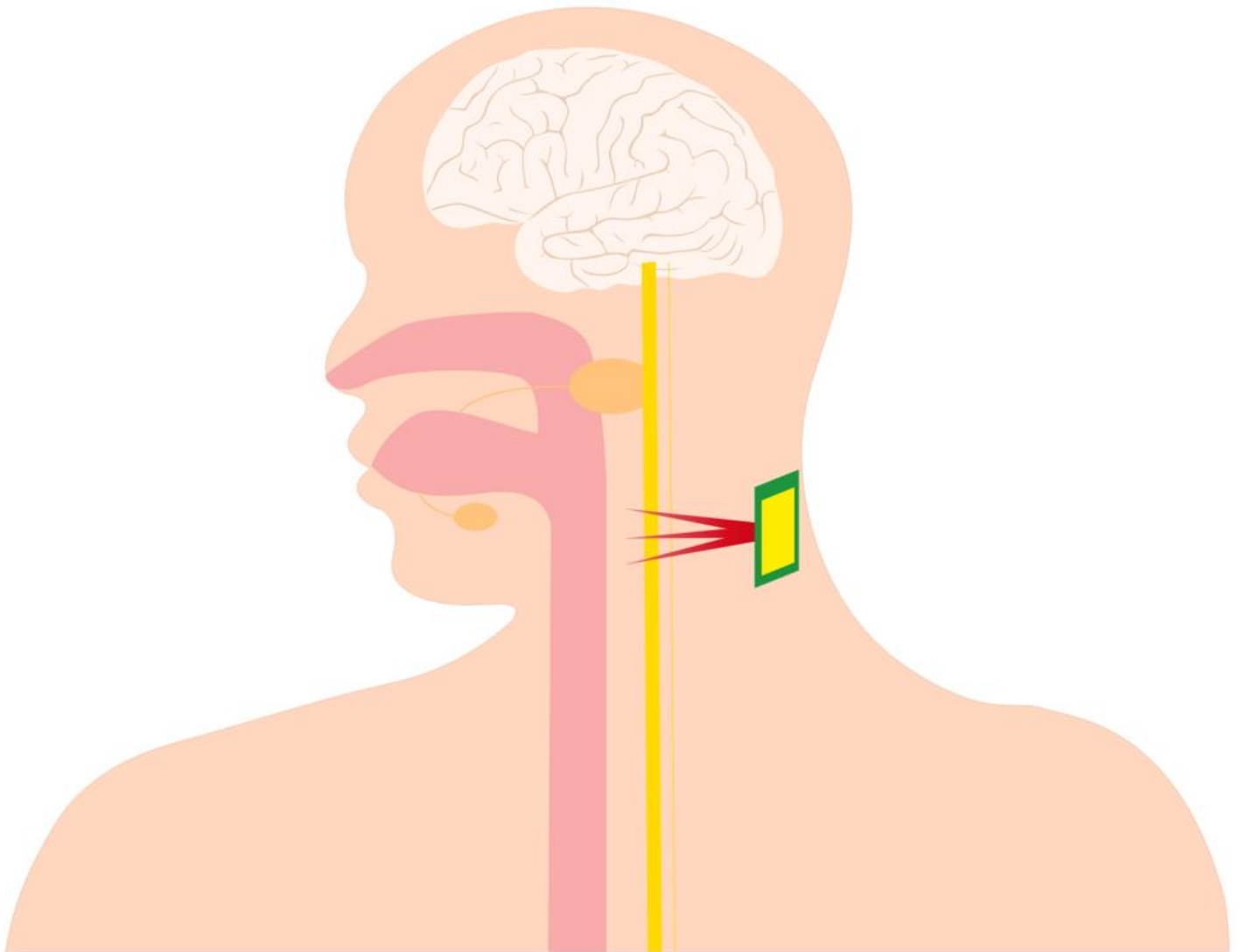Xinyu Yang

Ultrasound Transmitter for Non-invasive Vagus Nerve Stimulation

Ultrasound Transmitter for Non-invasive Vagus Nerve Stimulation

By

Xinyu Yang

in fulfilment of the requirements
for the degree of

Master of Science

in Electrical Engineering
at the Delft University of Technology

Student ID: 4873092
Supervisor: Dr. Tiago Costa, Prof. dr. ir. Wouter Serdijn
Thesis committee: Prof. dr. ir. Wouter Serdijn,   Bioelectronics, TU Delft
                       Prof. dr. ir. Ronald Dekker,   ECTM, TU Delft
                       Dr. Tiago Costa,   Bioelectronics, TU Delft

An electronic version of this thesis is available at http://repository.tudelft.nl/.

TUDelft

Content

List of Figures

List of Tables

Abstract

The main objective of this thesis is to design a transmitter channel which is implemented in a 2D phased-array ultrasound neuromodulation system aiming for the vagus nerve stimulation.

The vagus nerve has the widest distribution in the parasympathetic system and it oversees several important bodily functions such as the control of mood, immune response, digestion, and heart rate. The potential of vagus nerve neuromodulation in the medical area is huge with these functions and has been studied for decades. The traditional way of vagus nerve stimulation (VNS) is electrical stimulation which requires surgery to implant the stimulator under the skin. It has been proven that electrical VNS has therapeutic effects on many diseases such as rheumatoid arthritis, refractory epilepsy and depression. However, there are still several limitations to electrical VNS. Inside the vagus nerve, there are several fibers and some unwanted fibers will be stimulated because the spatial resolution of electrical VNS is relatively low. Side-effects including cough, throat pain, voice alteration, and dyspnea are caused due to inadequate spatial resolution. Compared with the electrical approach, focused ultrasound wave could provide a high spatial resolution and does not require any surgery. It has been proven that the ultrasound stimulation could modulate the neuron activity and applications have been conducted in the stimulation for the central and peripheral nervous system, including the vagus nerve. Using a 2-D phased array, the ultrasound VNS system could also tune its focal spot in a 3D space without any mechanical controlling stage, which enables the VNS device to be portable while remaining its precision. The detailed modulation mechanism of the VNS is still under discovery and the ultrasound VNS devices are of great potential with the advantages mentioned above. In an ultrasound VNS system, the transmitter is an essential section which determines the accuracy of spatial resolution and the amplitude of focal pressure. In this work, an ultrasound transmitter channel circuit is designed to be implemented in a 2-D phased array ultrasound VNS system.

Keywords: vagus nerve neuromodulation, ultrasound, transmitter, 2-D array

Acknowledgements

The past two years I have spent at TU Delft was a fantastic journey and an unforgettable memory in my life. During these two years, I have been helped by many people, my respected teachers, my dear classmates and my beloved parents. When my graduation thesis is about to be completed, and when I am about to finish my master program study, I would like to express my gratitude to the mentors, friends, and family members who supported and helped me in these two years.

First of all, I would like to thank Dr. Tiago Costa, who guided my work in my graduation thesis, for his sincere teaching, concern and help. He is my daily supervisor at TU Delft and provided me not only the guidance in the thesis work but also the consultative talk when I was nervous. It would impossible for me to accomplish the entire design in the past year without him.

Secondly, I would like to express my thanks to Prof. dr. ir. Wouter Serdijn, the chairman of the bioelectronics group at TU Delft. With his broad knowledge and experience in the circuit design, he always enlightens me during the monthly meeting.

I also want to thank the members of the bioelectronics group especially members of the ultrasound team. It is lucky to work in such a good team with talented and enthusiastic students. The weekly meeting provides a wonderful space for us to discuss our work and it inspired me a lot in my work.

Finally, I want to thank my parents. They provided me with both mental and financial support in the past two years. Although we could only communicate through the phone when I was in Nederland, their care and smiles are the largest motivation for me. After I went back to China, they kept on encouraging and warming me to help me finish the thesis work. I appreciated a lot and I am quite proud of having them in my life.

Chapter 1 Introduction

In this chapter, the essential effects and the potential of the vagus nerve neuromodulation will be illustrated first, followed by the introduction of different methods on the vagus nerve stimulation (VNS). After the comparison, the ultrasound VNS appears to be a better approach than the traditional electrical VNS. The possible mechanism and essential parameters of the ultrasound VNS will be explained in details. The transmitter in an ultrasound VNS system is of great importance and its design is the main purpose of this thesis work. Prior arts of the transmitter and the design specifications will be demonstrated after introducing the architecture of the transmitter. At last, the organization of the whole thesis is shown.

1.1 Vagus nerve
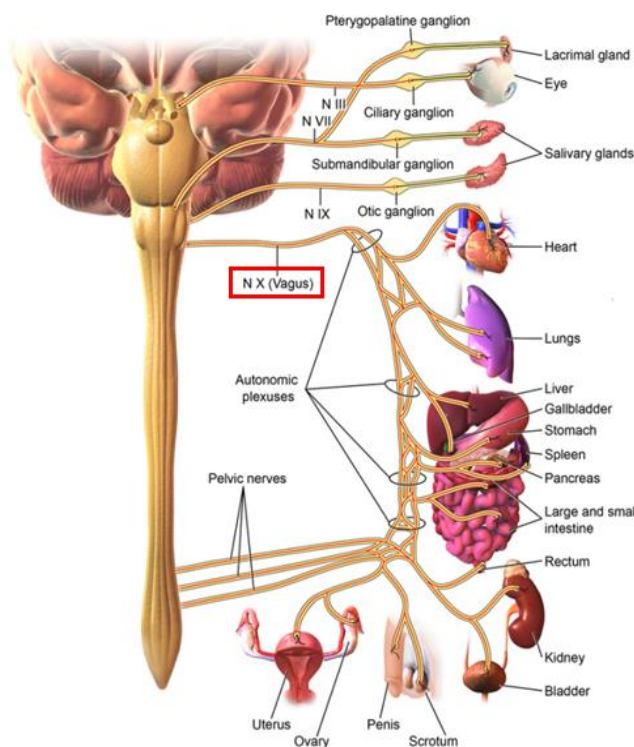
Organ	Function
Heart	Slows heart rate
Lungs	Cough reflex
Oesophagus	Modulation of peristalsis
Stomach	Regulation of gastric motility
Liver	Sense glucose levels; Syppress hepatic glucose production
Pancreas	Regulatory roles in the secretion of insulin and pancreatic exocrine
Small intestines	Modulation intestinal microenvironment for intestinal immune homeostasis

Figure 1.1 Vagus Nerve in Parasympathetic Nerve System [1] and related function[2]

The Vagus nerve is the 10th cranial nerve and the main contributor to the parasympathetic nervous system. It has the widest distribution in the body, as shown in Fig 1.1, from the brain to the abdomen, connecting the brain to the heart, lungs, stomach, and many other organs. These connections are built by the somatic and visceral afferent fibers, as well as general and special visceral efferent fibers in the vagus nerve. Thus, a vast array of crucial bodily functions are overseen by the vagus nerve, including control of mood, immune response, digestion, and heart rate[3]. The organs that the vagus nerve innervates and the related functions[2] have been studied by scientists are summarized and listed in Figure 1.1.

With these functions, the potential of vagus nerve neuromodulation in the medical area is huge and has been studied for decades. A relatively mature application is the VNS using in the treatment of refractory epilepsy and depression[4], which have also been approved by the Food and Drug Administration (FDA) in 1997 [5] and 2005 [6] separately. Further researches also illustrate the therapeutic effects of VNS in hypertension heart failure, rheumatoid arthritis, tinnitus, and stroke rehabilitation[7]. However, the inside mechanism of vagus nerve neuromodulation is not fully established and the efficacy varies from patient to patient with some side effects including cough, throat pain, voice alteration, and dyspnea[8]. Thus, the improvement in the VNS methods or devices would be vulnerable for a further understanding of the vagus nerve modulation and fewer side effects, which is also the motivation of this work. Historically, the VNS is conducted by an electrical stimulator, but focused ultrasound has also emerged to be an approach. The details and differences between these two methods are explained below.

1.2 Vagus nerve stimulation

1.2.1 Electrical VNS
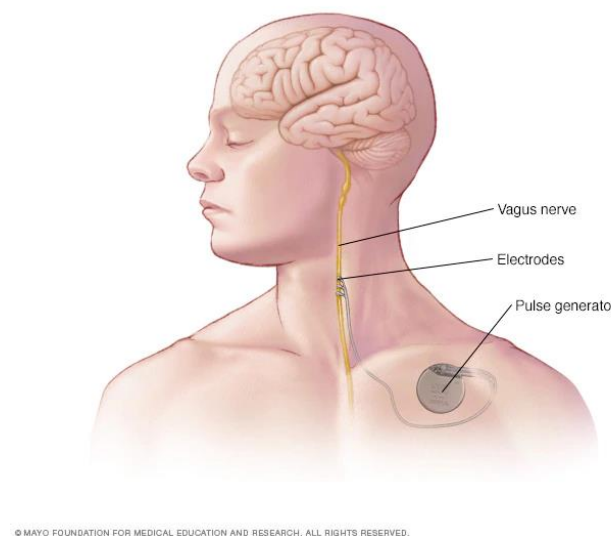
Figure 1.2 Electrical VNS[9]

As shown in Fig.1.2, normally an electrical stimulator is surgically implanted under patients' skin. The pulse generator delivers electrical pulses to cuff electrodes wrapped around the vagus nerve, near the neck where the vagus nerve is most accessible. The control of output current, signal frequency, pulse width, and on-time of signal could be tuned for different patients or different extent therapy. There are several commercial implantable VNS products, such as the SenTiva VNS Therapy by LivaNova[10] and the VNS Therapy system by Cyberonics[11]. The control of seizures improved over time[10], and according to[11], more than 60% of patients with refractory epilepsy gained 50% or more reduction in seizures.

