efficiency, and depends on R'_{on} , which can be considered a technology parameter. Transistor length L is kept to a minimum, since it would increase both C_{sw} and R_{on} . Optimal W_{opt} also depends on input resistance R_{in} . N refers to the number of PA units in the array; for the 7-bit AM resolution of this system, $N = 127$.

Since manual adjustment of C_{array} was required in order to implement parallel transformer combination in “hybrid mode”, R_{in} cannot be confidently calculated from equation 3.4, as would be the case for the other two operating regions. Besides, different effects including inductor ohmic resistance also influence the realistic value of W_{opt} . Thus, a parameter sweep was performed following these calculations, and a value of $W = 48$ was found to maximize efficiency for the system at this point. W in this case refers to the actual transistor width divided by the minimum transistor width of 80 nm allowed by the technology.

The above calculations assume that both the PMOS and NMOS transistors of the inverter have the same on-resistance. This was mostly achieved under typical process conditions by using a P-to-N width ratio of $W_p/W_n = 1$, implementing Super-Low Voltage Threshold (SLVT) PMOS transistors in conjunction with Regular Voltage Threshold (RVT) NMOS transistors for the inverters. Since equality between pull-up and pull-down on-resistances is also important for duty cycle propagation, this configuration was also used throughout the rest of the system.

Each of the PA inverters is preceded by a CMOS NAND gate. This gate receives the LO waveform and a single bit of thermometer code $AM < 1:N >$, signifying whether or not a specific PA unit is enabled. The LO is propagated towards the PA output for $AM < i > = 1$, whereas $AM < i > = 0$ results in its respective capacitor bottom-plate being held to DC. Thus, amplitude modulation is achieved according to the operating principles of a switched-capacitor PA as discussed in chapter 3.1. The AM encoder was implemented in the form of Verilog code.

The NAND gate is sized at $W = 12$, meaning a fanout of $F = 4$. This value was chosen for the buffer chain so as to provide decent driving strength while minimizing the total number of stages needed to ramp up from the tuner stage towards the PA. Thus, fewer losses occur and system efficiency is increased.

A “PA unit” or “PA cell” as seen in the illustration of figure 4.5 consists of these two gates, along with a unit capacitor. Each group of four PA units is driven by an inverter, also with a normalized width of $W = 12$ for a fanout of $F = 4$. Each group of four such inverters is, in turn, driven by an identical inverter, and so forth until the array input is reached. Each of the four arrays has a separate input, receiving the tuned LO waveform from its respective tuner block.

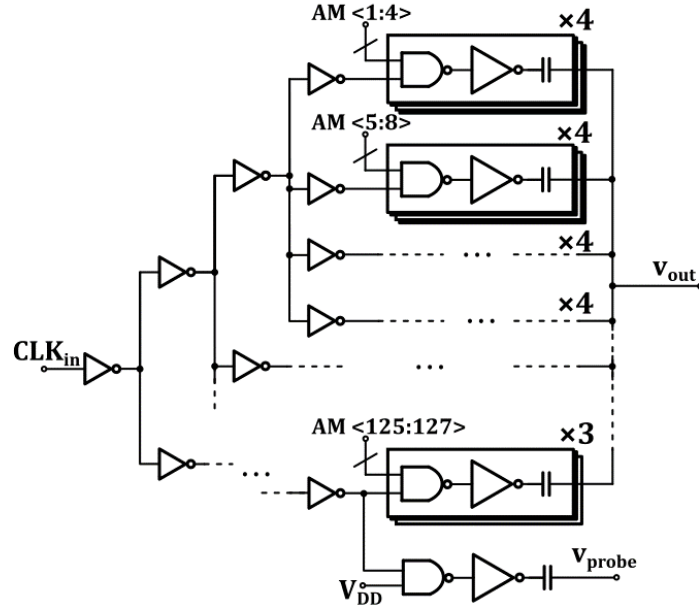


Figure 4.5: Schematic of PA Array.

Although only $N = 127$ units are needed for a 7-bit DAC, the 128th unit of each array acts as a probe for the feedback loop. This probe unit, built and sized identically to the rest, attempts to emulate the RF waveform as seen by the array capacitor bottom-plates, and thus present the most accurate depiction of its characteristics. Even if a feedback loop was not in use, a 128th unit, or at least part of one, would be necessary as a dummy in order to equalize capacitive loading throughout the array.

Each of the four PA array outputs was connected to a transformer input according to the “role” of its corresponding array as described in figures 2.19 and 2.20.

4.3: Tuners

In chapter 4.2, the design process of the tuning blocks was discussed, arriving at three separate multi-stage current-starve inverter designs; an NMOS-based duty cycle tuner, a PMOS-based duty cycle tuner, and a CMOS-based phase tuner which combines both aforementioned duty cycle tuners to shift both edges of a waveform. Selection of the appropriate blocks for each of the four RF waveforms driving the arrays was important, as maximum control should be achieved while minimizing complexity.

The proposed design has already been presented in figures 2.19 and 2.20. The system receives a differential square-wave RF waveform, meaning two separate inputs LO^+ and LO^- , each an inverted form of the other. According to the requirements of the “mirrored waveform” configuration, established in chapter 3.7.1, four waveforms $(v_\alpha, v_\beta, v'_\alpha, v'_\beta)$ need to be generated from these inputs. These waveforms need to reach the load through the SCPA arrays and transformers at a perfect duty cycle of $d = 50\%$, and with relative phases of:

$$\varphi_\alpha = 0^\circ, \quad \varphi_\beta = 240^\circ, \quad \varphi'_\alpha = 180^\circ, \quad \varphi'_\beta = 60^\circ \quad (4.4)$$

Phases φ_α and φ'_α are already set at a 180° difference by employing inputs LO^+ and LO^- . Thus, no phase tuner is needed in the path of v_α and v'_α . A case can be made for using a super-fine phase tuner on the v'_α in order to

ensure that the two do not drift apart in phase. However, if all circuitry is identical for both paths, no such issue should arise.

For phases φ_β and φ'_β , phase tuners are employed at the corresponding paths. Although it can be tempting to simply use an inverter to create v'_β out of v_β in order to only utilize one phase tuner, doing so will add extra delay and duty cycle corruption which is difficult to recover from. Instead, each path is equipped with a three-stage phase tuner, as described previously. The path of v_β receives LO^- as its input, as it is to be connected differentially to v_α , while v'_β is created from LO^+ .

4.3.1: Phase Detector

Duty cycle tuners are used for all four waveforms, as close-to-perfect duty cycles are required for proper harmonic suppression. A challenge is found in selecting between the PMOS- and NMOS-based current-starve designs presented previously. Although both influence the waveform duty cycle utilizing the same principle, their difference resides in the waveform edge being adjusted in each case.

If NMOS-based duty cycle tuners are used for all four paths, a problem arises regarding phase detection. In figure 4.6, qualitative plots for v_α and v_β are shown after both being subjected to NMOS-based duty cycle tuning. Their falling edges are shifted in time in order to facilitate duty cycle adjustment, and thus can be considered “unstable” edges, as they move during the tuning process. It should be noted that NMOS-based duty tuners in the context of this project influence the falling edge of the resulting output waveform as seen by the PA probe, since an odd number of inverter stages are employed.

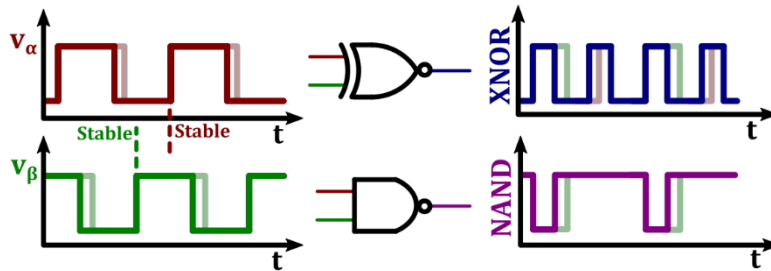


Figure 4.6: Illustration of Output Waveforms and Phase Detector Outputs for NMOS-Based Duty Cycle Tuner Implementation.

The issue can be seen when attempting to determine the phase difference between the two waveforms. A phase detector should only be able to discern the difference between the stable edges and discard the unstable ones, as taking them into account would result in false readings unless both duty cycles are perfectly equal. For example, if an XNOR or NAND gate is used as a phase detector, their resulting DC component fluctuates during duty cycle tuning, which can be perceived as “phase/duty coupling”. If simultaneous tuning of phase and duty cycle is to take place, this should be avoided.

The ideal logic function to be used as a phase detector in this case is $Y' = A + B'$, as it would only detect the area between the “stable” edges of figure 4.6. In practice, this would require an inversion of v_β , which might be troublesome to implement without adding extra delay to v_β compared to v_α prior to phase detection.

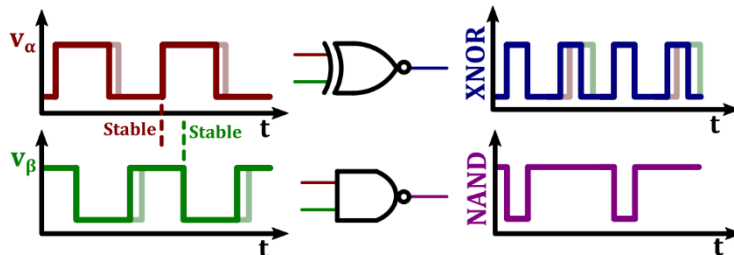


Figure 4.7: Illustration of Output Waveforms and Phase Detector Outputs for NMOS- and PMOS-Based Duty Cycle Tuner Implementation.

Instead, a PMOS-based duty cycle tuner is used for the path of v_β so as to influence its rising edge. In this case, a simple NAND gate can be used as a phase detector; its output remains constant throughout different settings for

the duty cycle tuners, as can be seen from the illustrations of figure 4.7. In this case, a duty cycle of $d = 5/6$ is seen at the output of the NAND gate for the desired phase difference of $\varphi_\beta = 240^\circ$. As with the previous configuration, an XNOR gate cannot be reliably used as a phase detector, since it still takes the “unstable” edges into account.

A PMOS-based duty cycle tuner is used for v'_β , with an NMOS tuner on v'_α . This ensures that a NAND gate can also be used in order to detect the phase difference between the two. Moreover, using identical NMOS-based tuners for both v_α and v'_α helps equalize the delay imposed upon both, so that they remain at a constant phase difference of 180° , as discussed previously. The same is true for the case of the PMOS tuners of v_β and v'_β , although only relevant for the purposes of minimizing tuner range as both are expected to suffer similar effects in the presence of specific variations.

4.3.2: Tuner Control

Figure 4.8 presents a block diagram of the implementation of tuners and phase detectors in this design.