However, there are several limitations of electrical VNS. First, it requires surgery, which might introduce more risk to the patients. Charging or replacing the battery would be a further problem. Second, currently, the spatial resolution of electrical stimulation is not high enough. Since the vagus nerve inside contains many fibers, or neuron axons, which execute different functions, a coarse stimulation activates some unwanted fibers and affects both the discovery of the vagus nerve modulation mechanism and the treatment effect. Although a non-invasive electrical VNS device also emerged (the gammaCore Sapphire[12]), the efficacy is affected, and spatial resolution is even worse when electrical signals have to pass through the skin. Besides, as a handheld device for patients, the gammaCore could not ensure the stimulation area is fixed on the vagus nerve during therapy. A real-time system would be preferred when VNS therapy is conducted. Despite the success of electrical VNS in the treatment of epilepsy and depression, more advanced VNS methods need to be developed for better spatial selectivity and neuromodulation efficacy. Ultrasound VNS emerged to be a good option.

1.2.2 Ultrasound VNS

Ultrasound is a pressure wave with a frequency higher than the human hearing range (>20kHz). Ultrasound is widely used in the medical area as a diagnostic imaging modality since the 1970s, and its potential for neuromodulation has also been developed in the past few decades. When the ultrasound wave is propagating in biological tissue, both mechanical and thermal effects are caused and could be used for therapeutic application[4]. High-intensity focused ultrasound (HIFU, intensity>200W/cm^2) has been used for non-invasive ablation[13], due to its deep penetration depth and millimetre spatial resolution. However, the thermal damage caused by HIFU is irreversible. In neuromodulation, on the other hand, low-intensity ultrasound (LIFU) is safe and reversible, thus being much more preferred. It is reported that LIFU in $0.5 - 100$W/cm^2 could still produce bioeffects while avoiding tissue heating[8]. More importantly, a 90 µm spatial resolution is now achieved using 43 MHz LIFU, which was utilized to stimulate neurons in the salamander retina[14], which indicates more potential for ultrasound VNS.

1.2.2.1 Possible mechanism

Although the complete mechanism of ultrasound neuromodulation is still uncovered, hypotheses (Fig 1.3) are established based on experiments and observations[15]. One proposed model is that the ultrasound acoustic pressure causes compression and rarefaction of the cell membrane. These mechanical effects on membranes and ion channels will change the membrane conductance and channel activity. Another hypothesis is that the mechanical effects caused by ultrasound acoustic pressure contribute to the formation of a bilayer sonophore. This gives rise to the mechanically originated displacement currents which alter changes in membrane capacitance and voltage. Either the mechanical-only model or the mechano-electric model results in affecting the voltage-mediated activity of ion channels and changing the neuronal membrane conductance. Further understanding of the ultrasound neuromodulation requires devices with high spatial selectivity so that a specific region or fiber of the nerve could be stimulated without activating unwanted objectives.
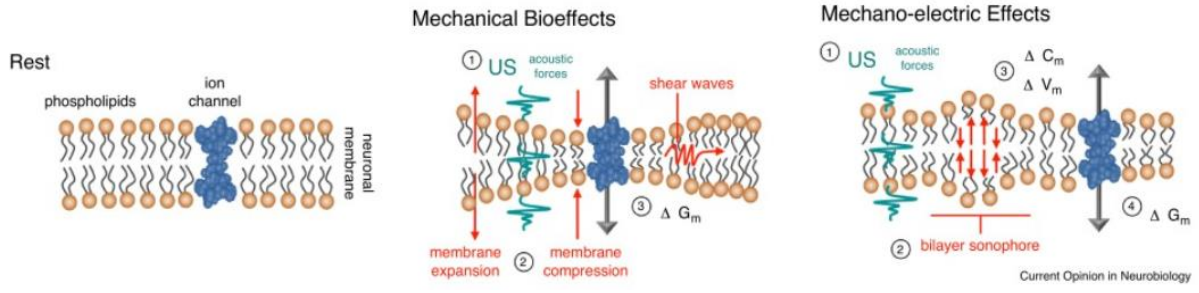
Figure 1.3 non-thermal modulation of neural activity by LIFU[15]

1.2.2.2 Structure and characteristics of an ultrasound system
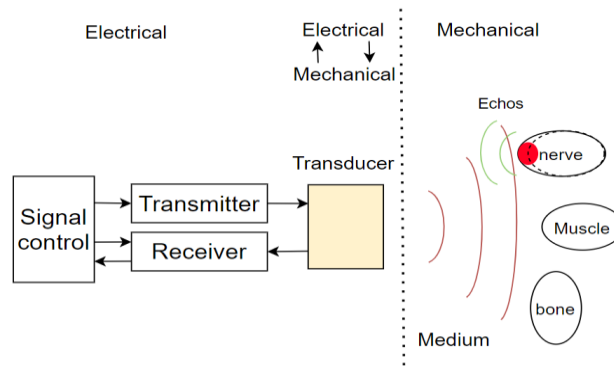
Figure 1.4 structure of an ultrasound system

As in Fig 1.4, a simple ultrasound system contains mainly four parts, the signal control block, the circuit section (transmitter and receiver), the piezoelectric transducer, and the medium area. The signal control is a computing unit where the input signal properties are pre-set and the received signals are processed. In neuromodulation, the bioeffects or treatment efficacy is mainly determined by the acoustic intensity and exposure time, both of the values pre-set in the signal control block. The acoustic intensity could be evaluated by the pressure amplitude, which is linearly proportional to the voltage amplitude of the electrical signal. As in equation 1.1, p is the output pressure, ε and ε_0 are the relative dielectric permittivity and the permittivity in the vacuum, k is the coupling factor, d_{33} is piezoelectric constant and t is the thickness of the piezoelectric layer. All of these parameters are the properties of the transducer material, except V is the applied voltage. In chapter 2, pressure amplitude is used to evaluate the ultrasound output performance, while in the circuit design part, the applied voltage is focused on. Users could determine the output pressure through the signal control block, but the applied voltage is dependent on the transmitter circuit.

$$p = \frac{\varepsilon \cdot \varepsilon_0 \cdot V}{k \cdot d_{33} \cdot t} \tag{1.1}$$

Exposure time is fully controlled in the signal control block. Each input signal could be scaled down to three layers, the total signal domain, burst domain, and pulse domain, as seen in Fig 1.5 [16]. In a total time (TT) for a whole signal, each unit is one burst interval (BI) which contains one burst duration (BD) and one inter-stimulus interval (ISI). The ISI is set to avoid a constantly-on stimulation causing accumulated temperature increase and possible damage to

the target nerve. During the ISI, ultrasound imaging could be conducted since it does not require a long burst duration compared to neuromodulation. Inside each BD, the pulses pattern is determined by different applications. For ultrasound neuromodulation, long bursts of continuous waves are applied to ensure the nerve is stimulated. While for imaging, a single-pulsed waveform is applied where the pulse length (PL) is fixed and it will repeat several times at a pulse repetition frequency (PRF). Depending on different stimulation purposes, the exposure time could be altered easily in the signal control block. In this work, both a continuous waveform (CW) and a single-pulsed waveform (PW) are provided in the pulse domain layer.
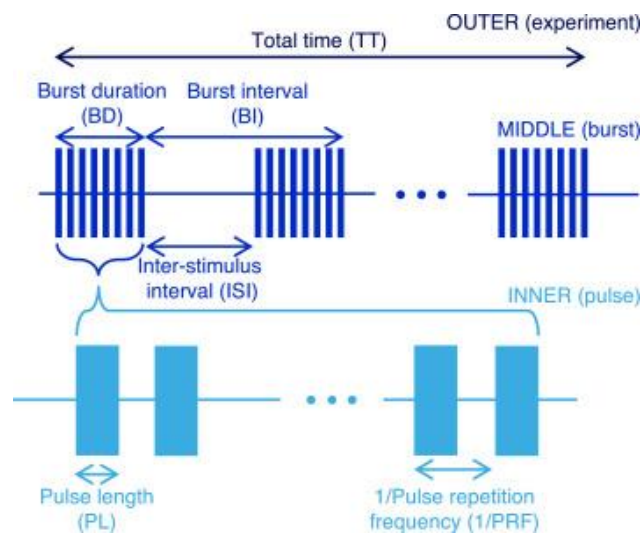
Figure 1.5 signal timing properties related to exposure time[16]

The circuit section is directly connected or attached to the transducer. In the traditional application of ultrasound stimulation, a single crystal bulky transducer is used. As in Fig 1.6, the focusing area is relatively large and the pressure intensity is equally distributed in the focusing area. Although acoustic lenses could be added to improve the focusing capability and spatial resolution, this kind of transducer requires an external mechanical stage to change the focusing region. Manually moving the transducer to change the focal area is possible but the precision is worse, which is only acceptable in some imaging applications. A phased array (either 1D or 2D) could use a technique called beamforming to improve the focusing ability. Since there are several small transducer elements inside one phased array, each of them could generate an ultrasound wave to the targeting position. Contributed to the beamforming, all the different waves from each element could reach the same pre-determined position at the same time where these waves constructively interfere to achieve a high pressure intensity and spatial resolution. The beamforming is conducted in the transmitter as a key design consideration in this work and will be explained in detail in the next section. In a phased array, the spatial resolution is dominated by the acoustic frequency, array size, and the focal depth. Both 1D and 2D array could be achieved in a small and wearable device since the circuit section is attached to each transducer element. 1D arrays are commonly found in commercial ultrasound imaging probes, however, compared to the 1D array, a 2D array could tune the focal spot at any geometric position in a 3D-volumetric space, while the focusing area for a 1D array is only a 2D surface plane. The simulated target is the vagus nerve which might move slightly during the stimulation. In addition, the relative position of the vagus nerve varies from individual to individual. Hence, a 2D transducer array and beamforming transmitter are the desired

structure for our design.
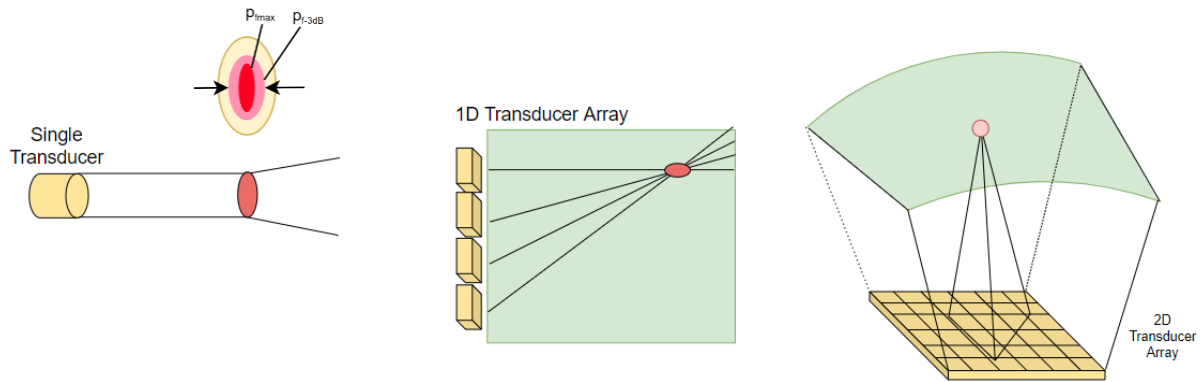
Figure 1.6 Focusing and tuning area for different transducer structure: single, 1D array, 2D array (each element in the 2D array is separated like in 1D array)

The last section of an ultrasound system is the medium area, where all the mechanical or thermal effects happen. Except for the mechanical reaction and related bioeffects caused by the acoustic pressure, reflection and absorption also are the dominant effects during the propagation of the ultrasound wave[17]. The acoustic wave propagates at different speeds in different mediums, such as the bone, nerve, or muscle, because the rigidity and density of them are different. Part of the acoustic wave reflects as echoes when it reaches a different material. These echoes carry information of their reflection boundaries, so they could be gathered and processed to form a diagnostic image, which allows the ultrasound imaging to be conducted at the same time so that a precise stimulation area is ensured. The absorption of the acoustic wave is due to the thermal energy conversion, which results in attenuation of the pressure amplitude. The attenuation is frequency and penetration depth related, which adds more limitations on the design of a transmitter.

1.3 Transmitter and beamforming

As introduced above, an ultrasound VNS system is a combination work of all sections, requiring programming in signal control, circuit design for transmitter and receiver, manufacture of the transducer, and physical analysis of the target medium. The transmitter circuit is responsible for the beamforming technique for a 2D phased array, driving each transducer elements, and providing different modes of waveforms for the signal control block to conduct different function, which is the key design task of this thesis work.

In a 2D phased array, each transducer element will have a channel of CMOS circuit attached to it, containing either transmitting or receiving circuit, or both of them. In Fig 1.7, a transmitter channel receives the signal control orders, generates the required electrical signal then passes it to the transducer interface, where the signal will be level-shifted to a higher level so that it can drive the piezoelectric material. The beamforming functions normally are not fully achieved inside the channel considering the area consumption. Here, the dimension of each channel (or the pitch size) should be half of the ultrasound wavelength to avoid grating lobes which can dramatically decrease the stimulation and focusing performance. Due to these area constraints, it would be quite challenging to interface both the transmitting circuit and receiving

circuit inside one channel. Also, some components could be shared for all channels, so in Fig 1.7 the beamforming function is split between the channel (inside) and the periphery of the array (outside). The detailed arrangement of the circuit components is based on the required beamforming function.
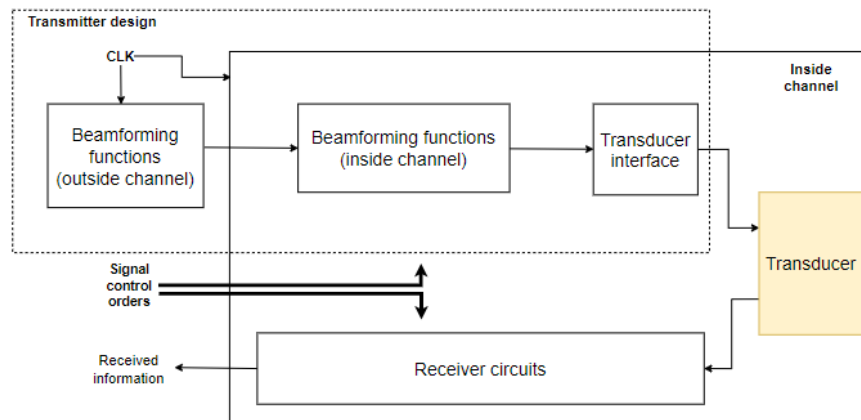
Figure 1.7 circuits implementation in the ultrasound VNS system
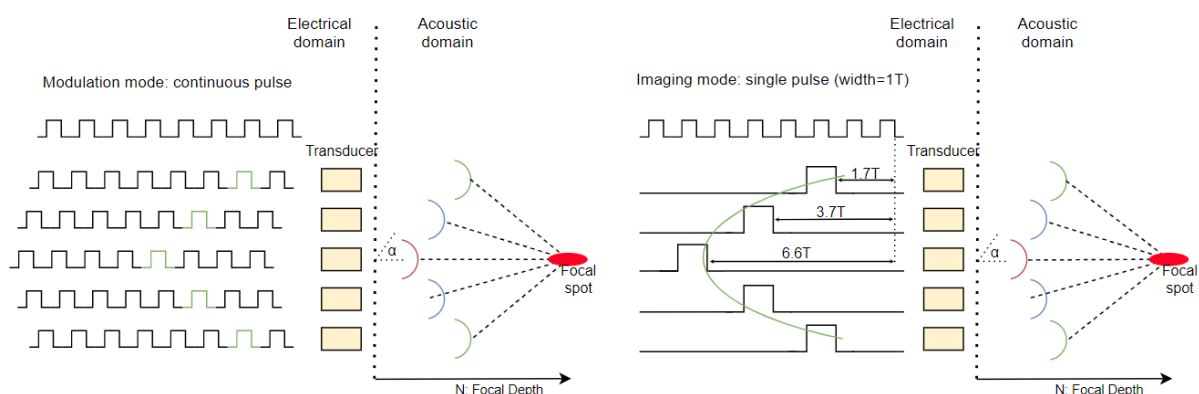
Figure 1.8 beamforming of the transmitter in a 2D array

Beamforming is a technique for directional wave transmission or reception. In Fig 1.8, a row/column of transducer elements is shown to focus at a pre-determined spot. Each element has a different position where the ultrasound wave is sent. Considering an ideal and isotropic medium, the travelling time of each ultrasound wave is dependent on the travelling distance and the sound speed, which is given in equation 1.2. Ultrasound waves from different elements would reach the pre-determined spot at the same time if these wave signals are triggered with the appropriate time delays.

$$t_{travel} = \frac{N}{c \cdot \cos \alpha} \tag{1.2}$$

The difference of waveforms in neuromodulation and imaging also represents in the time delay settings. For a continuous wave (CW), the delay difference among different channels is always smaller than one time period (0 to 1T) since the signal is periodic. The requirement is a precise delay resolution, especially for two adjacent elements. However, the single-pulsed wave (PW) for imaging has a delay difference in a long period (>nT, n=0,1,2,3…) but with the same time resolution as for the CW mode. Because for ultrasound imaging, the focal spot should be the same while the intensity should be kept low to avoid thermal damage.

The transducer interface determines the driving voltage which directly affects the pressure amplitude as in equation 1.1. The spatial resolution of the beamforming transmitter is determined by the focal depth (N), wavelength (λ), and array size (A), as in equation 1.3. All the exact values of the affecting variables will be gathered by the simulation results in chapter 2.

$$S_{res} = \frac{N\lambda}{A} \tag{1.3}$$

1.4 Prior art and Thesis objective

Efforts have been done in the design of CMOS transmitter using in a 2D phased array during the past few decades. Aiming in different applications, the achieving methods and performance of the transmitter also vary.

A miniature 2D phased array ultrasound transmitter[18] is capable to achieve a maximum focal pressure over 100 kPa with a 5 V supply at 0.5 cm depth in tissue without any acoustic matching layer. The spatial resolution could reach 200 μm level. This work is using a 10 MHz ultrasound wave and 26x26 transducer array, aiming for neuromodulation, and the total size of the CMOS chip is 5x4 mm^2. However, the stimulation output waveform is only a continuous wave, which is not compatible with imaging applications. An additional mode of wave could be included in this transmitter to have real-time surveillance of the focal spot. Besides, the pressure amplitude and spatial resolution could be further improved. For the human primary visual cortex, the excitation pressure level is in the MPa range[19], which might differ for the vagus nerve, but a higher upper limit for the pressure is always preferred. Controlling signals could tune the actual output pressure by turning off some of the array channel circuits.

Another transmitter system for ultrasound imaging[13] contains up to 4 types of pulses for different imaging functions. A 32x32 2D array is implanted in this work, which achieves an imaging depth of up to 7.5 cm and 60V driving voltage. Although the penetration is much deeper, the emphasized performance of this work is not on spatial resolution and only 64 of the transducer elements are connected with transmitting circuits. The transmitter of this work is mainly digital, containing some large area consuming digital blocks, such as a 10-bit comparator, which would also be an obstacle to implement both transmitter and receiving circuits inside one channel.

In our proposed system, the ultrasound VNS device is desired to be attached to the neck and conduct both neuromodulation and ultrasound imaging. Aiming for a wearable device, the transmitter design would have some severe requirements on the chip area and power. The heat dissipation must be carefully controlled to avoid any thermal damage. The limited area is a huge obstacle for integrating both transmitter and receiver inside one channel, which calls for some arrangement techniques such as the sparsity control of array channels. The precise geometrical focusing ability and pressure amplitude still need to be ensured under a small-sized chip. Both the physical understanding of the vagus nerve stimulation and the circuit design for a valid transmitter to achieve the VNS functions are required for this thesis work. A list of general specifications are given in the table below, further design variables will be

derived based on the MATLAB simulation in the next chapter. Although the final aim of this system is the VNS for human, currently our work is only targeting the animal models (mice and rats). The specifications below are also based on the research data of mice.

Performance	Specifications
Focal depth	5-7 mm
Pressure at focal spot	>1 MPa
Spatial resolution	~100 um
ultrasound frequency	>10 MHz
Power of each channel	<1 mW
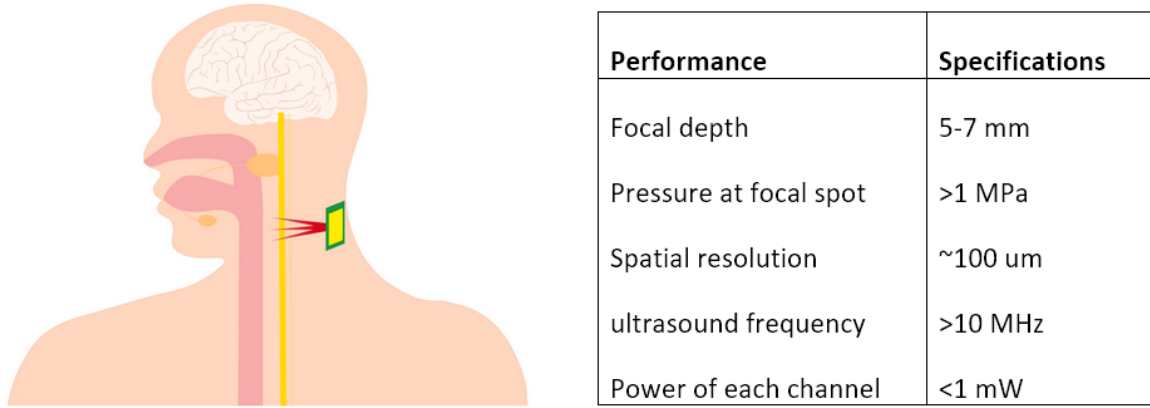
Figure 1.9 proposed device and the general specifications

1.5 Thesis organization
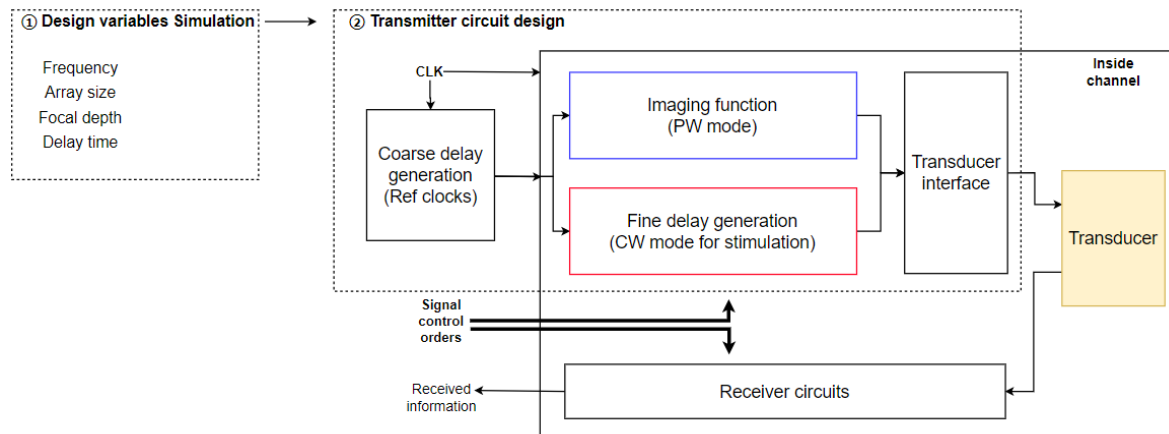
Figure 1.10 Transmitter design blocks

This thesis work presents the design of an ultrasound transmitter system for vagus nerve stimulation, which includes the MATLAB system-level simulation and the design of four circuit block, the coarse delay generation clocks, the stimulation CW wave generation block, imaging function block, and the transducer interface block. Containing the introduction chapter, there are seven chapters in total. The other chapters are:

Chapter 2 demonstrates the MATLAB simulation for the whole stimulating system. In this chapter, the neuromodulation and imaging function will be simulated in MATLAB to get the detailed values for the required design variables, including frequency, array size, geometrical focal depth, and the time delay configuration.

9

Chapter 3 illustrates the design of the coarse phase control block, which is a delay-lock loop including a phase detector, a charge pump, and a voltage-controlled delay line.

Chapter 4 discusses the fine delay control generation. Here the phase interpolator is chosen.

Chapter 5 introduces the embedded ultrasound imaging function block which is fully digital. Several digital components are included, such as counter, comparator, multiplexer, and register.

Chapter 6 shows the interface circuit for the transducer. It is responsible for driving the delayed signal to a higher level to activate the transducer.

Chapter 7 demonstrates the whole system and concludes the contributions and future perspective of this thesis work.

Chapter 2 Stimulation model and Matlab simulation

In this chapter, a MATLAB simulation model that could achieve the transmitter functions are built and simulated. Based on the research data of the vagus nerve, specific design requirements for the circuit design and signal properties setting are derived, including the values of ultrasound frequency, array size, focal depth, and delay time. Additionally, the study of array sparsity is also tested to provide a potential solution for output pressure control and area saving.