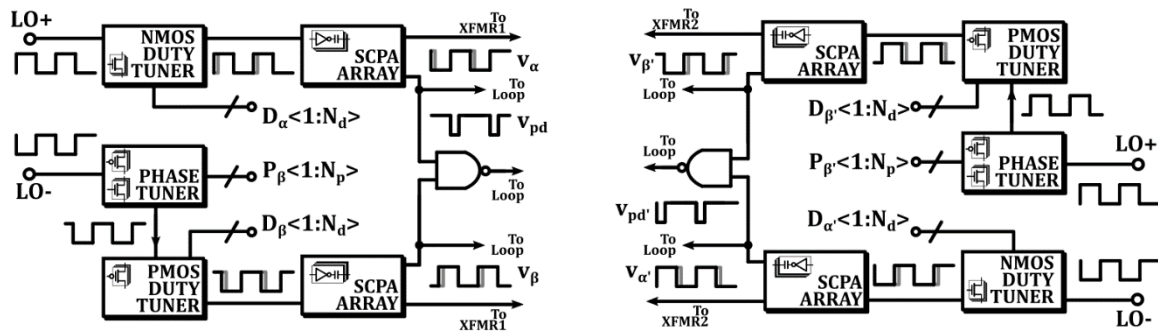


Figure 4.8: Block Diagram of Tuner Blocks and Phase Detectors.

Most of what has been already discussed is present in this diagram. The tuner blocks are pictured in greater detail in figures 3.31, 3.32, and 3.33. Six digital codes are used in total to control the tuners. The “SCPA array” block used is that of figure 4.5, with each block featuring a “probe” output leading towards the feedback loop, both directly and through a phase detector. Two identical NAND gates are used as phase detectors, so as to provide independent tuning of phase and duty cycle, as described previously.

Implementation of the tuners required a largely “trial-and-error” process. Although design equations 4.15 and 4.16 proved useful as a starting point, it was immediately apparent that different process corners reacted differently to different configurations. Increasing code D by one, for example, results in a much larger change in duty cycle for the SS (Slow NMOS, Slow PMOS) process corner, than for the FF (Fast NMOS, Fast PMOS) corner. This results in glitches at coarse transitions, in the form of either abrupt steps or instances of non-monotonicity.

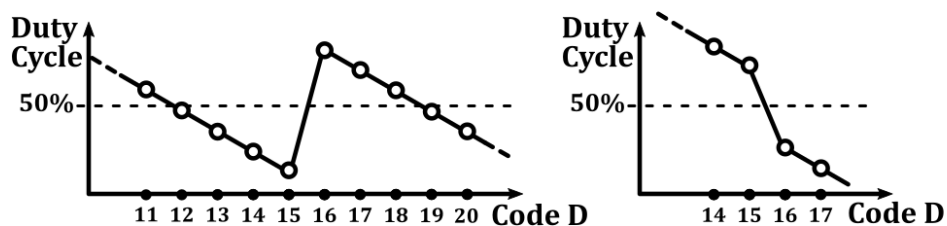


Figure 4.9: Illustrations of Tuning Curve Non-Idealities. Left: Non-Monotonicity. Right: Large Step.

It was discovered that non-monotonous steps posed a smaller threat to loop functionality than large steps in the right direction. Figure 4.9 displays qualitative plots of segments of the duty cycle tuning curve in both scenarios. If the desired duty cycle value of $d = 50\%$ is found during a non-monotonous transition, in this example between codes 15 and 16, it is likely that the same value is also found for codes D prior to and following the transition for non-monotonous glitches. The same is not true for large monotonous steps, for which the desired value can never be achieved in the same scenario.

Therefore, when adjusting the sizing of coarse tuning stages, priority was given to minimizing step size during coarse transitions, at the cost of non-monotonous transitions for certain process corners. Figure 4.10 shows simulated waveform duty cycles versus duty cycle tuner control code D for five process corners, for the NMOS-based tuner used in the project. This refers to the waveforms as seen at the PA probes, and therefore by the tuning loop.

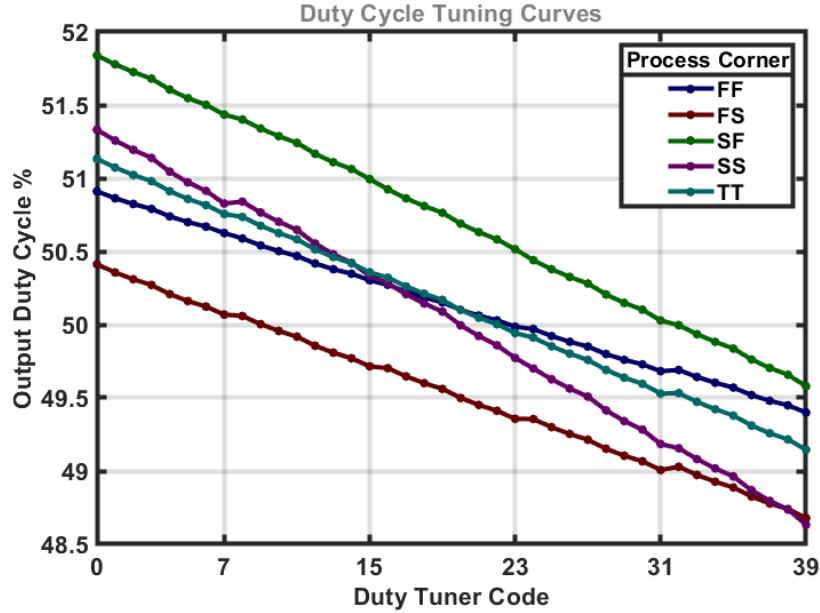


Figure 4.10: Simulated Tuning Curves for NMOS-Based Duty Cycle Tuner Across Process Corners.

The PMOS-based equivalent curves are almost identical albeit inverted, since enabling more PMOS units results in an increase in duty cycle. Sizing adjustment was performed independently for NMOS- and PMOS-based duty cycle tuners, although the final transistor sizes were largely similar between the two.

No explicit capacitor C_L was used in either design, relying in proper sizing of the succeeding inverter to provide appropriate capacitive loading. In order to center the curve towards $d = 50\%$, sizing was adjusted for the PMOS transistor of the NMOS-based coarse current-starve inverter (M_{pc} in figure 3.23) so as to provide a duty cycle offset. Inversely, the NMOS size was adjusted on the PMOS-based tuner.

Besides the aforementioned difference in slope, a significant offset seems to emerge for corners FS and SF, where pull-up and pull-down strengths are at their most diverse. In order for the system to be able to achieve $d = 50\%$ for every corner, the duty cycle tuning range was adjusted to 40 states, facilitated by 8 fine states multiplied by 5 coarse states. Thus, referring to the nomenclature used in figures 3.33 and 4.8:

$$N_{df} = 7, \quad N_{dc} = 4, \quad N_d = N_{df} + N_{dc} = 11$$

Thus, an 11-bit compound code $D < 1:11 >$ is used to control each duty cycle tuner, with components $D_f < 1:7 >$ and $D_c < 1:4 >$ being thermometer codes. An encoder was created in Verilog in order to translate decimal system numbers from 0 to 39 into the appropriate format.

Table 8.1, present in the appendix, displays how different decimal values are translated into tuner control codes for both NMOS- and PMOS-based duty cycle tuners.

4.3.3: Tuner Performance

From figure 4.10, and considering the nominal process corner TT, the duty cycle tuning range is $d_{max} - d_{min} = 1.987\%$, translating into a time shift of $8.28 ps$ for an LO frequency of $f_{LO} = 2.4 GHz$. Over 39 steps, this means that an average step size of $212 fs$ for the relevant waveform edge is achieved thanks to the current-starve architecture of the tuner.

Since non-linearities and harsh coarse transitions are present, some additional figures are relevant. For TT, the minimum fine-tuning step size is equal to $175 fs$, with a maximum step of $252 fs$. Coarse transitions range from a large step of $292 fs$ to a non-monotonous step of $11 fs$, which can be simply considered a redundancy.

The smallest fine-tuner step size is found for the FF corner, at 129 fs . The largest fine-tuner step occurs for the SS corner and is equal to 390 fs . The harshest monotonous coarse transition is equal to 315 fs , found in the SF corner. The largest non-monotonous coarse transition occurs for FS, at 91 fs .

Even at its point of highest non-linearity, a step of 315 fs would mean a worst-case deviation of 0.08% from the ideal duty cycle value. Based on the theoretical calculations plotted in figure 2.13, this would not be enough to cause any issues for suppression of either $HD2$ or $HD3$.

Despite the ability of the duty cycle tuners to correct duty cycle propagation errors due to process variation, it should be noted that the tuners provide sensitivity of their own. This means that the large disparity among corners seen in figure 4.10 would not be as significant in the absence of duty cycle tuners.

In fact, an early version of the design consisted of three current-starve stages for the duty cycle tuners. When the coarsest stage was added, it was revealed that it would cause high duty cycle disparity among corners, thus necessitating its own use for tuning range extension. After it was removed, the tuning range of the two-stage solution could still be made sufficient for all corners, as seen in figure 4.10.

4.3.4: Phase Tuners

The phase tuners were constructed using a three-stage design, as pictured in figures 3.31 and 3.33. The two finest stages, named “super-fine” and “fine”, were made by combining the tunable pull-down and pull-up networks of the NMOS- and PMOS-based current-starve inverters used for the duty cycle tuner designs, respectively. Thus, the same number of tuner states were used; 40 states facilitated by 8 super-fine and 5 fine states.

The coarsest stage, named “coarse”, consists of a tapped-inverter delay line featuring transmission gates, only one of which is active at a time, according to coarse phase code P_c in spatial unary format. A total of 8 coarse stages were used for the desired tuning range, bringing the total number of states to 320. Compound phase tuning code $P < 1: N_p >$ thus consists of:

$$N_{psf} = 7, \quad N_{pf} = 4, \quad N_{pc} = 8, \quad N_p = N_{psf} + N_{pf} + N_{pc} = 19$$

Table 8.2, present in the appendix, displays proper encoding of decimal values into phase tuner control codes.

The same principles as with the duty cycle tuner were also applied in this case. Non-monotonous coarse transitions were preferred over large monotonous ones, and the desired phase difference of $\Delta\varphi = 240^\circ$ was to be achievable for all process corners. The tuning curves were centered towards the desired value by adding a number of inverters prior to the coarse stage. The simulated phase tuning curves are displayed in figure 4.11.

In contrast with the duty cycle tuning curve of figure 4.10, the most problematic corners in this case are FF and SS, wherein both pull-up and pull-down delays are significantly lower and higher than typical, respectively. Corners FS and SF are more in line with TT, in terms of both slope and offset. It should be mentioned that the tuning curve for the SS corner is not properly displayed due to a simulation error. Only few points are available for coarse transitions.

More importantly, large monotonous steps can be seen for both fine and coarse transitions. This has been a result of poor implementation, which was brought upon by the challenging nature of this design.

The presence of a rigid tapped-inverter coarse stage added limitations to the design process. The delay provided for each coarse step depends on inverter delay, itself dependent on inverter on-resistance and capacitive loading, as described in equation 3.18. Inverter transistor width influences both parameters as per equation 4.1, and does not allow for unlimited reduction of inverter delay, as would be desired.

The fine stage needs to accommodate the step size of the coarse stage, as well as the range of the super-fine stage. As illustrated by figure 3.24, an upper limit exists on the number of steps that can exist per stage, since on-resistance is dominated by the inverter itself. Increasing the tuning range on this intermediate stage thus requires an increase in capacitive loading, implemented by proper sizing of the subsequent inverter.

Doing so would, however, also increase fine stage step size, which would then require an equal increase in the tuning range for the super-fine stage. Given that the same limitations apply on this current-starve stage as well, an increase in loading capacitance would also be required, thus sacrificing overall tuner resolution by increasing its minimum step size.