2.1 Stimulation model
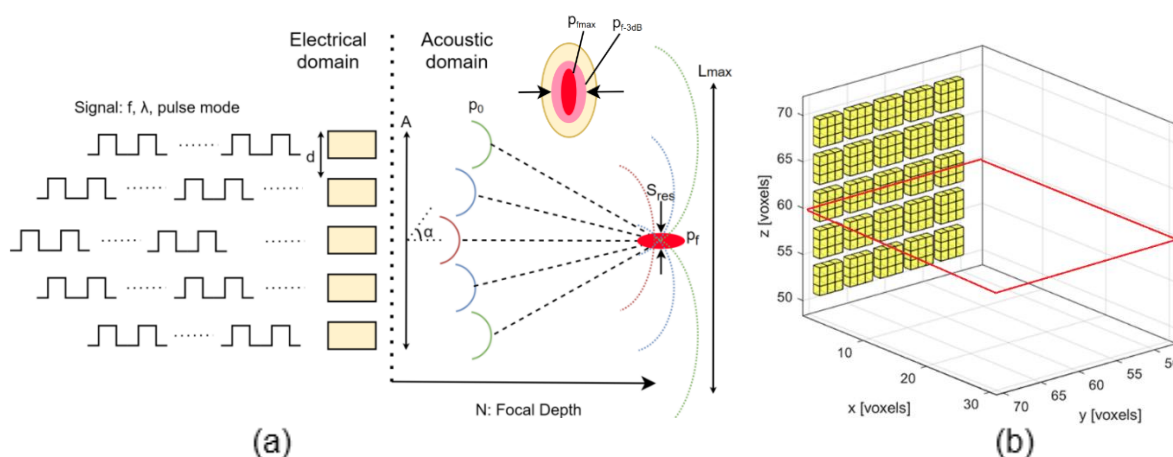
Figure 2.1 MATLAB simulation model

As shown in Fig 2.1, the whole process and its 3D space model from beamforming to pressure sensing in the ultrasound neuromodulation are built by MATLAB code using the k-wave toolbox. In Fig.2.1(a), the electrical domain is the signal generating section, where the code sets the transmitter properties (wave modes, array size, frequency, focal depth). In the acoustic domain, which is also shown in Fig.2.1(b), it corresponds to the medium where the acoustic wave propagates. We assume the medium is soft tissue and isotropic, with sound speed $c = 1540 \, m/s$, medium density $\rho = 1063 \, kg/m^3$, and attenuation coefficient $\alpha = 1 \text{dB/cm·MHz}$. The transducer array is located in the y-z plane of medium space, and the acoustic wave, which propagates in all directions, is sensed by a sensor area in the x-z plane (the red square in the middle of the medium space). The whole space is composed of small-block grids with dimensions 30um*30um*30um. This grid size allows the configuration of each transducer element to be a 3x3 block and the pitch size (distance between the centre of two adjacent elements) to be 120um. The configuration for grid and pitch size is the same as previous work in [18] and is reused here in the consideration of simulation time, channel area, and fabrication process. Smaller dimensions such as 15um have been tested for the grid, but the simulation time is increased to several hours. In [18] only neuromodulation function is implemented, but in this work, more circuit components should be included in one channel. Keeping the pitch size to be 120um, which means a channel area of 120um*120um, is an appropriate and conservative consideration. Additionally, the transducer must be diced into several elements to

11

allow the different time-delay settings for each channel, which requires a fabrication process on a single piezoelectric material. The dimension of the edge between two elements could not be too small to fabricate, and 30um is proved to be an achievable value as in [18]. The main limitation of the pre-set pitch size is on the ultrasound frequency. As mentioned above, the pitch size should be smaller than half-wavelength to avoid side lobes. If so, from equation 2.1, then the ultrasound frequency could only be smaller than 6.4 MHz.

$$c = f\lambda \tag{2.1}$$

However, according to simulation, if the pitch size is larger than half-wavelength but smaller than one-wavelength, the existing side lobes do not have a significant effect on the focusing performance, especially in a large array. The limitation of the 120um pitch size now is mitigated and allows a frequency range from 6.4 MHz to 12.8MHz. Figures of simulation results will be shown later in this chapter. After introducing the model, the variables we need to simulate and their requirements are explained below.

2.1.1 Variables under simulation

In the simulation model, the ultrasound frequency (f), size of the phased array (A), geometrical focal depth (N), and the time delay should be simulated to get appropriate values to fulfil our specifications mentioned in chapter 1.

2.1.1.1 Frequency

The ultrasound frequency (f) is the essential variable affecting the performance of both neuromodulation and imaging function. The following equations illustrate the relationship between frequency and each performance factor.

$$S_{spatial} = \frac{N\lambda}{A} \tag{2.2}$$

$$p = p_0 e^{-\alpha x} \tag{2.3}$$

$$\alpha = af^b \tag{2.4}$$

In Eq 2.1 and 2.2, frequency is directly related to the spatial resolution of the focal spot. Spatial resolution is defined as the -3 dB dimension of the focal spot width as in Fig 2.1 (a). A smaller value of the $S_{spatial}$ means a higher spatial resolution. Since the speed of sound is fixed, higher frequency gives a shorter wavelength and thus a finer spatial resolution. Eq 2.3 is the amplitude of focal pressure, where p_0 is the original amplitude, α is the attenuation coefficient and x is the displacement (penetration depth) of the ultrasound wave. Here the equation is for the ultrasound wave sent by every single element, and the total focal pressure is an accumulated value of all the waves. Because the distance from each element to the focal spot is different, the attenuation for the ultrasound wave from different elements is also different and has to be considered separately. Attenuation is highly frequency dependent and mainly caused by absorption. Since the acoustic wave causes compression and rarefaction when it is propagating in soft tissue, higher frequency gives rise to more vibration which is partly converted to random vibrational heat energy. This process is defined as absorption. Additionally, other effects such as scattering and beam divergence also contribute to the attenuation and are frequency-dependent. The attenuation coefficient is proportional to the frequency as in Eq 2.4, where a and b are material-related coefficients. As mentioned above, the medium in our

model is soft tissue thus the attenuation coefficient is $\alpha = 1$dB/cm·MHz. In the propagation of ultrasound waves, the higher the frequency, and the deeper it propagates, the more attenuation on the pressure amplitude. The simulation of frequency should find a high-frequency value but not cause too much attenuation.

2.1.1.2 Array size

In the MATLAB model, the grid system we used only determines the dimension for each transducer element, but the array size is the dimension of the whole array which relates to the number of elements included. In this work, to achieve a 3D volume space stimulation, a 2D phased array is chosen, and its shape is square. The Array size (A) is given in NxN formation. In the future development, the different shape of the array could be used in different applications. The performance factors related to array size are the spatial resolution in Eq 2.2 and the focal pressure in Eq 2.5.

$$p_f \propto \frac{p_0 A}{N} \tag{2.5}$$

Here the p_f is the accumulated pressure amplitude in the focal spot and p_0 is the amplitude for each element. A larger array contains more transducer elements, which means more waves will be sent to the focal spot, and hence the increased pressure. In the focal spot, the highest-intensity area has a more rapid increase than the surrounding region, which results in a finer spatial resolution. Increasing the array size is a boost in both spatial resolution and focal pressure. However, there are several upper limits for increasing the array size. Firstly, our proposed device is a tiny and wearable one, aiming for the VNS for humans but will start with tests on mice. The dimension of the array would be better on the few-millimetre scale. The circuit design also suffers from a larger array. The distance between the focal spot and the centre transducer elements in the array is different from the distance between the focal spot and the corner elements. This difference increases with the rising of the array size. In pulsed-wave mode, the centre elements should send the wave much later than the corner ones to ensure all the waves reach the focal spot at the same time. So the larger array requires a large delay time which is processed and stored in the circuit components. The number of circuit components or the size of them will also be increased. Additionally, a large delay time value occupies more bits in the data transfer and processing, so the time of one pulse transmit in all channels will be longer and the rate of imaging will be affected. In the simulation, we should find the minimum array size which fulfils the spatial resolution and pressure specifications.

2.1.1.3 Focal depth

The actual focal depth of the phased array is different from the geometrical value we set in the MATLAB code. This difference is called the focal shift. To ensure the actual focal spot is in the pre-set location, an appropriate geometrical value is needed to be simulated. The focal shift is inevitable for a focusing system with a low Fresnel number $N_{Fres} < 100$. The Fresnel number reflects the relative contribution of focusing and diffraction effects for an aperture focusing model and is defined [20]:

$$N_{Fres} = \frac{r^2}{\lambda N} \tag{2.6}$$

Here r is the radius of the aperture, and it corresponds to the dimension of the array in our

system. λ is the wavelength and N is the focal depth. The focal shift is explained in the model of an aperture system in Fig 2.2. The focal shift will always exist if the focusing wave is not a perfect full spherical wave as in Fig 2.2 A. The red point in the middle is the source of the blue waves and the red waves are the reversed signal. The red wave will always focus on the middle position because of its circular symmetry. When this symmetry is broken such as the red focusing wave is only part of a spherical wave, the focusing position will change since there is information lost [20]. This partial wave focusing situation is similar to the aperture focusing model and our phased array focusing system. As shown in Fig 2.2B, the convergence and diffraction both exist in the spreading out of a wave in the aperture focusing model. The ratios of them are $\theta_{conv} = r/N$ and $\theta_{diff} = \lambda/r$ separately. If the convergence is dominating, the output waveform is more focusing and the actual focal spot will be close to the geometrical value. Otherwise, the focal spot will shift more back to the aperture as in Fig 2.2C.
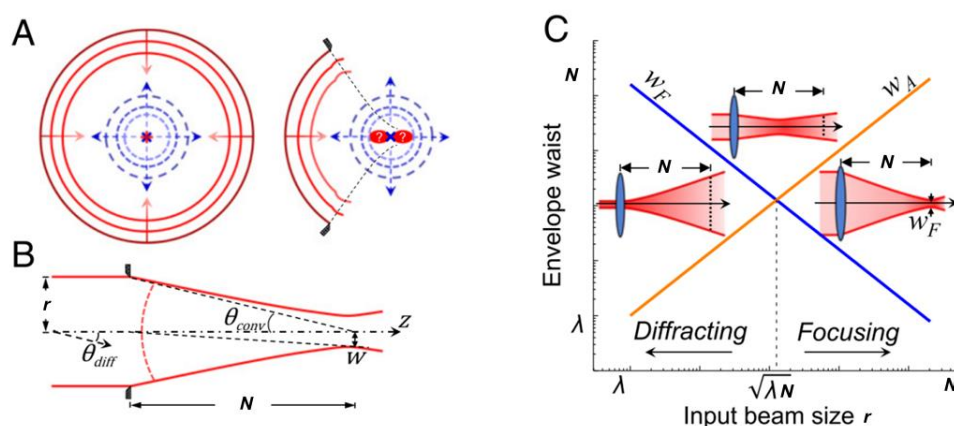
Figure 2.2 Focal shift model [20]

In our case, assuming the minimum geometrical setting is 5mm and the minimum wavelength is 120um (corresponding to 12.8MHz) and the diameter of the array is 2.5mm as in [18], the ration $\theta_{conv}/\theta_{diff} = 10.4$. The output is more focusing but the actual focal depth will always be smaller than the geometrical value. The simulation aims to find the appropriate geometrical value which allows the actual focal depth to meet the specification.

2.1.1.4 Time delay

As mentioned in chapter 1, for neuromodulation, the input signal should be a burst of continuous wave and the phase difference among different channels is $0{\sim}1T$. For imaging, the input signal should be the pulsed wave and the phase difference is larger than one period (nT, n=1,2,3...). After setting a fixed f, A, and N, the delay values for each channel will be calculated in MATLAB and have to be achieved in the circuit design. In MATLAB, these delay values are calculated in a continuous value range, which gives the exact phase each channel needs. However, in the circuits, the generated delay time is based on a unit delay value, and all the delay should be integer multiples of this value. The unit delay value (Δt) is defined as the minimum delay or the time resolution. A computation step called quantization should be conducted to convert the continuous values to quantized values. As shown in Fig 2.3, before quantization, the delay time for each element is a precise value in a continuous curve, while after quantization, the delay time is an approximate number to the precise value and is the integer times of Δt.
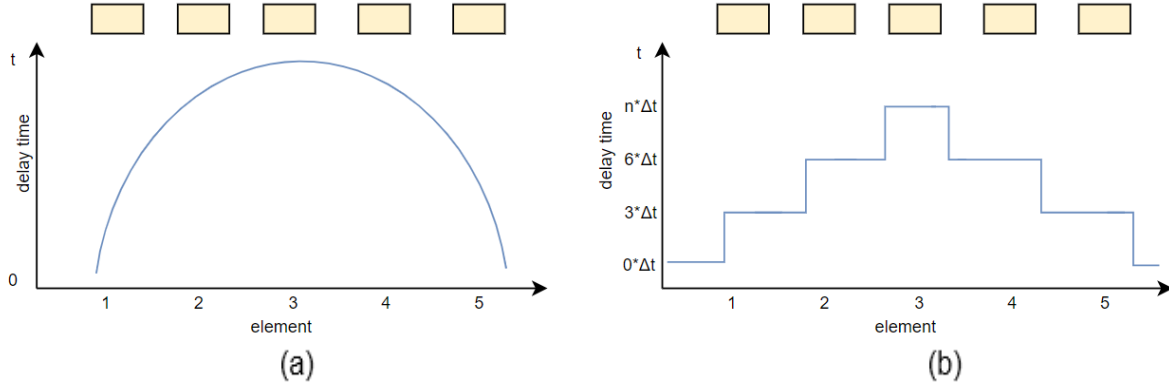
Figure 2.3 Quantization (a)before quantization (b)after quaantization

Δt is given by:

$$\Delta t = \frac{T}{2^n}, n = 1, 2, 3...$$ (2.7)

T is the period of the ultrasound wave, and n should be simulated. The quantization process will degrade the spatial resolution if n is small, but a large n will increase the complexity and area of the circuit. The appropriate value for n is the minimum integer to keep the spatial resolution the same as the value before quantization. After deriving the minimum delay from the quantization bit n, all the delay values could be expressed using Δt as the unit. The delay time for neuromodulation and imaging could be given by:

$$0 \leq t_{delay} < 2^n \cdot \Delta t$$ (2.8)

$$0 \leq t_{delay} \leq a \cdot T + b \cdot \Delta t, \ a, b = 1, 2, 3...$$ (2.9)

Eq 2.9 is for imaging, where $a \cdot T + b \cdot \Delta t$ indicates the largest phase difference among all the channels. Normally this difference is between the centre element and the corner element, but the exact value for a and b depends on the array size and focal depth. Also, to achieve a 3D volumetric focusing ability, a $\pm 15°$ range from the middle of the sensing plane is simulated. The largest phase difference should cover different steering angles.