These factors were taken into account during the earliest iteration of the phase tuner design, as the pre-configured duty cycle tuner transistors were re-purposed for the phase tuners. This version would thus display much more palatable tuning curves than those of figure 4.11. However, an overhaul of all transistors used in the system was necessary due to a design oversight discovered late during the project, as a result of which the tuning curves were predictably damaged.

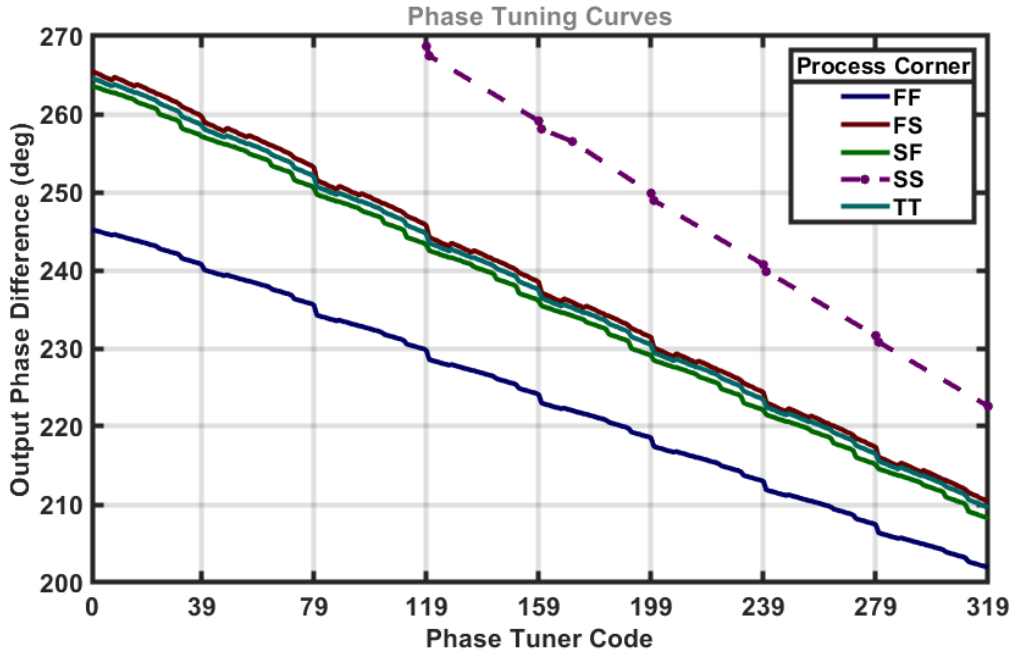


Figure 4.11: Simulated Tuning Curves for Phase Tuner Across Process Corners.

Although the duty tuners were easy to repair, the increased complexity of the phase tuners posed more of a challenge. Work on optimizing phase tuner curves continued throughout the final moments of the project, yet the issues pictured in the final version of figure 4.11 were not resolved in time.

A possible improvement would include increasing stage delay and reducing step number for the tapped-inverter stage, thus relegating it to a “super-coarse” stage, with a new “coarse” current-starve stage succeeding it. By increasing the available degrees of freedom, a smoother tuning curve could be achieved without sacrificing tuner resolution. However, this coarse stage would increase variation sensitivity, and would likely require high capacitive loading, with implications for “phase/duty coupling” as discussed in chapter 4.2.3.

Regardless, relevant characteristics of the phase tuning curve are as follows:

For TT and $f_{LO} = 2.4 \text{ GHz}$, a range of 55.08° is achieved over 319 steps for an average of 0.17° , or 200 fs per step, with a minimum super-fine-tuning step of 132 fs and a maximum of 220 fs , slightly smaller than its duty tuner counterpart. Fine transitions range from a large monotonous step of 641 fs to a non-monotonous 247 fs step.

Coarse transitions are highly non-ideal, with a maximum step of 1445 fs . Although the slopes and step sizes are rather consistent for different corners, it should be noted that FS contains an even harsher monotonous step of 1850 fs , whereas the largest step for corner SF is instead a more acceptable 1001 fs .

Even in this favorable case, however, a step of 1001 fs means a worst-case deviation of 0.72% from the ideal value. Based on the calculations of figure 2.13, suppression of the third harmonic would be inadequate if the ideal value were to be located within this step.

4.3.5: “Phase/Duty Coupling”

“Phase/duty coupling”, i.e., the effect of a change in phase code on duty cycle, can be quantified by measuring output duty cycle throughout the phase tuning range. This is pictured in the plots of figure 4.12.

For all corners, a visible change in duty cycle occurs as phase code is swept. Although negligible during super-fine changes, fine transitions cause significant jumps since, due to smaller baseline transistors, more radical changes in on-resistance occur for each step. Thus, mismatch between PMOS and NMOS on-resistances is more visible. Coarse transitions, on the other hand, effectively “reset” the duty cycle since all current-starve units are disabled as the next tapped-inverter step is enabled.

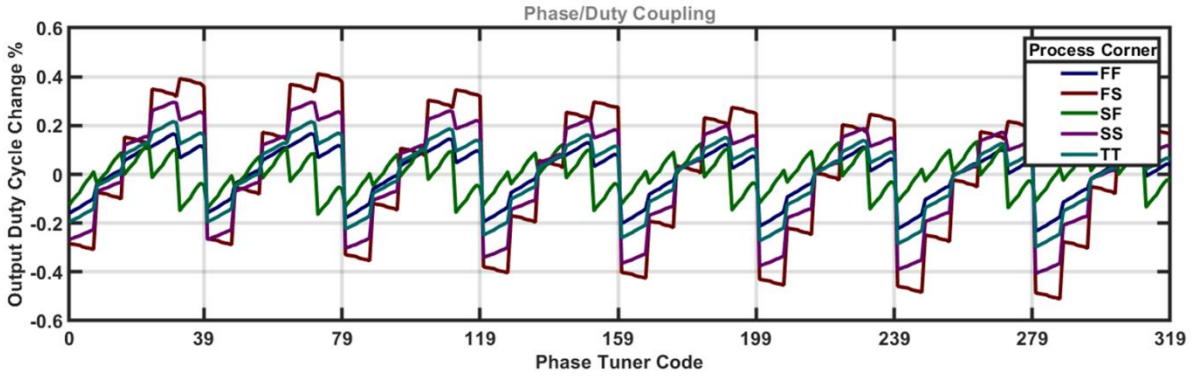


Figure 4.12: Simulated Plot of Duty Cycle vs. Phase Tuner Code, Across Process Corners.

Interestingly, the SF corner seems to suffer the least from this effect, signifying that the PMOS-to-NMOS width ratio used throughout the project might have been too low; duty cycle is better propagated when “faster” PMOS transistors are used. This is corroborated by the fact that FS, the opposite corner, displays the largest amount of “phase/duty coupling”.

In every case, however, this coupling effect is rather insignificant. Not only is the overall change small enough to be quickly recovered from by the duty cycle tuners, it is also usually small enough to be irrelevant as is. Apart from the most adverse of transitions, these deviations from the ideal duty cycle value are not significant enough to jeopardize harmonic suppression.

4.3.6: Harmonic Performance

Despite the shortcomings of the phase tuner, its performance with regard to harmonic suppression is adequate, at least when coarse transitions are avoided. Figure 4.13 displays plots for output $HD2$ and $HD3$, as extracted from simulations of the system for different process corners. Since two phase tuners exist for the four waveforms, code P_β for the left-hand tuner of figure 4.8 was swept, whereas $P_{\beta'}$ was kept constant, manually re-tuned for each corner. The positions of $P_{\beta'}$ are shown in figure 4.12 as dotted vertical lines.

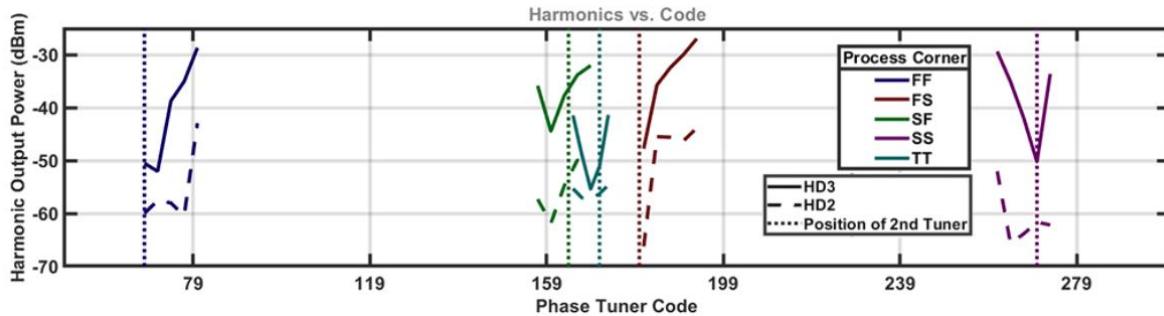


Figure 4.13: Harmonic Power vs. Phase Tuner Code, Across Process Corners.

From the above plots it can be seen that adequate harmonic suppression, meaning output component power of < -41 dBm for both second- and third-order harmonics, is achievable for all corners. Although the minima are not always found for $P_\beta = P_{\beta'}$ in this example, it is obvious that this could be achieved via a more precise selection of $P_{\beta'}$ values.

$HD2$ is consistently lower than $HD3$ in all cases, and any fluctuation of $HD2$ versus P_β is a result of “phase/duty coupling”, proven by these simulations to not be significant enough to threaten $HD2$ suppression. In the case of the FF corner, a coarse transition is present in the sweep range, yet $HD2$ remains within specification. The same is true for the SF corner, although it should be noted that the coupling effect is less potent at these corners according to figure 4.12.

In the case of TT, a total of eight values of P_β exist, for which $HD3 < -41$ dBm. This number is smaller for different corners, although this could be attributed to a mismatch between P_β and $P_{\beta'}$, or imprecise duty cycle tuning.

Figure 4.14 is generated from the same simulation, with the X-axis now representing measured phase difference between waveforms v_α and v_β . Interestingly enough, the phase difference value for which $HD3$ minima are found is somewhat lower than the ideal theoretical value of $\Delta\varphi = 240^\circ$, with all five corners achieving minimum distortion

for differences between 238.6° and 239.3° . This can be attributed to imperfections in the output waveforms, due to the PA transistors or the matching networks.

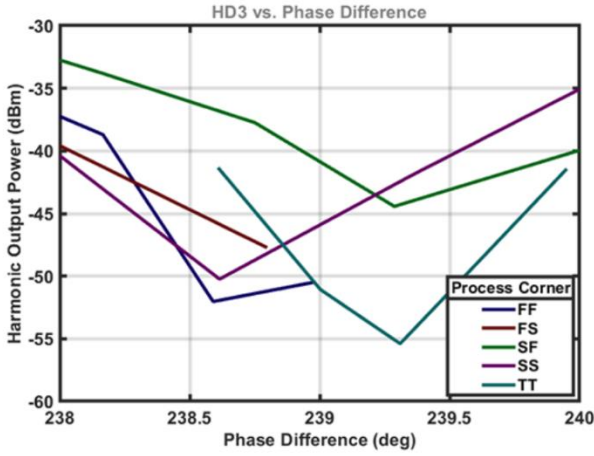


Figure 4.15: HD3 vs. Measured Phase Difference, Across Process Corners.