2.1.2 Simulation setup

Since the spatial resolution and pressure are affected by several variables mentioned above, it would be hard to simulate all the variables at the same time. Each variable will be simulated in a reasonable range while the other variables will be fixed to constant values. The frequency will be simulated first because it determines the wavelength which affects the performance of A and N. Due to the pitch size limitation, the frequency will be swept from 7MHz to 12MHz with a step of 1MHz. The other factors are fixed at $A = 25 \times 25$ (the array is 25*25), $N = 7mm$ (the setting focal depth).

The array size will be simulated next, from 25*25 to 75*75 increasing by 10 elements each time. The other factors are fixed at $f = 12MHz$, $N = 7mm$. The actual focal depth is affected by both the wavelength and the array dimension, so it will be simulated after f and A. The geometric value is tested from 3mm to 8mm with 1mm per step. The other factors are $A = 25 \times 25$, and $f = 12MHz$.

15

After all of these variables are fixed, the quantization bit will be simulated from $n = 1$ until the appropriate value is found. Then the minimum delay and the largest delay difference could be extracted from the time delay matrix in MATLAB.

2.2 Simulation results

2.2.1 Output figures on the sensing plane (f, A, N)

The performance factors are extracted from the output figure of the sensing plane which is placed in the middle of the 3D model. The figure illustrates the pressure amplitude distribution, where the pressure value is normalized to the amplitude of a single element p_0. The highest value of this dimensionless pressure appears at the focal spot and is defined as p_f. Another factor called pressure gain is defined as p_f/p_{array} which illustrate the pressure amplification from the total phased array to the focal spot. With this pressure gain, how much of p_0 is required to achieve the desired focal pressure will be easy to calculate. The spatial resolution is measured as the -3dB dimension of the focal spot width. The distance from the array to the focal spot is the actual focal depth.
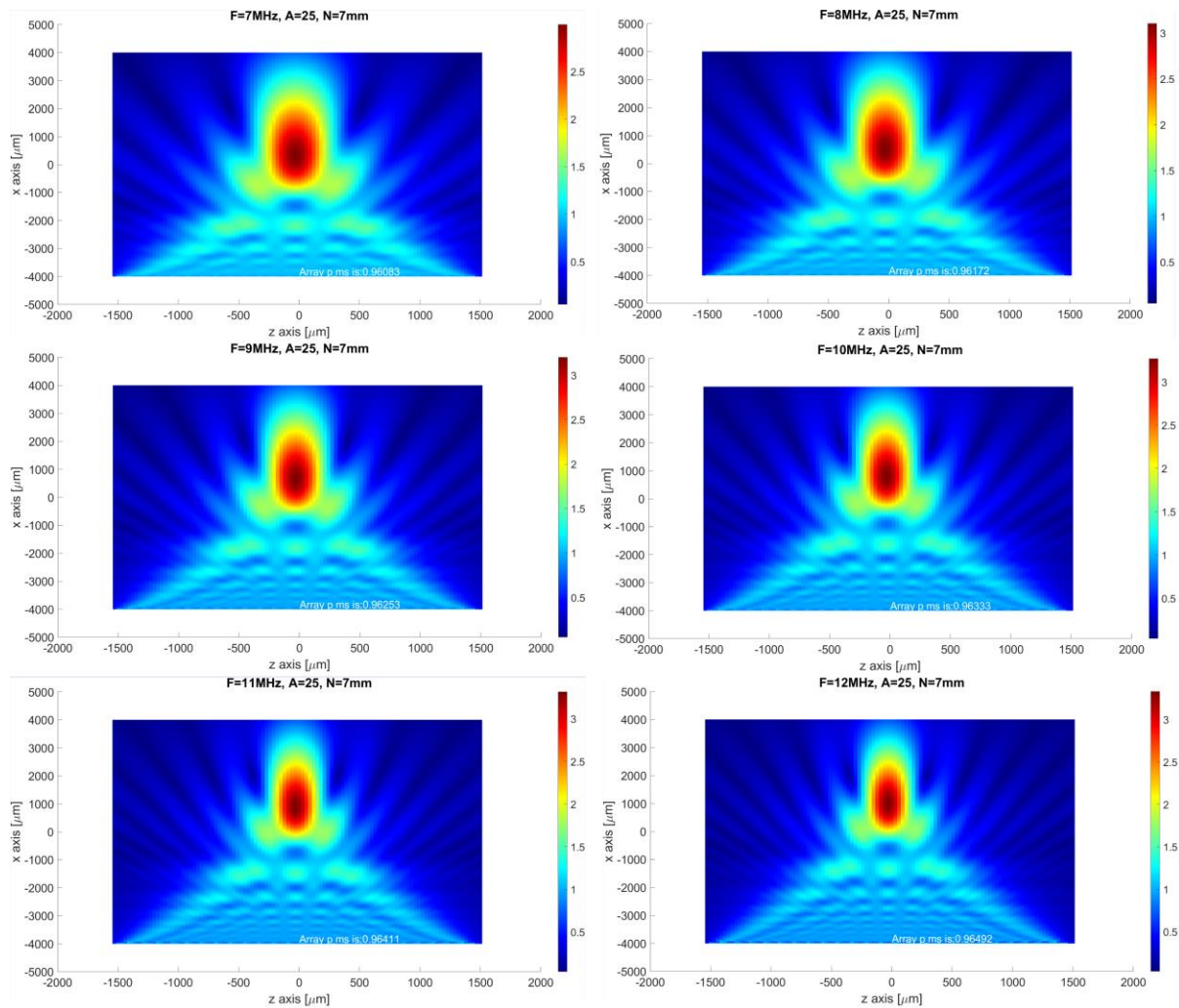
(a) Frequency simulation from 7 to 12MHz
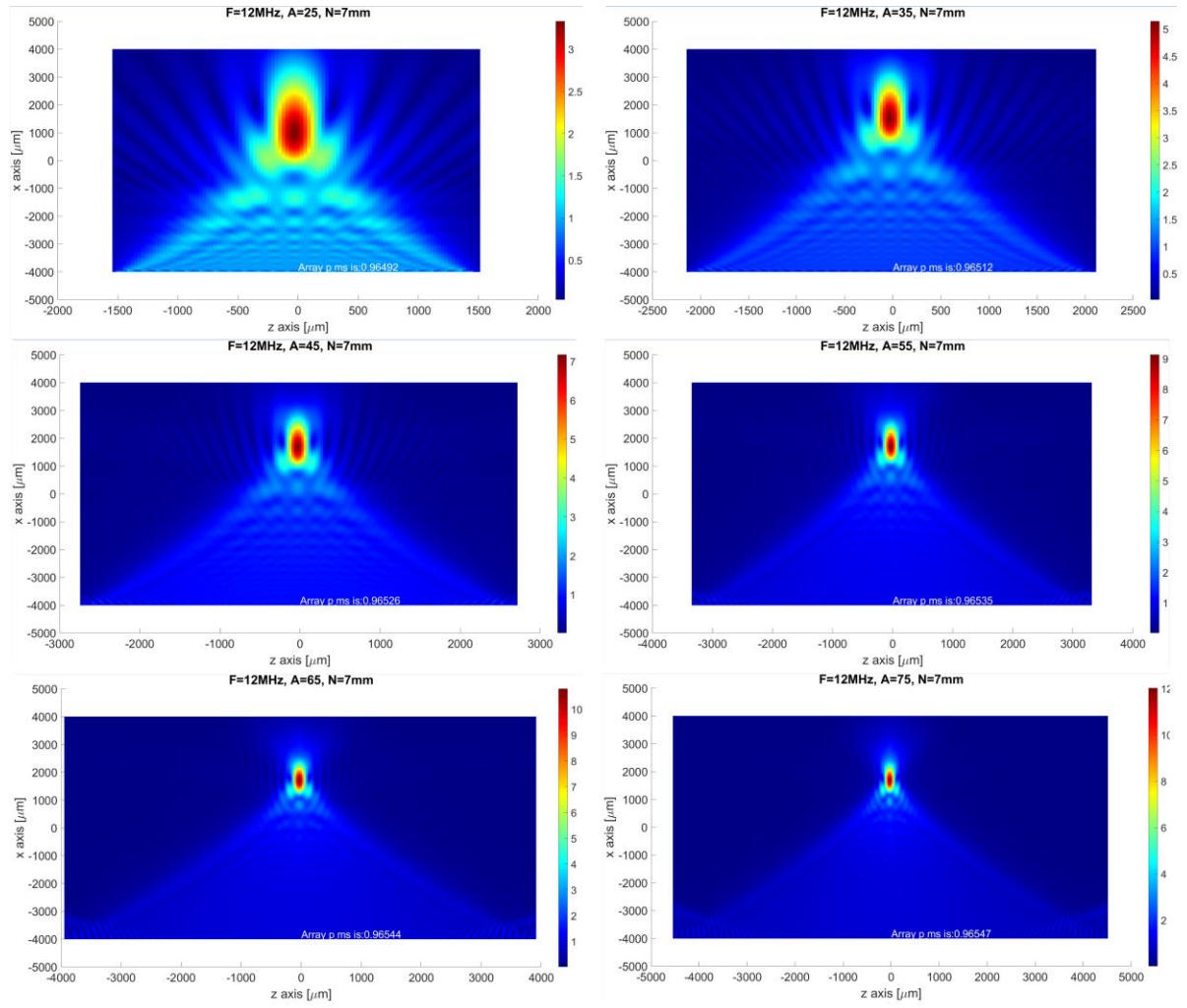
(b) Array size simulation from 25*25 to 75*75
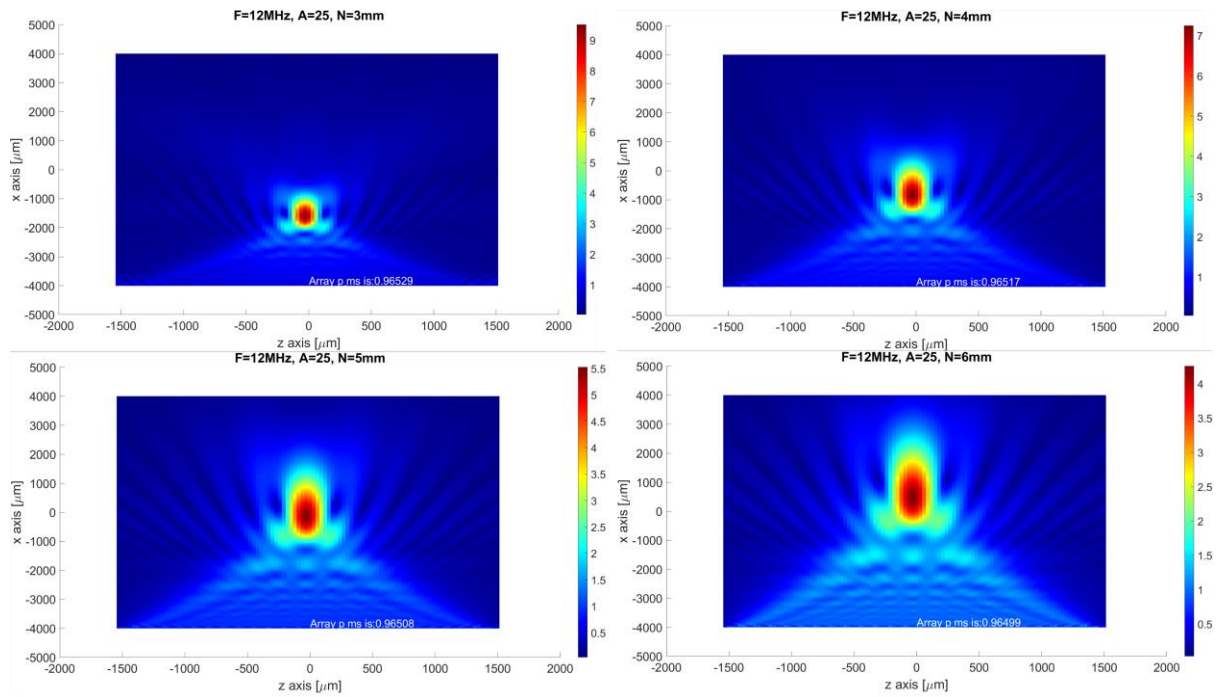
(c) Focal depth(geometrical) simulation from 3mm to 6mm
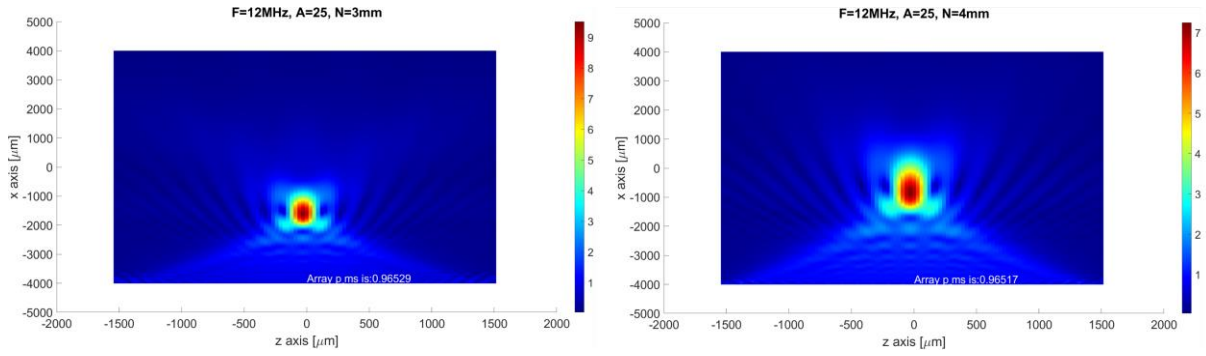
(d) Focal depth(geometrical) simulation from 7mm to 8mm

Figure 2.4 Output figures of the sensing plane

From a general perspective, in the simulated frequency range, the size of the focal spot decreases with higher frequency, while the pressure gain is increased. The side lobe appears to be an issue although the beam size is reduced with higher frequency. However, after the array size simulation, it was found that the focusing performance improved dramatically with a larger array. The side lobe becomes acceptable when the array dimension is larger than 45. It is because the array size has a larger tuning range which allows more elements to focus at the focal spot. The pressure at the focal spot is boosted more than the regions of side lobes, thus the relative ratio of $p_{sidelobes}/p_{focal}$ is reduced. As for the focal depth, the actual value does increase according to a larger geometrical value, but the focusing performance is worse due to a deeper position suffers from more attenuation. A set of more analytical figures will illustrate the detailed variations of the focusing performance using the data extracted from Fig 2.4.