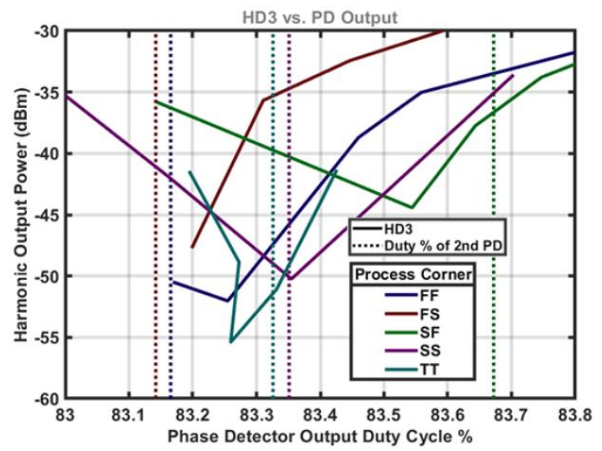


Figure 4.14: HD3 vs. Phase Detector Output Duty Cycle, Across Process Corners.

Since the desired value for $\Delta\varphi$ differs from the theoretical ideal, the sizing of the NAND gates used as phase detectors was adjusted to output a duty cycle of $d = 5/6$ at a phase difference of $\Delta\varphi = 239.6^\circ$, so that a voltage of $5V_{DD}/6$ can still be used as a reference for the loop, at least in TT conditions. Figure 4.14 presents a third set of plots from the same simulations as the two presented previously. In this figure, the X-axis represents the duty cycle of the phase detector output waveform v_{pd} . The dotted line displays the duty cycle of v_{pd} for each corner.

It can be seen that the goal of bringing the point of minimum $HD3$ towards $d_{pd} = 5/6$, or 83.33%, was mostly achieved for TT. The selected value for $P_{\beta'}$ in SS also results in $d_{pd'} \approx 5/6$, with an $HD3$ minimum at $d_{pd} \approx 5/6$ as well. The rest of the corners offer little relevant information due to $P_{\beta'}$ mismatch. However, this error has unintentionally presented a picture for phase mismatch tolerance. If, for example, an erroneous value of $d_{pd'} = 83.17\%$ is reached as a result of the feedback loop for FF, a “correct” value of $d_{pd} = 83.33\%$ between waveforms v_α and v_β will still keep $HD3$ within specification.

4.4: Feedback Loop

The tuning feedback loop, as described in chapter 4.3, is responsible for acquiring the duty cycle and phase difference of a pair of waveforms, and making the necessary adjustments to the tuners in order to achieve values as close to ideal for these characteristics.

According to a basic version of the logic pictured in the flowchart of figure 3.35, the loop must compare these characteristics to references, and increase or decrease the relevant tuning code until the reference value is reached. After this point all codes are saved, and later averaged, resulting in the final tuning codes used for transmission. A digital clock of $f_{CLK} = 40 \text{ MHz}$ was used for the logic.

The NAND gates of figure 4.8 output waveforms $v_{pd,pd'}$ whose duty cycle is related to the phase difference between its two inputs, and are thus used as phase detectors. Connecting a simple first-order RC low-pass filter at this output results in isolation of its DC component, which is relative to its duty cycle.

A comparator can then compare this DC component with a reference DC voltage, allowing the logic to determine whether the corresponding phase difference should be increased or decreased. Similarly, the duty cycles of waveforms $v_{\alpha,\beta,\alpha',\beta'}$ themselves can also be measured through separate LPFs, comparators, and references.

These comparators, along with all subsequent loop logic, were coded in Verilog and were not synthesized. “Implementation” of the loop, in this case, refers to designing the logic itself within the limitations of the system and should be viewed as a starting point for future research, rather than a complete feature of the system.

For the following descriptions, a generic loop branch dedicated to either a duty cycle tuner will be used as an example, outputting code C which can refer to either D or P . Duty cycle loop branches are almost identical to phase branches, and any differences between them will be noted.

4.4.1: Logic

A memory component is also needed in order to compare adjacent comparator outputs. If a “correct” value $C[n]$ for duty cycle or phase is reached, comparator output $CMP[n]$ is expected to change from “1” to “0”, or the inverse. An “edge detector” is thus utilized to signify the occurrence of this event. In the Verilog implementation, this has the form of a two-cycle accumulator, which will add together the two most recent comparator outputs: $ACC[n] = CMP[n] + CMP[n - 1]$. This has three possible results; “2”, “1”, and “0”.

Using duty cycle as an example, $ACC[n]$ values of “2” and “0” mean that the comparator output is consistently at “1” and “0” respectively, and thus the duty cycle must be reduced or increased accordingly. An accumulator output of $ACC[n] = “1”$ means that a “correct” value for duty cycle or phase can be found between the two most recent tuning code values: $D[n]$ and $D[n - 1]$. After this point, the tuning code would ideally be expected to oscillate around this “correct” value, even in the absence of an edge detector.

The edge detector is, however, used to trigger the “wait-and-average” function, as described in chapter 4.3. Although the delay between the tuner blocks and the PA probe is insignificant compared to a clock period, the RC low-pass filter present between the probe and the logic can provide large delays, depending on its time constant $\tau = RC$. A scenario similar to that of figure 3.38 may arise, in which case the averaging function becomes absolutely necessary.

In order to combat this delay even further, a “roll-back-and-hold” function was also implemented in the Verilog code. Considering that a delay equivalent to X clock periods is needed for a change in duty cycle or phase to be “noticed” by the comparator after going through the LPF, it can be assumed that an edge in the comparator output waveform was brought upon by the value used for the tuning code $C[n - X]$, X cycles prior.

Thus, in the presence of an “1” at the accumulator output $ACC[n]$, the code is rolled back by X cycles to its past value of $C[n - X]$. It is then held for X additional values, so as to allow time for the LPF to settle. Tuning resumes as usual after this point. The held value is essentially “tested” by serving as the starting point for the remainder of the tuning process.

After a number Y of occurrences of $ACC[n] = “1”$, all saved C values are averaged. The held values are not taken into account more than once during averaging, as that would skew the result towards possibly erroneous or inaccurate values. The “roll-back-and-hold” function only serves to reduce tuning duration by reducing the amplitude of the oscillation of C by effectively “pulling” it towards its presumed equilibrium position.

Moreover, different values are used for the gain G of the tuning during the process. G refers to the number by which code C is increased or decreased when deemed to be lower or higher than ideal, respectively; $G = |C[n] - C[n - 1]|$. Although a value for $G = 1$ provides maximum precision by changing the appropriate quantity by as little as possible, it can lead to long tuning times, especially in the case of the large number of steps present in the phase tuner.

Thus, a number of “gain modes” is implemented in the Verilog code. When tuning begins, an initial value of 0 is given to codes D and P , and the LPF capacitors are considered to be devoid of any charge. Thus, a large gain of G_1 is used during mode 1, reducing the time needed for the codes to approach their ideal values, as well as for the capacitors to charge towards their expected voltage. Each value for D or P is held for X cycles in this mode, so that the LPF may settle between code changes.

As soon as $ACC[n] = “1”$ during mode 1, the first instance of a “roll-back-and-hold” occurs, and mode 2 begins. An intermediate value of $G_2 < G_1$ is used for tuning, so as to quickly move towards the “correct” C value which has triggered the edge in the comparator waveform. When $ACC[n] = “1”$ reoccurs, mode 3 with $G_3 = 1$ is enabled, and fine-tuning begins, as described previously. The first instance of $ACC[n] = “1”$ during mode 3 triggers the “wait-and-average” function, which is finalized after Y rollbacks.

Figure 4.16 presents a block diagram of the feedback loop. Flowcharts and pseudo-code describing the logic are displayed in figure 8.1, present in the appendix. The loop contains three branches; two duty cycle branches, for an NMOS- and a PMOS-based duty cycle tuner, and a phase tuner branch. These are terminated by appropriate encoders, which operate according to tables 5.1 and 5.2.

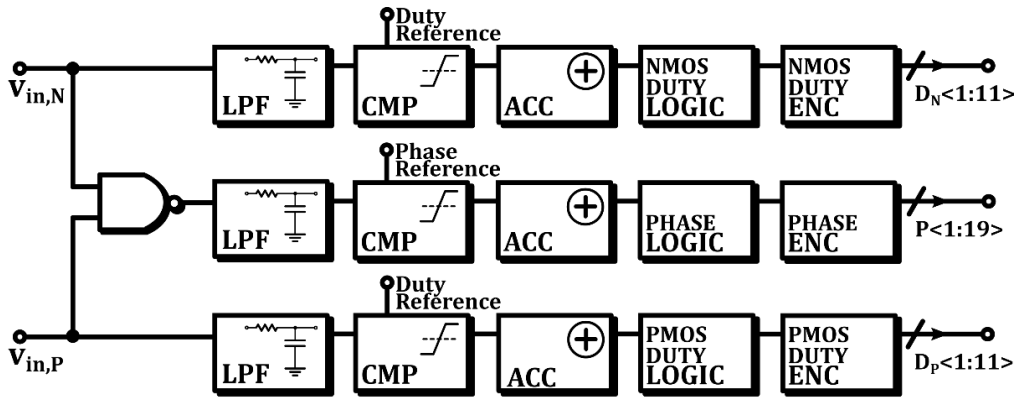


Figure 4.16: Block Diagram of Feedback Loop.

The logic flowchart of figure 4.16 condenses the description presented above. A difference between the three subtypes of logic is noted; in contrast with the phase and NMOS curves of figures 4.9 and 4.10, PMOS-based duty cycle tuners increase waveform duty cycle when D is increased. Therefore, the reaction of the logic to different accumulator outputs varies depending on the tuner being controlled. Although not pictured in the flowchart, limits also exist for the minimum and maximum code that can be outputted based on tuning range. These limits again depend on the tuner being controlled.

Flag S is enabled when C values start being stored. T refers to the number of rollbacks performed whilst in mode 3, and averaging takes place when $T = Y$. The waiting duration as the LPF settles is given in clock cycles by variable Z , which is equal to settling delay X , unless fine-tuning is being performed.

4.4.2: Parameters

The values for variables Y , G_1 , and G_2 were selected according to a self-imposed specification of under $5 \mu s$ for the tuning duration. In order to retain accuracy, parameter Y must be maximized so that more data is available for averaging. $Y = 12$ was found to be the upper limit for Y given the tuning duration constraints.

The remaining tuning time is used for modes 1 and 2, with their respective gains for the phase branch being $G_1 = 45$ and $G_2 = 9$. This configuration allows for quick approximations of the desired tuning values, while steering clear of phase coarse transitions. An alternative approach would be to use $G_1 = 40$ and $G_2 = 8$, the exact step sizes for coarse and fine phase transitions, but using an initial value of $P = 20$, thus keeping maximum distance from these problematic points. For the duty cycle tuners, $G_1 = G_2 = 2$ due to their small range.

Parameter X can be calculated as the number of clock periods needed to match the settling time of the low-pass filter. In order to decide on a value for X , LPF parameters must first be established. The purpose of the filter is to isolate the DC component of the probe waveform or the phase detector output, so that comparison with the DC reference can take place.

However, an RF component remains in the LPF output due to its finite cutoff frequency and roll-off. The magnitude of this component should remain low, so as to not threaten the accuracy of the loop. The qualitative plots of figure 4.17 can illustrate the trade-off between overlap and settling time. A low RC time constant for the LPF will result in quick settling, but the strong high-frequency component due to the increased bandwidth results in overlap between consecutive values of tuning code, thus reducing accuracy. On the other hand, setting RC too high can provide very distinct voltage levels among different codes at the cost of increased settling time, increasing total tuning duration.

An RC constant of $\tau = 50 ns$ was selected, facilitated by $R_{LPF} = 50 k\Omega$ and $C_{LPF} = 1 pF$. This setup allowed for minimal overlap between voltage levels, while keeping total tuning time under $5 \mu s$. A value of $X = 5$ was selected for the tuning loop. For a clock of $f_{CLK} = 40 MHz$, this means that the waiting period dedicated to settling is equal to 2.5 LPF time constants. Larger values for X would “over-correct” by rolling back to older, more inaccurate values of code C , not to mention a longer overall tuning duration.

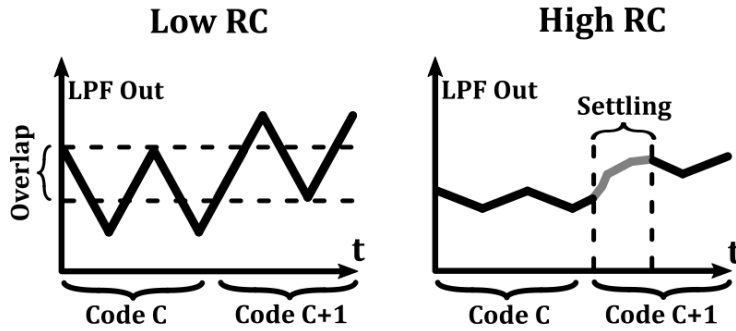


Figure 4.18: Illustrations of LPF Output for Low RC and High RC Configurations.

Due to the RF component present in the LPF output, another issue arises, as illustrated by the qualitative plots of figure 4.18. When the LO frequency is an integer multiple of the clock frequency, in this case for $f_{LO} = 2.4 \text{ GHz}$ and $f_{CLK} = 2.44 \text{ GHz}$, the two are “synchronized” in the sense that the clock edges always coincide with the same moment during the RF period. Thus, a “correct” value for code C will be “seen” by the comparator as consistently too high or too low, prompting an unnecessary correction.

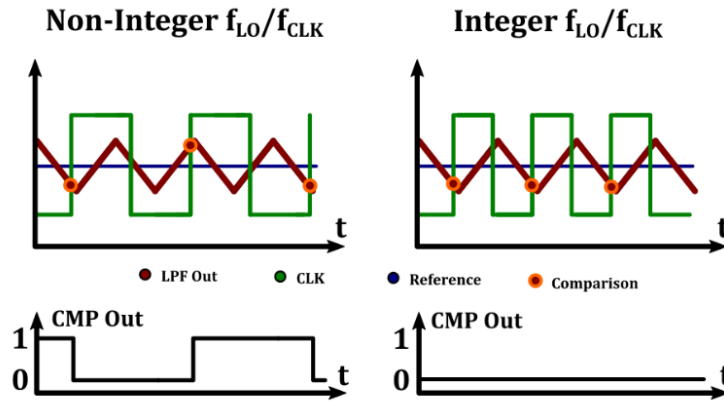


Figure 4.17: Illustrations of Relevant Waveforms for Non-Integer and Integer LO-to-CLK Frequency Ratios.

In contrast, the same value of C for a non-integer f_{LO}/f_{CLK} ratio will result in a desired oscillation at the comparator output, as the moments of comparison can “find” the LPF waveform either above or below the reference. Although this issue is most likely a result of the high ideality of the loop implementation, wherein the logic is in the form of Verilog code and no thermal noise is considered, it did pose a significant problem during design and is part of the reason for implementation of the “roll-back-and-hold” function.

The comparator references of $V_{ref,d} = V_{DD}/2$ and $V_{ref,p} = 5 V_{DD}/6$, referring to duty cycle and phase, were implemented using simple resistive voltage dividers. Although a more sophisticated bandgap-based reference or similar might be preferable in a complete implementation, it should be noted that any supply voltage variation in the present implementation would affect both the waveforms and the references in the same manner, thus allowing for comparison.

When the tuning loop is enabled, all six loop branches are tuned simultaneously and independently, in order to minimize tuning duration. A case could be made for a successive implementation, wherein the phase tuners can be operated first, and the duty cycle tuners would follow after phase codes have been finalized. This could be advantageous in the presence of “phase/duty coupling”, and could also allow for as little as one loop branch to be used for all tuning operations, depending on the time budget.

4.4.3: Operation

Simulation of the feedback loop was extremely time-consuming, due to the complexity of the system, simulation accuracy required, and overall long simulation times. Most of the debugging was performed using faster time-scales, referring to clock frequency, filter bandwidth, and transient simulation duration. A full-scale simulation was performed in typical conditions for $f_{LO} = 2.4 \text{ GHz}$, and lasted several days. The plots of figures 5.8 through 5.11

present the progressions of the six tuning codes throughout the tuning process as generated from the aforementioned simulation.

Figure 4.19 displays the codes controlling the six tuners. For the phase tuners, the initial “mode 1” high-gain region can be seen in effect until about $t = 1 \mu s$, at which point a desirable value is crossed, and the first rollback is performed. Mode 2 lasts for a single cycle, as both codes are reduced by an intermediate amount, and are quickly rolled back to their earlier values, signifying that the optimal code value is found close to $P = 180$ for both phase tuners.

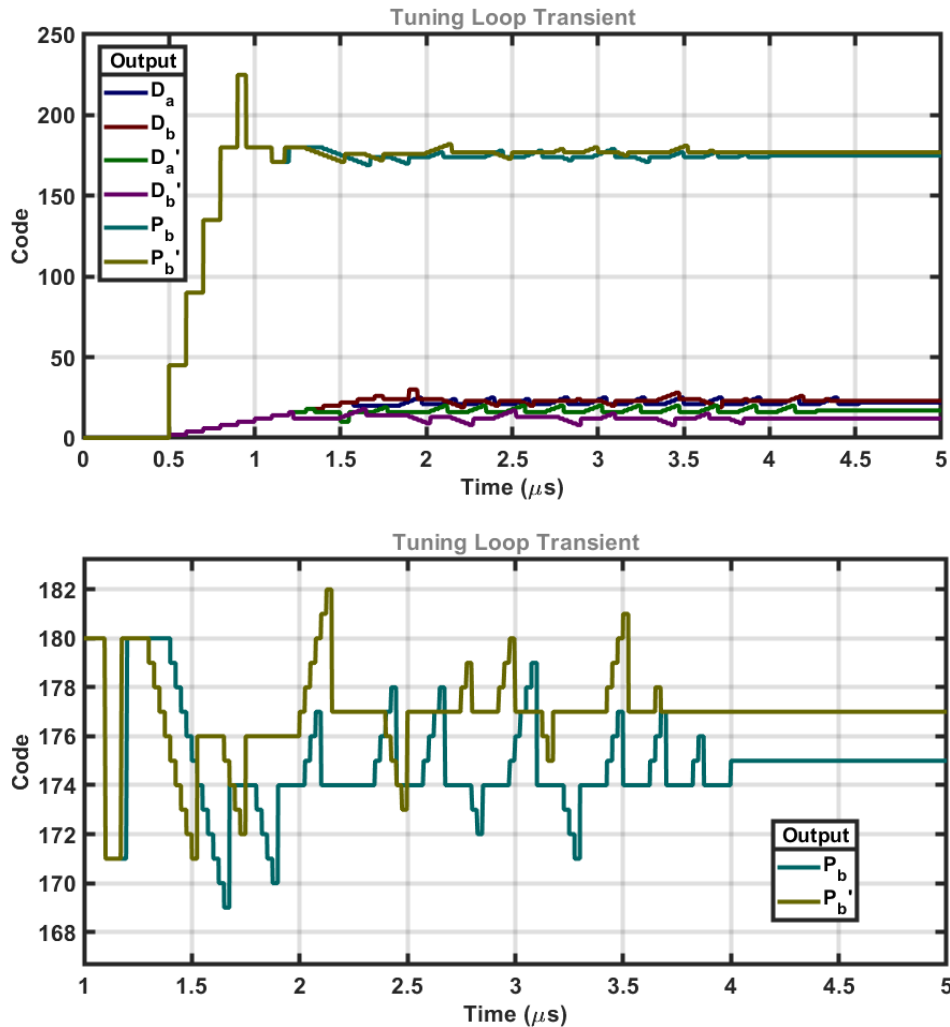


Figure 4.19: Simulated Time-Domain Plots of Phase Tuner Codes During Feedback Loop Tuning. Bottom: Zoomed In.

Fine tuning begins when “mode 3” is enabled, with each clock cycle inducing a reduction or increase of only $G_3 = 1$. In each case, a rollback of $X = 5$ cycles is performed when a desirable value is detected, and tuning resumes after another $X = 5$ cycles. Both code transients present a fair amount of oscillation, which is kept lower thanks to the “roll-back-and-wait” feature as established previously.

It can be seen that code P_β displays an upward tendency after rollback. This is taken into account by the averaging feature, as demonstrated from the fact that a higher value than the presumed “center of oscillation” is chosen as final phase code.

The corresponding plots for the four duty cycle tuner codes are seen in figure 4.20.

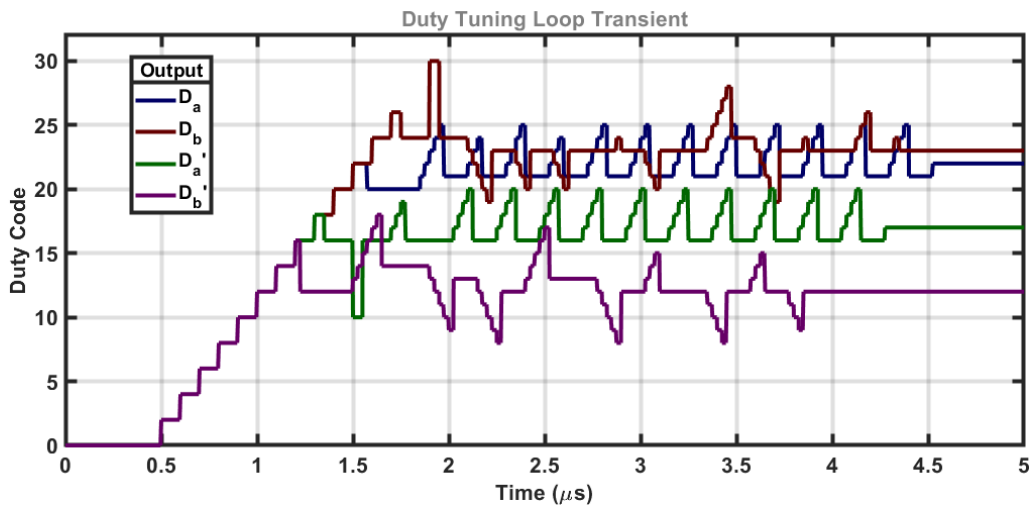


Figure 4.21: Simulated Time-Domain Plots of Duty Cycle Tuner Codes During Feedback Loop Tuning.