2.2.2 Detailed analysis of focusing performance
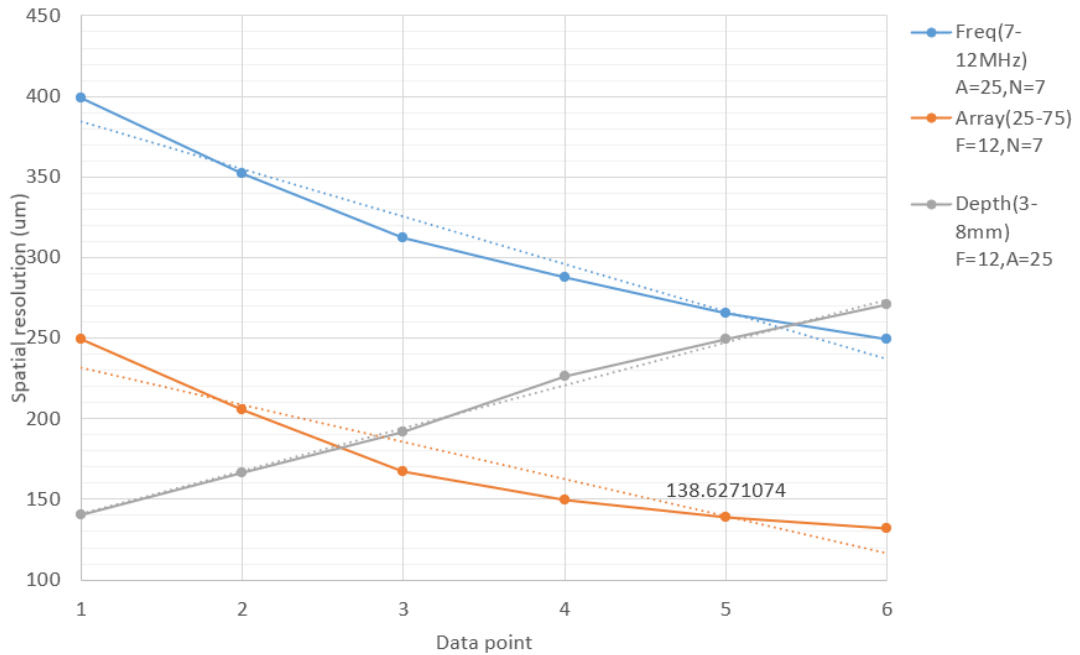
(a) Spatial resolution
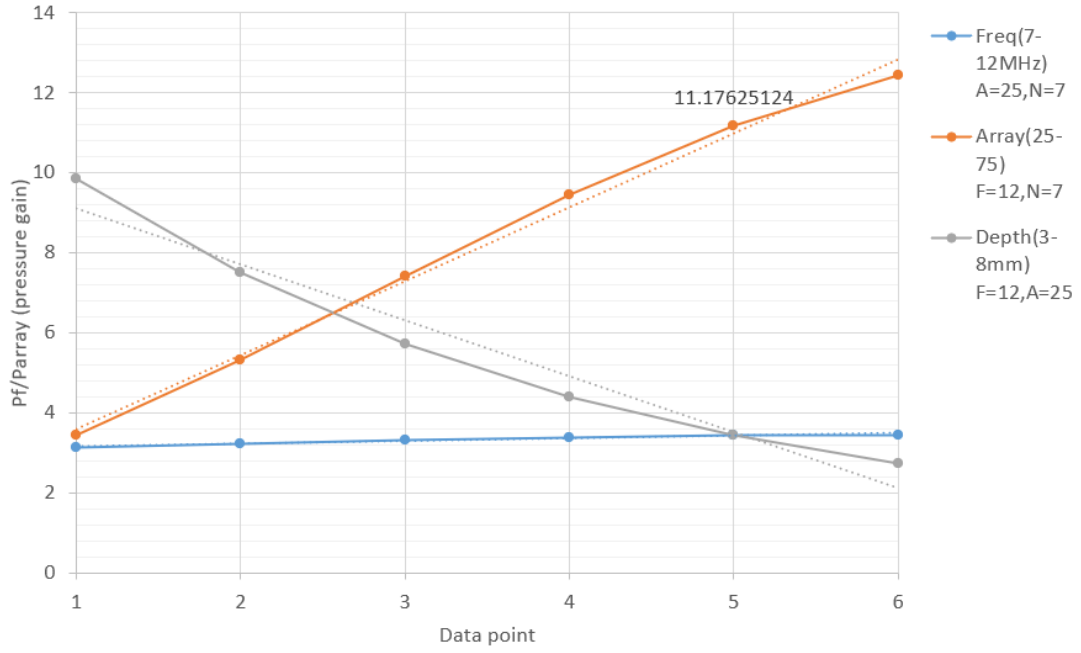
(b) Pressure gain p_f/p_{array}
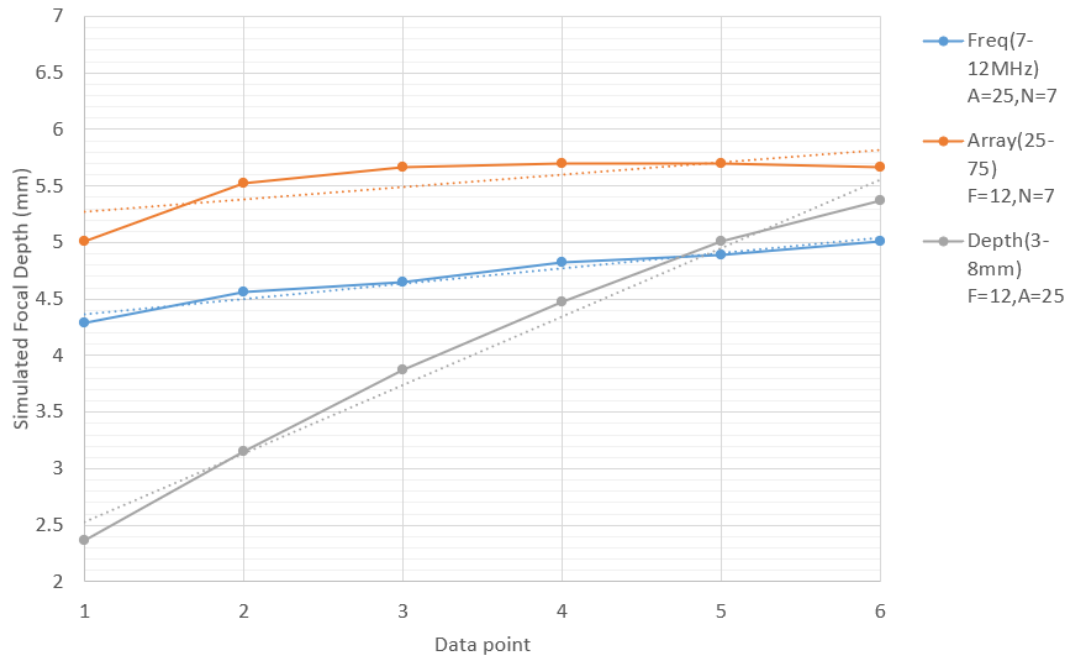
(c) Focal depth (actual)

Figure 2.5 Performance variations

In Fig 2.5, the data for three output performance factors are extracted from the figures of sensing plane. The variation curves demonstrate how the performance factor varies according to different variables. The comparison between the slope of the variation curves and the linear fit lines illustrates the variation speed.

The spatial resolution gets improved with either higher frequency and larger array, but the rate

of improvement is reduced gradually. The reduction is due to the attenuation both for the frequency curve and the array size curve. Higher frequency causes more vibration hence more energy converted to heat and lost. For a larger array, the transducer elements at the corner have the longest path to the focal spot, so these elements suffer from more attenuation because of the long distance. However, the slope of the focal depth curve does not change too much. Except for the attenuation due to a deeper penetration depth, the focusing ability is also changed. As discussed above, the ratio of convergence and diffraction $\theta_{conv}/\theta_{diff}$ is affected by the wavelengths, diameter of the array and the geometrical focal depth. Here only the focal depth is changed which results in a smaller $\theta_{conv}/\theta_{diff}$, so the output pattern is more diffracting with a deeper depth. From the result of spatial resolution, the frequency and array size are expected to be as large as possible while the focal depth should be kept minimum as long as it reaches the depth of the vagus nerve.

The pressure gain shows an obvious trend with the array size or the focal depth while it varies little with the frequency. It has been mentioned above that a larger array contains more elements for focusing and a deeper focal spot will suffer more attenuation on the pressure amplitude. Considering the limited tuning range for the geometrical focal depth, a large array is the best approach to get a higher pressure gain among these three variables. A higher pressure gain will release the requirement on the original pressure p_0 produced by a single element. Since p_0 is determined by the transducer material and the circuit inside the channel, it will be discussed later in chapter 6.

The result of the actual focal depth is mainly based on the ratio $\theta_{conv}/\theta_{diff}$, where $\theta_{conv} = r/N$ and $\theta_{diff} = \lambda/r$. When the frequency is swept from 7MHz to 12MHz, the wavelength is reduced from 220um to 128um and θ_{diff} becomes smaller, hence the actual focal depth moves to a deeper position. The array size is increased from 25*25 to 75*75, corresponding to the diameter from 1.5mm to 4.5mm. However, the actual focal depth only increases when A spans from 25*25 to 65*65 and it drops when A becomes 75*75. Extra tests using 85*85 and 95*95 arrays are conducted. The focal depth has a reduction of 30um and 90um compared to the 65*65 array. One possible reason for this phenomenon would be the precision of the measurement of the focal spot. Because the output figure is composed of several 30um*30um grid points, the actual highest intensity point might be smaller than this scale. When the array size is increased, the total area of the focal spot becomes smaller and the highest intensity point shifts to a smaller and more precise position. When the neuromodulation is conducted, a -3dB dimension area of the focal spot is recognized as the stimulation region, so this shift will not affect the modulation process. Further understanding of this phenomenon would require the experiment using the actual device. Also, according to the simulation result, the key parameter of controlling the actual focal depth is the geometrical value, not the array size.

After all these simulations, the ultrasound frequency is chosen to be 12MHz. The array size is 65*65 which also considering the simulation time. When the array is increased from 65*65 to 75*75, the spatial resolution only improved 6um but requires an extra 40% of the simulation time. The total dimension of the array would be 7.8mm*7.8mm with a 65*65 array which is

relatively large. Further increasing the array size does not have a significant improvement and will increase the burden on circuit design and wire routing. When the f and A are determined, further test on the geometrical focal depth has been conducted. To achieve the expected focal depth from 5-7mm, the geometrical value should be 6.2mm(resulting in 5.01mm depth and 130um resolution) and 8.7mm(resulting in 7.05mm depth and 160um resolution).

2.2.3 Time delay extraction

As mentioned above, the values of delay time created by the MATLAB code should be quantized in the form of integer multiples of the minimum delay Δt. The period of the ultrasound wave is based on the 12MHz frequency and is approximated to 84ns for a convenient calculation in circuit design. After the simulation, the quantization bit is chosen to be 6, so the minimum delay is:

$$\Delta t = \frac{84ns}{2^6} = 1.3125ns \qquad (2.10)$$

The simulation figures with quantization bit = 6 are shown below:

(a)   (b)

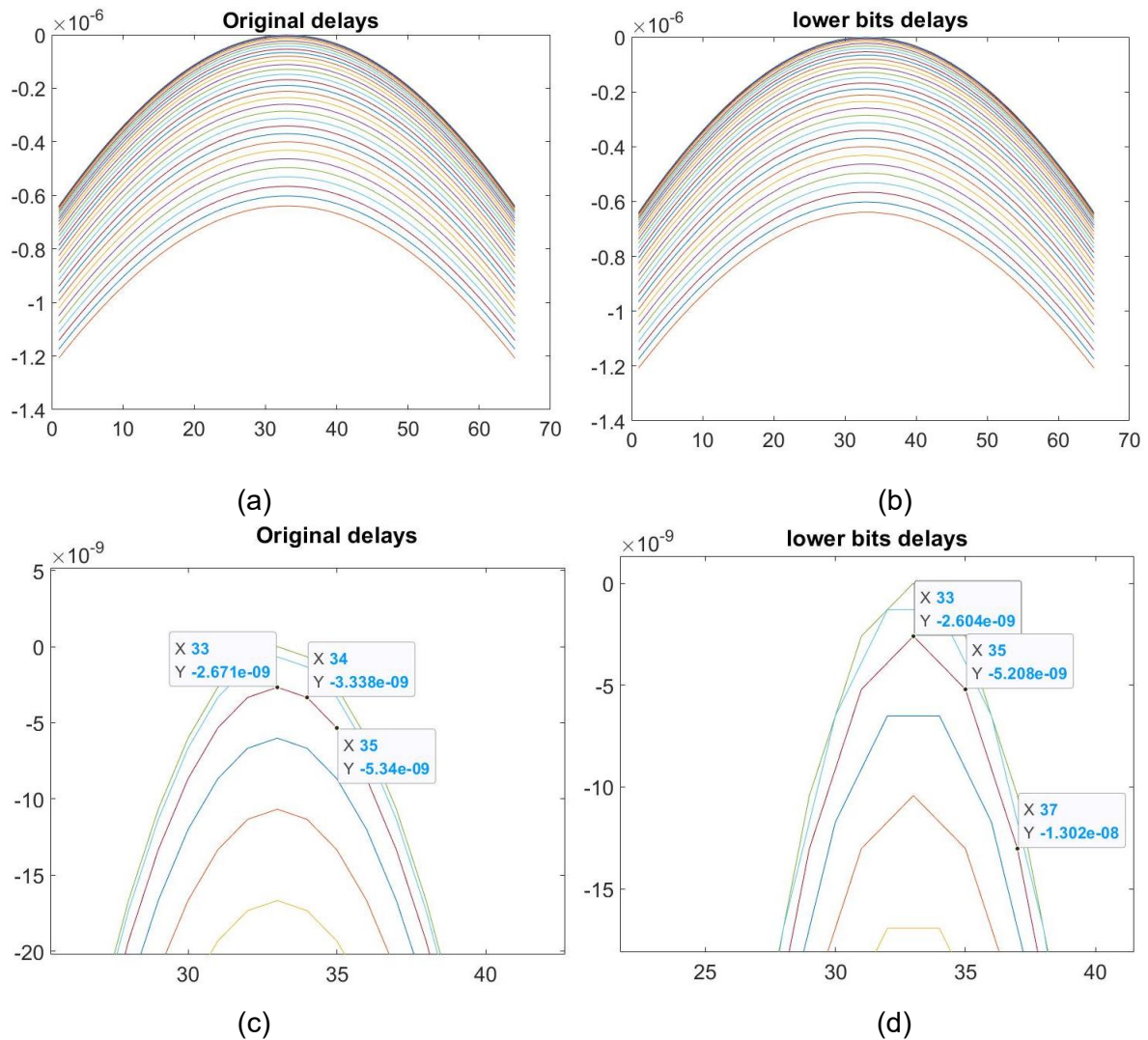(c)   (d)

Figure 2.6 delay time before and after quantization (a)before quantization (b)after quantization (c)zoomed-in before quantization (d) zoomed-in after quantization

From Fig 2.6, the zoomed images show that after quantization the values of delay time are

multiples of the minimum delay. From a general view in (a) and (b), the overall delay curves are similar to those before quantization, so the performance factors will not be degraded by the quantization. The data in Table 2.1 proves that the performance especially the spatial resolution is similar to the one before quantization.

Table 2.1 Performance before and after quantization

Quantization	Frequency	Array size	Depth(geometrical)	Depth(actual)	Spatial resolution	Pressure gain
Before	12MHz	65*65	7mm	5.7mm	138.63um	11.18
After	12MHz	65*65	7mm	5.7mm	138.31um	11.17

After quantization, the range of delay for the neuromodulation is $0 \sim (2^6 - 1)\Delta t$ with a resolution of Δt. For imaging, the delay resolution is also Δt and the largest delay value is $22 \cdot T - 7 \cdot \Delta t$. These are the actual requirements for the circuit design.

All the variables from MATLAB simulation have been determined and listed below:

Table 2.2 Values for the simulated variables

Variables	Symbol	Value
Frequency	f	12MHz
Period(approximated)	T	84ns
Array size	A	65*65
Focal depth(geometrical)	N	6.2~8.7mm
Delay resolution	Δt	1.3125ns
Maximum delay	$22 \cdot T - 7 \cdot \Delta t$	1.83us

2.3 Sparsity

The channel area is a strict requirement for the circuit design, especially for the phased array with both transmitting(Tx) circuit and receiving(Rx) circuit. From some previous work such as [13] and [21], the Tx and Rx circuits are not embedded in the same channel. Also, for ultrasound imaging, the required spatial resolution and focal pressure are not as high as for the neuromodulation. Separating the Tx and Rx in different channels would be a good approach to release the area requirement for circuit design in each channel. The sparsity simulation is set to discover the performance factor variations when some of the transmitting channels are randomly turned off. These turned-off channels could be used to embed only Rx circuits if the spatial resolution and pressure are not reduced too much. The simulation has bee conducted by setting randomly 5% to 95% of the transducer element to be turned off. The other variables

are fixed at $f = 12MHz$, $A = 65 * 65$, $N = 7mm$.
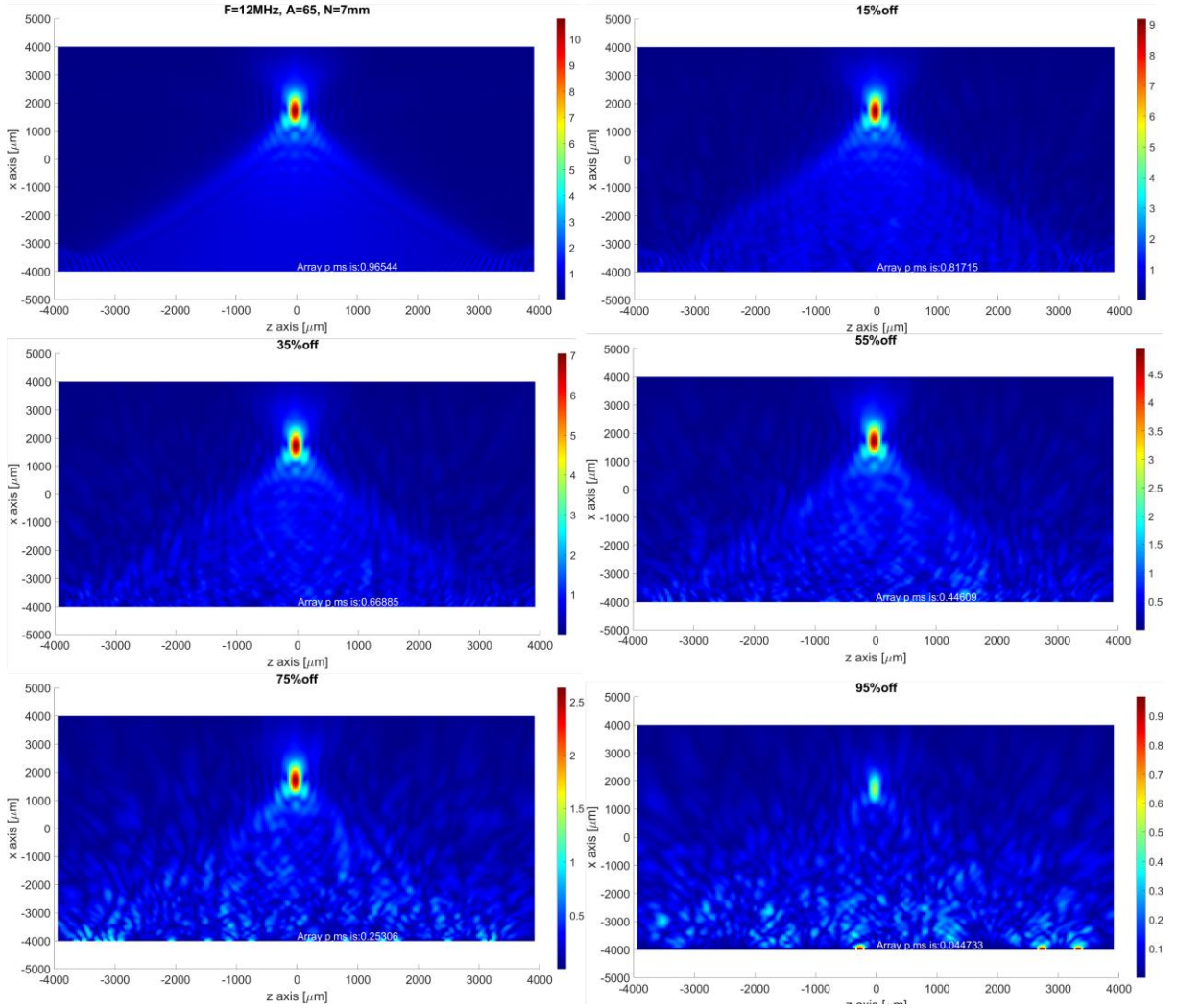
Figure 2.7 Output figures of the sparsity simulation

From Fig 2.7 it could be seen that with the increase of sparsity extent, the focal spot does not change the position, while the intensity is getting lower. The exact values of the performance are compared to the original one(no sparsity) to indicates the deviation in Fig 2.8. The spatial resolution and focal depth barely change in a range of $\pm 2\%$, and the pressure gain changes in a range of $\pm 15\%$. However, the actual focal pressure is linearly controlled by turning off some of the channels, without a change in the driving voltage. From this simulation, the random sparsity only modulates the focal pressure but keeps the focusing ability unchanged. The arrangement of separating Tx and Rx in different channels is allowed as long as the highest pressure for neuromodulation is adequate. It is even possible to leave some channel blank and use the area for adjacent channel interface circuit if that circuit occupies a large area. Another benefit of neuromodulation is the efficacy could be easily controlled by turning some of the channels for different patients. Further development of sparsity could be conducted on different positioning of the Tx and Rx channels or integrating Tx channels with different frequency.
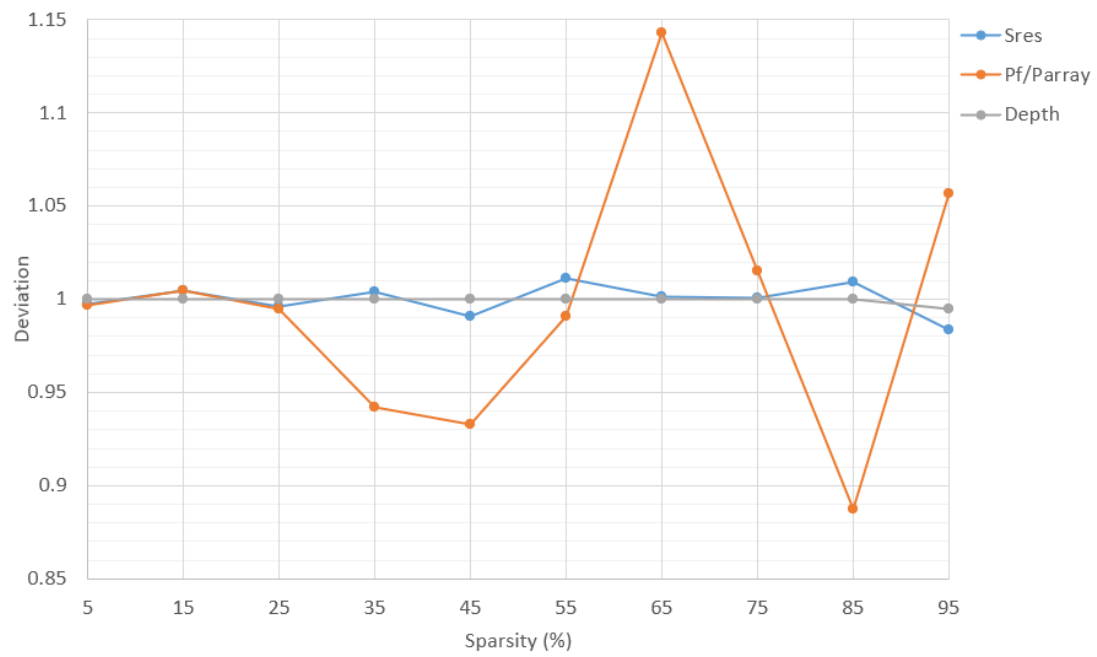
Figure 2.8 Deviation of the performance factors due to sparsity
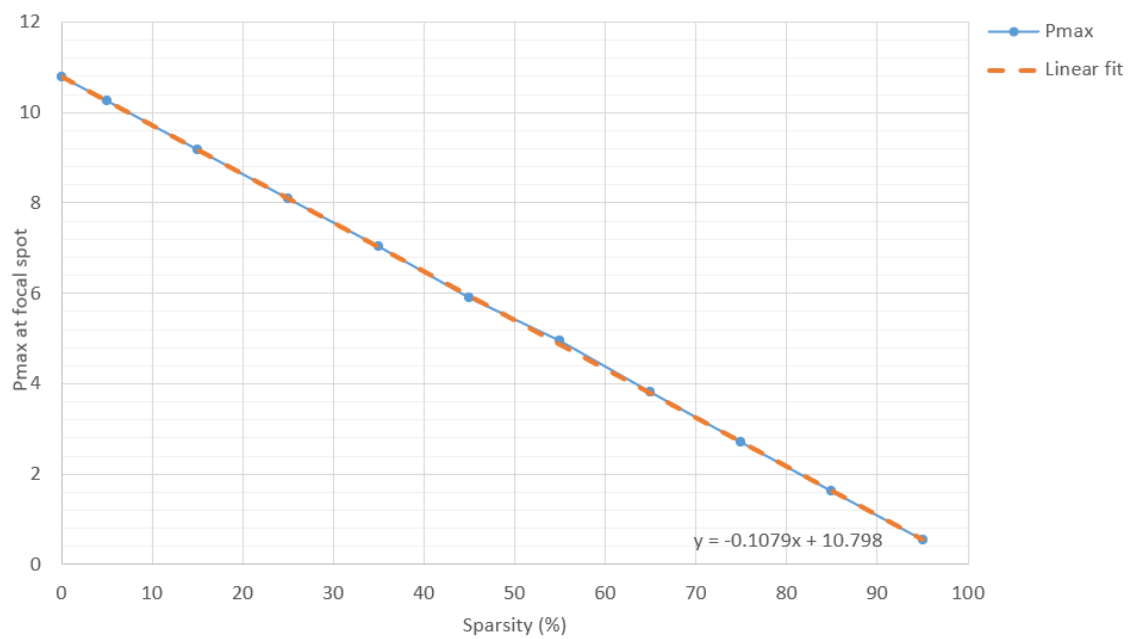
Figure 2.9 Modulation of sparsity on focal pressure

Chapter 3 Coarse delay generation

From this chapter, the circuit design to achieve the expected time delay will start. Chapter 3 and 4 are the circuits to generate the delay for neuromodulation, while chapter 5 is responsible for the circuit to generate the delay for imaging function. In this chapter, firstly the arrangement of coarse delay section and fine delay section will be introduced. Then the fundamentals of the delay locked loop (DLL) will be explained. The design of the DLL will be separated into three sections, each section containing one main component of the DLL. At last, an output simulation for the whole DLL will be shown.