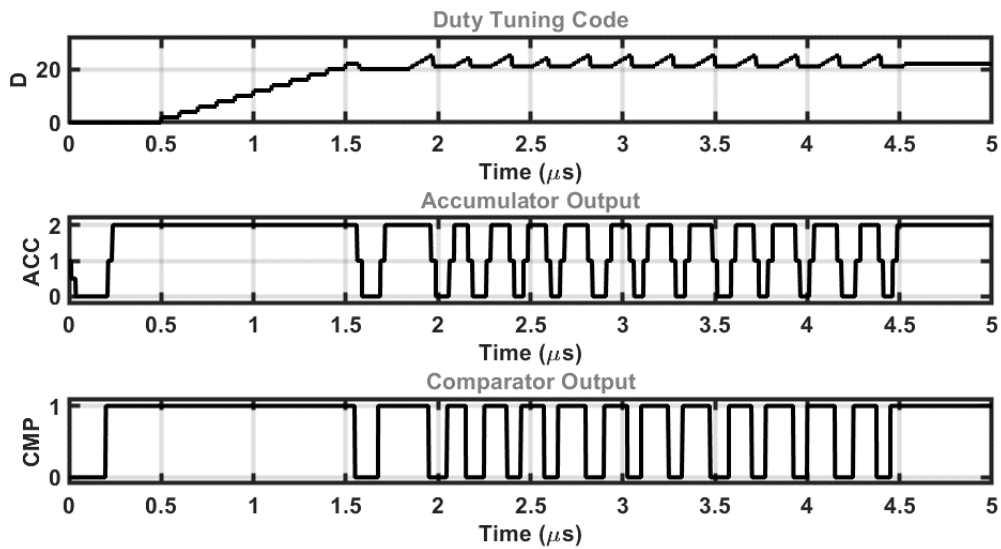


Figure 4.20: Simulated Time-Domain Plots of Code D_α and Related Accumulator and Comparator Outputs During Feedback Loop Tuning.

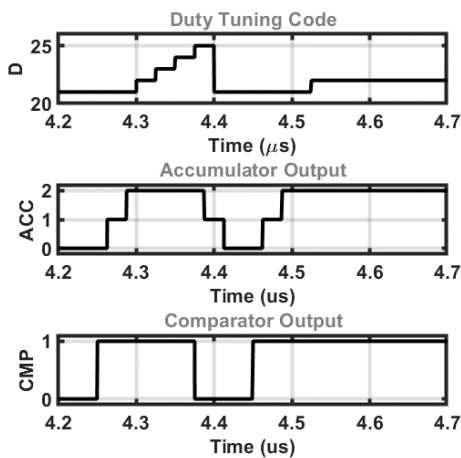


Figure 4.22: Zoomed-In Version of Fig.4.21.

Figures 5.10 and 5.11 display the progression for code D_α , along with the outputs of its related accumulator- and comparator-equivalent logic blocks. It can be seen that an edge in CMP triggers $ACC = 1$, which in turn triggers the “roll-back-and-wait” function. Instances of $ACC = 2$ outside the “wait” period prompt an increase in D_α , whereas all instances of $ACC = 0$ in this case have occurred within the “wait” period and are thus not taken into account.

For the two most prominent codes during the process, $D_\alpha = 21$ and $D_\alpha = 22$, it can be seen that the corresponding comparator outputs are $CMP = 0$ and $CMP = 1$, respectively. Thus, either one is viable as a final code value since a duty cycle of 50% can be found between the two. Here, the final code after averaging is $D_\alpha = 22$.

The remaining loop branches operate in the same way for all six tuners. The final codes result in decent harmonic suppression, although it is not presented here due to lack of extensive “full-scale” testing.

5: SIMULATION RESULTS

Implementation of the system was performed as described in the previous chapter. A number of simulations were performed to gauge its functionality and performance. Due to its adjustable nature, manual tuning of phase and duty cycle was performed in order to achieve the results presented below, with the feedback loop disconnected throughout.

5.1: PA Performance

Tables 5.1 and 5.2 present some relevant simulation results for an LO frequency of $f_{LO} = 2.4 \text{ GHz}$, at peak input amplitude, in the TT process corner and at a temperature of 25°C . The resulting output spectrum is shown in figure 5.1, and a period of the output time-domain waveform is displayed in figure 5.2.

Fundamental Output Power	12.62 dBm
2nd Harmonic Output Power	-56.27 dBm
3rd Harmonic Output Power	-51.07 dBm
System Efficiency	42.75%
Total Power Consumption	42.77 mW
Tuner Power Consumption	4.68 μW

Table 5.1: Simulated Figures for $f_{LO} = 2.4 \text{ GHz}$.

Waveforms	Duty Cycle	Phase Difference	Phase Detector Duty Cycle
v_α	50.055%	239.26°	83.326%
v_β	49.984%		
v'_α	50.023%	238.92°	83.332%
$v_{\beta'}$	50.068%		

Table 5.2: Simulated Waveform Characteristics for $f_{LO} = 2.4 \text{ GHz}$.

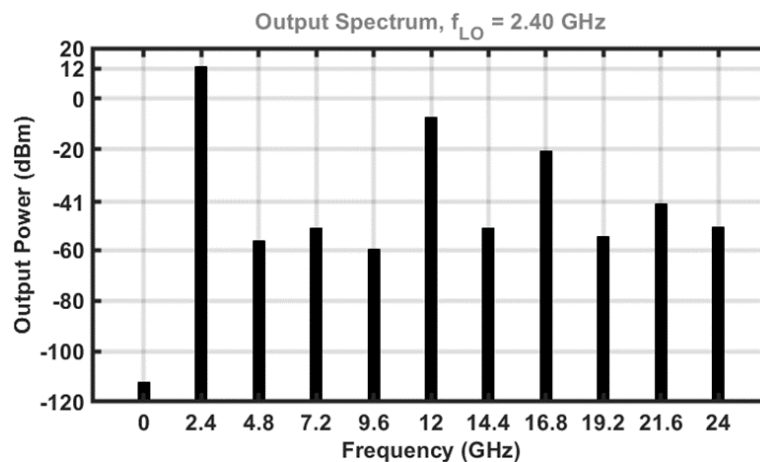


Figure 5.1: Simulated Output Spectrum for $f_{LO} = 2.4 \text{ GHz}$.

The system conforms to set specifications for fundamental output power $\approx 13 \text{ dBm}$, as well as harmonic distortion of $< -41 \text{ dBm}$ for the third harmonic, along with even-order harmonics. Predictably, odd-order harmonics other than the explicitly suppressed $HD3$ are rather high, due to inadequate filtering provided by the matching frequency response. This distortion can also be seen in the time-domain output waveform.

The system consumes 42.77 mW of power, with a peak system efficiency of 42.75%. Only $4.68 \mu\text{W}$ of power are consumed by the tuner blocks, with the rest dissipated by the PA arrays and passive components. Tuner codes have been adjusted for optimal duty cycle and phase, with both phase detectors displaying the desired duty cycle of $d_{pd} \approx 5/6$ in this configuration.

Tables 5.3 and 5.4, along with figures 5.3 and 5.4 present the corresponding results for $f_{LO} = 2.5 \text{ GHz}$.

Fundamental Output Power	11.99 dBm
2nd Harmonic Output Power	-46.03 dBm
3rd Harmonic Output Power	-56.02 dBm
System Efficiency	39.63%
Total Power Consumption	39.92 mW
Tuner Power Consumption	6.89 μW

Table 5.3: Simulated Figures for $f_{LO} = 2.5$ GHz.

Waveforms	Duty Cycle	Phase Difference	Phase Detector Duty Cycle
v_{α}	49.992%	236.76°	83.438%
v_{β}	49.631%		
v'_{α}	50.022%	236.47°	83.454%
$v_{\beta'}$	49.880%		

Table 5.4: Simulated Waveform Characteristics for $f_{LO} = 2.5$ GHz.

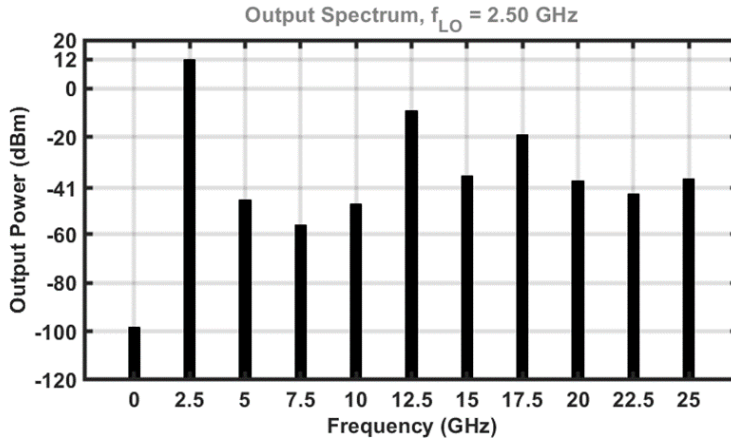


Figure 5.2: Simulated Output Spectrum for $f_{LO} = 2.5$ GHz.

An error unfortunately occurred during manual tuning of duty cycle for v_{β} and $v_{\beta'}$ for this simulation, resulting in a higher value for second harmonic output power, although it remains within specification. Due to this error, the remaining results do not offer much information, although it can be demonstrated that the system operates in a very similar fashion for both simulated values of f_{LO} .

5.2: AM Performance

Since the system operates as an RF-DAC, it is important to measure the effects of amplitude modulation. Figure 5.3 presents plots for system efficiency, fundamental output power, and total power consumption as a function of input amplitude, or code AM . These were generated from a simulation for $f_{LO} = 2.4$ GHz.

The system is highly linear, with a third-order intercept point of $IIP3 = 42.3$ dB, as can be seen from figure 5.6. Total power consumption is reduced for decreasing AM values, although not significantly. Comparing output fundamental power to total power consumption results in a figure for system efficiency. Peak efficiency is limited by dissipation in the active and passive components of the system, whereas switched-capacitor dissipation also plays a role at power back-off, leading to an efficiency of 22.52% at half input amplitude. Attempts to increase power back-off efficiency were made through proper impedance matching, although implementation was less than ideal.

Figure 5.4 displays time-domain plots of the output waveform for different values of code AM , where the amplitude can be seen to predictably increase for higher values of AM . However, zooming in at the zero-crossing produces the right-hand plots of figure 5.4, where an inconsistency can be seen for different values of AM . This can be considered $AM - PM$ distortion, since a receiver would use these zero-crossings to determine the phase of the waveform.

By that metric, $AM - PM$ distortion is quantified in the plot of figure 5.6. Over 3.5° of phase shift are measured throughout the AM range. Depending on the application, this could be problematic with regard to transmission quality.

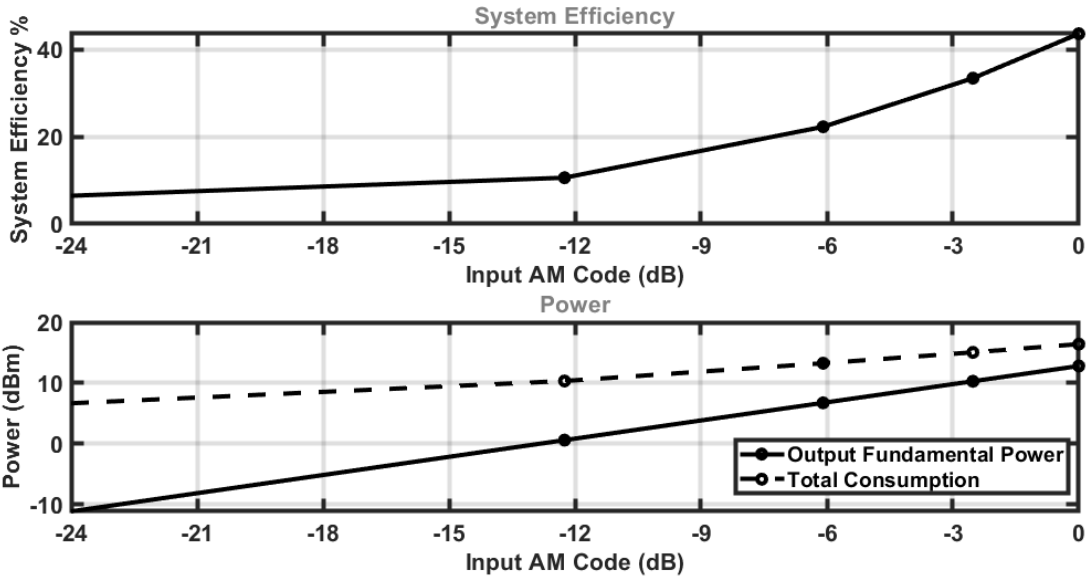


Figure 5.3: Simulated System Efficiency, Output Power, and Power Consumption.

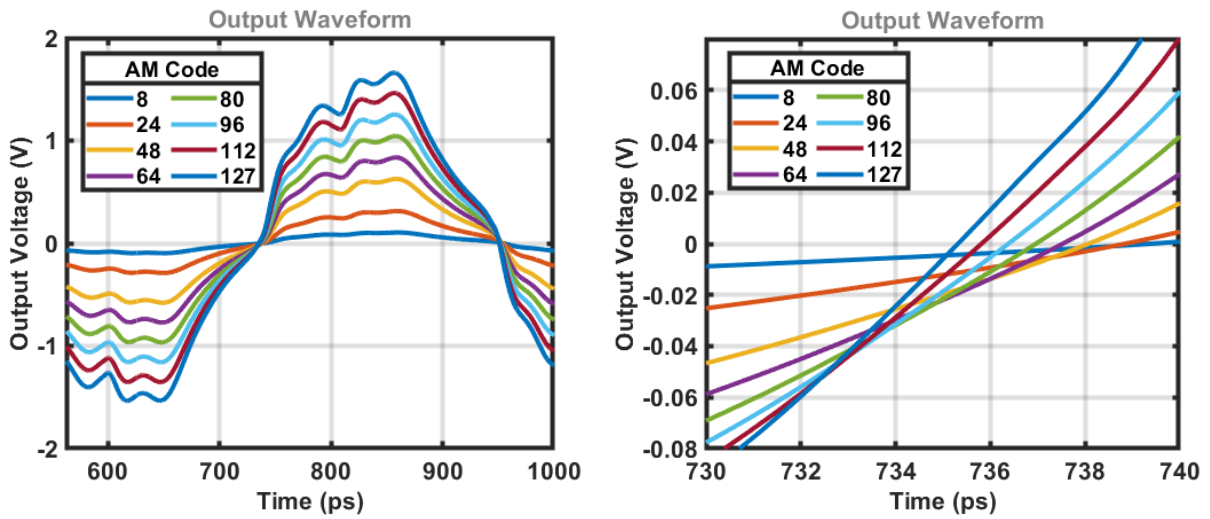


Figure 5.3: Simulated Output Waveforms for Different Input Levels. Right: Zoomed In.

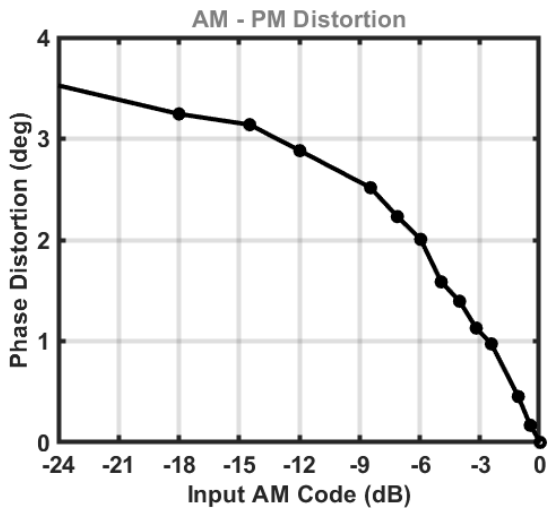


Figure 5.4: Simulated Plot of Phase Distortion vs. Input Amplitude.

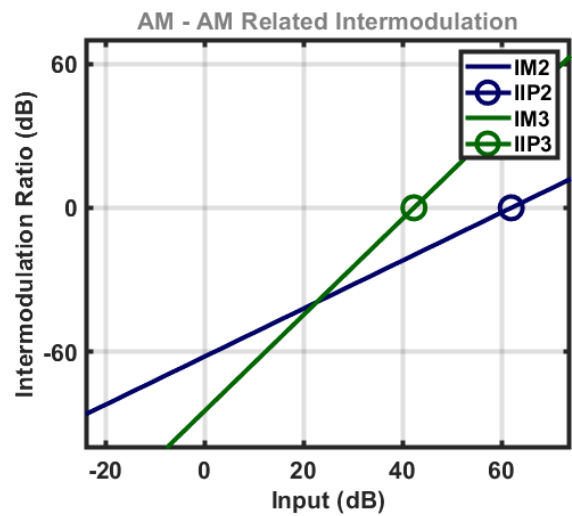


Figure 5.6: Second- and Third-Order Intermodulation Ratios.

6: CONCLUSION

6.1: Thesis Summary

The switched-capacitor power amplifier architecture offers a highly linear, highly efficient solution for integrated wireless RF transmitters. Consisting of digitally controlled units, the system operates neatly as an RF-DAC by employing capacitive voltage division. Two major issues are, however, inherent to this topology; reduced power back-off efficiency due to switched-capacitor power dissipation, and spectral impurity due to harmonics present in the pulse waveforms generated by the class-D units.

In this thesis project, a switched-capacitor power amplifier was designed in 22nm technology, operating between the frequencies of 2.4 GHz and 2.5 GHz, with a fundamental output power of 12.62 dBm. This thesis provides detailed documentation of the concepts, design decisions, challenges, and issues encountered during the project.

In order to suppress even- and third-order harmonics, the input RF waveform was manipulated in order to apply an appropriate phase shift, resulting in an output power of under -51 dBm for the second and third harmonic. Due to the high sensitivity to duty cycle and phase mismatch, precision tuners were designed, mostly employing a current-starved inverter topology. Despite some non-linearity issues related entirely to implementation, the tuners were able to achieve an average DTC resolution of below 212 fs under typical conditions.

As there is low tolerance for waveform amplitude mismatch due to transformer layout asymmetry, four waveforms were employed. These are used to drive the four PA arrays, which are connected to two output transformers. A total of six digitally controlled tuners are used for calibration of the four waveforms.

Since PVT variation can drastically alter the waveform characteristics, thus jeopardizing the quality of harmonic suppression, the tuners can be calibrated autonomously with the use of a feedback loop. This sub-system is equipped with digital logic designed as Verilog code and tested within the system, and is able to monitor the duty cycle and phase difference of the waveforms and perform appropriate adjustments to the tuning codes. In its current state, the loop is able to perform simultaneous calibration of all six tuners in under 5 μ s.

In order to increase power efficiency at lower input amplitudes, the “no shunt” impedance matching method employed in this project allowed for smaller capacitors to be used in the PA array units, thus reducing switched-capacitor power dissipation as well as required chip area. The absence of a shunt capacitor did, however, result in increased power for odd harmonics of the fifth order and above.

6.2: Comments and Recommendations for Future Work

The design proposed and described in this thesis cannot be considered a “*production-ready*” system, nor is it presented as such. Due to the wide scope of the project, focus was inevitably shared among multiple aspects during the process of research, design, and implementation. As a result, time constraints have limited the implementation to its current state.

This thesis aims to provide proof of concept for its main design components; namely the precision tuners, high-efficiency matching technique, and real-time calibration loop. However, a number of improvements can be proposed, pertaining to both optimizing the design as presented and building upon it, regarding potential future attempts at a complete implementation.

The primary design challenge regarding the tuning blocks is to improve linearity during coarse transitions. Tuning curves display different slopes for different process corners, and attempting to perfectly match fine tuners to coarse tuners for all corners might be all but futile. In the case of the duty cycle tuner curves, as seen in figure 4.10, care was taken to keep this inevitable non-linearity within tolerance, with preference for non-monotonous transitions over large, low-resolution steps. The same principle should be applied when attempting to repair the phase tuner curves of figure 4.11 by increasing the range of the fine/super-fine sub-curve. Despite the design expressions presented, the optimization process may unfortunately be delegated to “trial-and-error” when multiple process corners are involved.

Regarding the “no shunt” matching technique, the theoretical calculations and preliminary test simulations presented in this thesis have displayed potential for increasing power back-off efficiency, although the veracity of this promise should be further investigated. Regardless, the viability of this technique could also be an issue, given the fact that additional, lossy passive filtering should be employed due to the resulting pronounced high-order

harmonics. It is likely that a “partial-shunt” alternative could emerge as a compromise, resulting in increased array capacitance in order to improve matching frequency response as needed.

The “mirrored waveforms” method employed in this project introduced potentially unnecessary complexity to the design, as well as doubling the required chip area. It became necessary to include in the system due to layout mismatch concerns, however whether or not this was exacerbated by the matching technique was inconclusive. Making it possible to achieve similar harmonic performance by employing only two waveforms would be a very welcome improvement to the system. Achieving equal or higher output power should not be an issue considering the preliminary test simulation results displayed in picture YYY.

The output transformer used in the final design was selected in favor of other designs, as it allowed for decent performance regarding both efficiency and harmonics. Initial focus on “4-way” transformers was proven to be time- and resource-intensive, necessitating the use of a simpler solution, albeit one that does not necessarily comply to the “no shunt” paradigm as presented. Future research on this aspect could result in a viable matching configuration that operates further into the “quasi-current mode” region.

The feedback loop logic is only implemented as Verilog code which would need to be eventually converted to circuit level. Additional delays introduced should be accounted for in the logic code, since the “roll-back-and-hold” function parameters are related to total delay time. The tuning algorithm itself can be improved, perhaps with more focus on avoiding coarse transitions during the tuning process.

Finally, a complete implementation would require the design of a layout. Although a partial layout was designed for this project, post-layout simulation results were inconclusive albeit optimistic, and were not deemed important enough to present in this thesis. It should be noted that parasitic capacitance related to trace proximity can be catastrophic with regard to tuning curves, and should be minimized even at the cost of increased chip area.

7: REFERENCES

- [1] S. Yoo, J. S. Walling, E. C. Woo, B. Jann, and D. J. Allstot, "A Switched-Capacitor RF Power Amplifier," *IEEE J. Solid-State Circuits*, vol. 46, no. 12, pp. 2977–2987, Dec. 2011, doi: 10.1109/JSSC.2011.2163469.
- [2] R. Winoto, "Digital Radio-Frequency Transmitters: An Introduction and Tutorial," *IEEE Solid-State Circuits Mag.*, vol. 10, pp. 70–80, 2019, doi: 10.1109/MSSC.2018.2844605.
- [3] P. J. Green, "Peak-to-average power ratio and power amplifier back-off requirements in wireless transmissions," in *TENCON 2017 - 2017 IEEE Region 10 Conference*, 2017, pp. 630–633, doi: 10.1109/TENCON.2017.8227938.
- [4] A. Ba, V. K. Chillara, Y. Liu, H. Kato, K. Philips, and R. B. Staszewski, "A 2.4GHz class-D power amplifier with conduction angle calibration for –50dBc harmonic emissions," in *2014 IEEE Radio Frequency Integrated Circuits Symposium*, Jun. 2014, pp. 239–242, doi: 10.1109/RFIC.2014.6851708.
- [5] European Telecommunications Standards Institute, "Electromagnetic compatibility and Radio spectrum Matters (ERM); Wideband transmission systems; Data transmission equipment operating in the 2,4 GHz ISM band and using wide band modulation techniques; Harmonized EN covering the essential requirements of article 3.2 of the R&TTE Directive." 2012.
- [6] B. Yang, E. Y. Chang, A. Niknejad, B. Nikolić, and E. Alon, "A 65nm CMOS I/Q RF power DAC with 24–42dB 3rd harmonic cancellation and up to 18dB mixed-signal filtering," in *2017 Symposium on VLSI Circuits*, Jun. 2017, pp. C302–C303, doi: 10.23919/VLSIC.2017.8008517.
- [7] C. Huang, Y. Chen, T. Zhang, V. Sathe, and J. C. Rudell, "A 40nm CMOS single-ended switch-capacitor harmonic-rejection power amplifier for ZigBee applications," in *2016 IEEE Radio Frequency Integrated Circuits Symposium (RFIC)*, May 2016, pp. 214–217, doi: 10.1109/RFIC.2016.7508289.
- [8] V. Vorapipat, C. Levy, and P. Asbeck, "2.8 A Class-G voltage-mode Doherty power amplifier," in *2017 IEEE International Solid-State Circuits Conference (ISSCC)*, Feb. 2017, pp. 46–47, doi: 10.1109/ISSCC.2017.7870253.
- [9] M. Mizokami, T. Uozumi, Y. Yamashita, K. Shibata, and H. Sato, "A 43%-efficiency 20dBm sub-GHz transmitter employing rise-edge-synchronized harmonic calibration with 33.3% duty cycle," in *2017 Symposium on VLSI Circuits*, Jun. 2017, pp. C304–C305, doi: 10.23919/VLSIC.2017.8008518.
- [10] H. C. Luong and J. Yin, "Chapter 2: Transformer Design and Characterization in CMOS Process," in *Transformer-based design techniques for oscillators and frequency dividers*, Springer, 2016.
- [11] A. Ba et al., "A 1.3 nJ/b IEEE 802.11ah Fully-Digital Polar Transmitter for IoT Applications," *IEEE J. Solid-State Circuits*, vol. 51, no. 12, pp. 3103–3113, Dec. 2016, doi: 10.1109/JSSC.2016.2596786.
- [12] M. Maymandi-Nejad and M. Sachdev, "A monotonic digitally controlled delay element," *IEEE J. Solid-State Circuits*, vol. 40, no. 11, pp. 2212–2219, Nov. 2005, doi: 10.1109/JSSC.2005.857370.
- [13] B. I. Abdulrazzaq, I. A. Halin, S. Kawahito, R. Sidek, S. Shafie, and N. A. M. Yunus, "A review on high-resolution CMOS delay lines: towards sub-picosecond jitter performance," *SpringerPlus*, vol. 5, 2016.
- [14] S. Kak, "Spread Unary Coding," *ArXiv*, vol. abs/1412.6122, 2014.
- [15] J. S. Park, S. Hu, Y. Wang, and H. Wang, "A Highly Linear Dual-Band Mixed-Mode Polar Power Amplifier in CMOS with An Ultra-Compact Output Network," *IEEE J. Solid-State Circuits*, vol. 51, no. 8, pp. 1756–1770, Aug. 2016, doi: 10.1109/JSSC.2016.2582899.
- [16] Y. Yin, L. Xiong, Y. Zhu, B. Chen, H. Min, and H. Xu, "A compact dual-band digital doherty power amplifier using parallel-combining transformer for cellular NB-IoT applications," in *2018 IEEE International Solid - State Circuits Conference - (ISSCC)*, Feb. 2018, pp. 408–410, doi: 10.1109/ISSCC.2018.8310357.
- [17] L. Ye, J. Chen, L. Kong, E. Alon, and A. M. Niknejad, "Design Considerations for a Direct Digitally Modulated WLAN Transmitter With Integrated Phase Path and Dynamic Impedance Modulation," *IEEE J. Solid-State Circuits*, vol. 48, no. 12, pp. 3160–3177, Dec. 2013, doi: 10.1109/JSSC.2013.2281142.
- [18] W. Simb, A. Scholtz, and D. Kehrer, "Design of Monolithic Integrated Lumped Transformers in Silicon-based Technologies up to 20 GHz," 2000.

8: APPENDIX

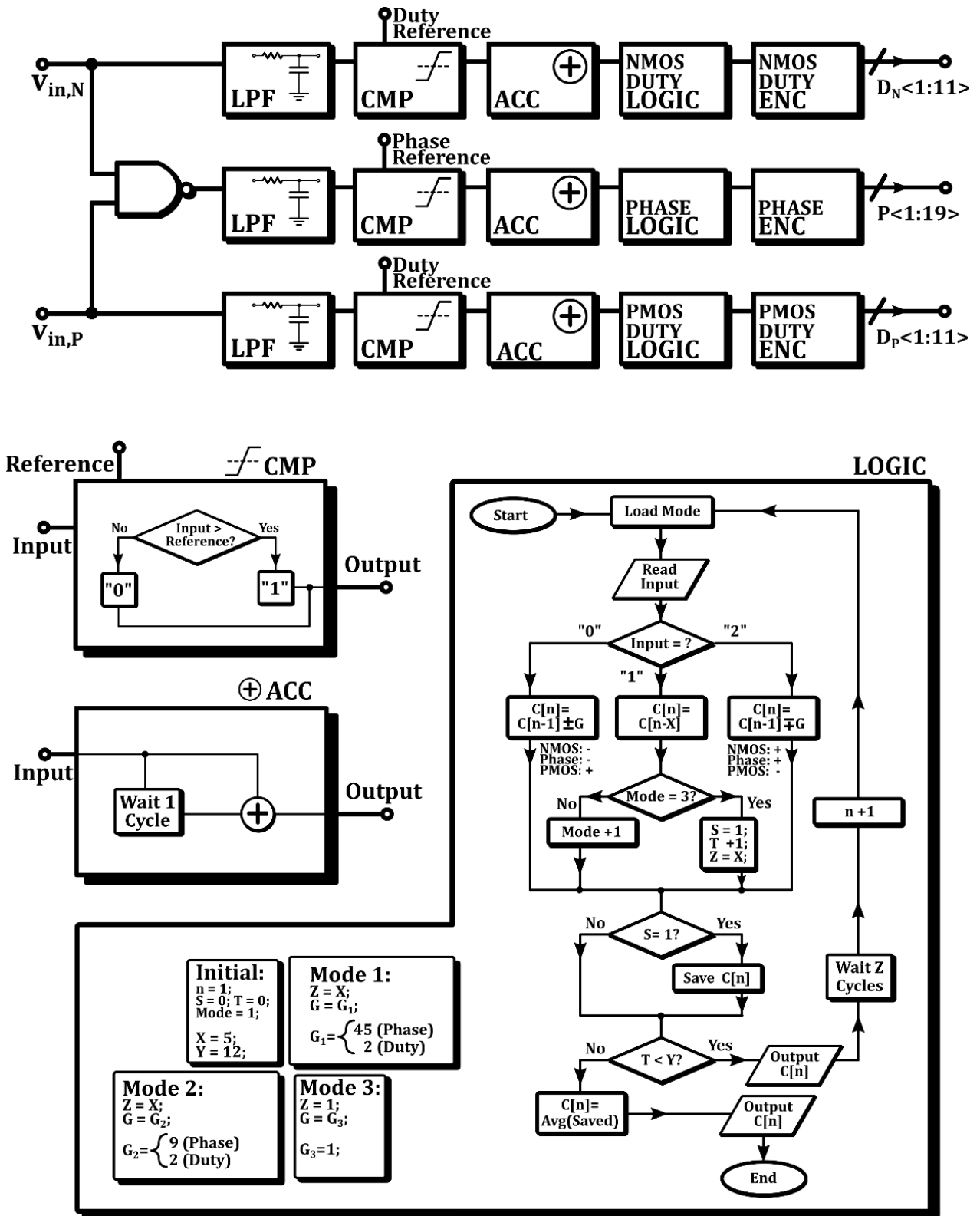


Figure 8.1: Block Diagram of Feedback Loop, with Flowchart and Pseudo-Code of Logic.

Code D	NMOS Duty Tuner			PMOS Duty Tuner		
Decimal	Coarse	Fine	Compound D_n	Coarse	Fine	Compound D_p
0	0000	0000000	00000000000	1111	1111111	11111111111
1	0000	0000001	00000000001	1111	1111110	11111111110
2	0000	0000011	00000000011	1111	1111110	11111111100
7	0000	1111111	00001111111	1111	0000000	11110000000
8	0001	0000000	00010000000	1110	1111111	11101111111
9	0001	0000001	00010000001	1110	1111110	11101111110
15	0001	1111111	00011111111	1110	0000000	11100000000
16	0011	0000000	00110000000	1100	1111111	11001111111
17	0011	0000001	00110000001	1100	1111110	11001111110
23	0011	1111111	00111111111	1100	0000000	11000000000
24	0111	0000000	01110000000	1000	1111111	10001111111
31	0111	1111111	01111111111	1000	0000000	10000000000
32	1111	0000000	11110000000	0000	1111111	00001111111
39	1111	1111111	11111111111	0000	0000000	00000000000

Table 8.1: Encoding Table for Duty Cycle Tuners.

Code P	Phase Tuner			
Decimal	Coarse	Fine	Super-Fine	Compound P
0	00000001	0000	0000000	00000001000000000000
1	00000001	0000	0000001	00000001000000000001
7	00000001	0000	1111111	00000001000011111111
8	00000001	0001	0000000	00000001000100000000
9	00000001	0001	0000001	00000001000100000001
16	00000001	0011	0000000	00000001001100000000
32	00000001	1111	0000000	00000001111100000000
39	00000001	1111	1111111	00000001111111111111
40	00000010	0000	0000000	00000010000000000000
79	00000010	1111	1111111	00000010111111111111
80	00000100	0000	0000000	00000100000000000000
120	00001000	0000	0000000	00001000000000000000
280	10000000	0000	0000000	10000000000000000000
319	10000000	1111	1111111	10000000111111111111

Table 8.2: Encoding Table for Phase Tuners.