3.1 Delay/phase map arrangement

As discussed above, the required delay range for the neuromodulation is $0\sim(2^6-1)\Delta t$ where the resolution $\Delta t = 1.3125ns$. Because in the neuromodulation, the wave signals are continuous, this delay range is corresponding to a phase map from 0 to 2π. To achieve 2^6 values of the delay time, dividing the 2π phase into a coarse section and a fine section is essential. Because the channel area is limited, it would be impossible to implement all the delay values inside the channel. Generating all the delay values outside the channel faces the obstacle of complicated wire routing. A reasonable arrangement of the phase map is shown in Fig 3.1, where the coarse delay is from $0°$ to $180°$ with a step of $45°$ and the fine delay is in a $45°$ range with a step of $5.625°$. The five clocks from $0°$ to $180°$ will be generated outside the channels as the reference clocks and then are sent to each channel, where two adjacent reference clocks will be selected to generate a finer delay. A polarity control bit could be used to invert the input reference clocks for each channel which completes the phase map from 0 to 2π.

Figure 3.1 Phase map of delay generation

The requirements for the reference clocks in the coarse delay generation are the precision and stability. The minimum delay for the transmitter is $\Delta t = 1.3125ns$, so the maximum delay error should be $\Delta t/2 = 656.25us$ to avoid the delay value changing to the next step. The power

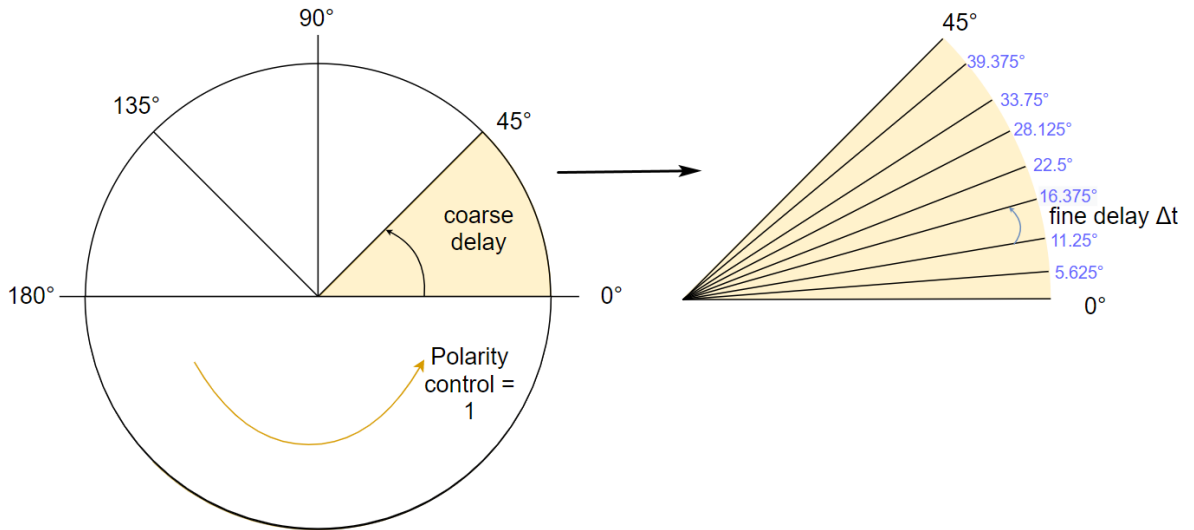and area limitations are relatively low because the coarse generation block is placed outside the channels and shared by all of them. The DLL is chosen to generate these reference clocks in our design, due to its advantages of low phase noise, no jitter accumulation and multiple-phases output generation.

3.2 Fundamentals of DLL

3.2.1 Architecture and operating principle

DLL is a precise delay generation circuit widely used in many timing applications, such as clock generation and signal synchronization[22]. Similar to the phased-locked loop(PLL), DLL is also using negative feedback to align the phase of a delayed output signal to a given input signal. The main difference between the DLL and the PLL is the architecture of how the output signal is delayed. In PLL, a voltage-controlled oscillator(VCO) is implemented while in DLL a voltage-controlled delay line(VCDL) is utilized. Due to the absence of the VCO, the DLL can provide a more excellent jitter performance than the PLL since the random timing error does not accumulate from cycle to cycle. As a first-order loop circuit, the DLL is also more stable and easy to design than the PLL. A block diagram illustrating all the components of a DLL is shown below.
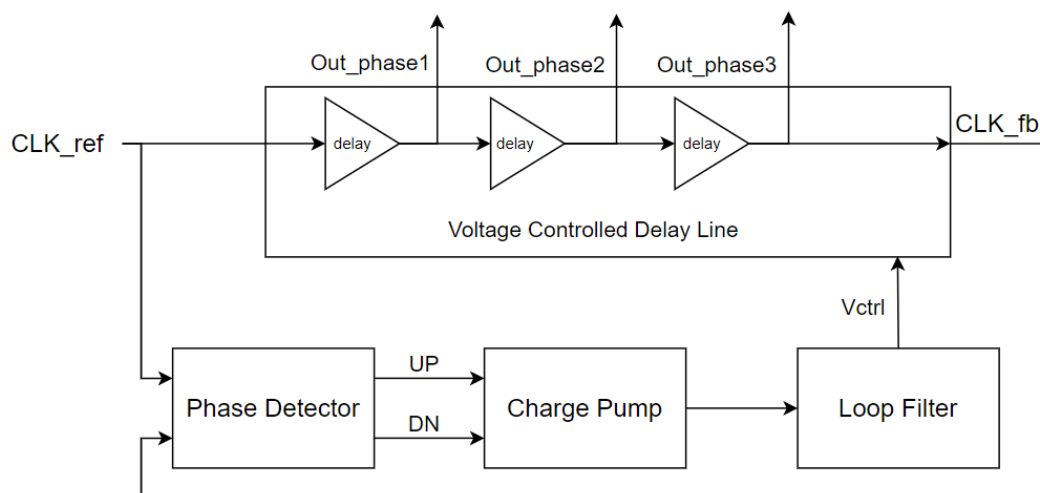
Figure 3.2 The block diagram of DLL

A typical DLL contains mainly three blocks, the phase detector(PD), the charge pump(CP) and the voltage-controlled delay line(VCDL). The loop filter is normally a single capacitor and it is considered together with the charge pump. The operation principle of a DLL is: The original clock (CLK_ref) first propagates through the VCDL to generate a delayed clock (CLK_fb) which will be sent back as one of the inputs of the PD. The PD will compare the phases of two input clocks and convert the phase difference to UP and DN signals which indicate the direction and magnitude of the phase error. This phase error will be converted in either voltage or current form in the CP to charge or discharge the loop filter capacitor. The voltage accumulated on the LF capacitor, also the control voltage for the VCDL, will increase or reduce based on the magnitude of the phase difference. The VCDL is composed of several delay stages as in Fig 3.2, and the delay time of each stage is equal and controlled by the Vctrl. Thus the total delay

on the CLK_fb will also vary according to the Vctrl until the CLK_fb is aligned to the CLK_ref at a certain phase difference. The phase difference should be $n\pi, n = 0,1,2 \ldots$ when the CLK_fb is aligned, and the exact value is dependent on the application. Once the output signal CLK_fb is aligned, each delay stage inside the VCDL could provide a delayed signal and the phases are equally distributed with a step of $n\pi/N$, where N is the number of stages.

From the architectures, there are two types of DLLs. The difference between them is the input signal of the VCDL. One architecture is shown in Fig 3.2 where the input signal of VCDL is the same clock signal of the phase detector. This architecture is often used for frequency synthesis, clock generation and signal synchronization. In the other architecture, the CLK_fb is generated by a signal CLK_ref1 through the VCDL and then compared to a separate CLK_ref2 through the PD. This architecture is often used for clock recovery circuits. In our design, the purpose is to generate reference clocks with multiple phases, so the first type is preferred.

From the circuit implementation, the DLLs could be divided into analog DLLs and digital DLLs. In Fig 3.2 it is a typical analog DLL. The digital DLL is also based on the negative feedback loop but the function blocks are made of digital circuit elements. For example, in a digital DLL, the charge pump is replaced by a control logic circuit such as a counter. The counting bits adds or reduces according to the UP/DN signal and then change the delay time of the delay line. The delay time variation of a digital DLL is also quantized which requires large bits of digital components to have a fine delay resolution. A low-bit digital DLL will result in great jitter at the output clocks. In an analog DLL, the delay control inside VCDL is continuous and sensitive to the control voltage, which allows more precise control of the output signals. Although the power consumption of an analog DLL is higher than a digital one and the design complexity is much enhanced, the analog DLL could provide a better jitter performance which is essential for our reference clocks generation. The problem of power consumption is released because the DLL is located outside the transmitter channel which results in a relatively low average power in each channel. Thus, an analog DLL is preferred in our design.

A loop dynamics model of the analog DLL is shown below for further analysis[23]. In Fig 3.3, the input phase Φ_{in} is the original phase of signal CLK_ref, and the output phase Φ_{out} is the phase of the delayed signal CLK_fb. K_{PD} is the gain of the phase detector with a unit of Volts/rad. I_{CP} is the charging current in the charge pump and T_{REF} is the period of CLK_ref. K_{PD} is the gain of voltage-controlled delay line with a unit of rad/Volts. The loop filter is assumed as a single capacitor with its transfer function as $F(s) = 1/sC$.
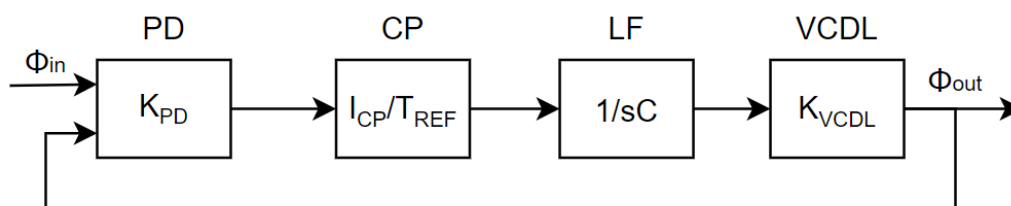

Figure 3.3 Loop dynamics of analog DLL[23]

The output phase could be written as:

$$\Phi_{out} = (\Phi_{out} - \Phi_{in}) \cdot \mathrm{K}_{PD}\mathrm{K}_{VCDL} \cdot \frac{\mathrm{I}_{CP}}{sC} \cdot f_{REF} \qquad (3.1)$$

The transfer function of this loop is:

$$H(s) = \frac{\Phi_{out}}{\Phi_{in}} = \frac{1}{1 + \dfrac{s}{\omega_N}} \quad, \omega_N = \mathrm{K}_{PD}\mathrm{K}_{VCDL} \cdot \frac{\mathrm{I}_{CP}}{2\pi C} \cdot \omega_{REF} \qquad (3.2)$$

It can be seen that the DLL is a single-pole system thus is unconditionally stable. The loop bandwidth ω_N is linear to the frequency of input signal CLK_ref as long as the other parameters are constant. If the input signal frequency is variable, higher ω_{REF} will cause a higher ω_N, resulting in a faster acquisition speed of the system. The acquisition speed, or the lock-in time, is the time DLL takes to correct any disturbance in the loop and reaches a stable state (the output CLK_fb is aligned to CLK_ref). However, choosing the loop bandwidth is a trade-off between output phase noise, size of the filter and the lock-in time. A higher ω_N will bring a shorter lock-in time at the expense of more phase noise at the output. In our design, a shorter lock-in time means the focusing ultrasound wave could be generated faster. Although a short lock-in time is desired, it is not at the top priority. The delay time precision of the output clocks is the most essential, but several non-idealities in the operation will degrade the timing performance of DLL.

3.2.2 Non-idealities of DLL

In any realistic circuit, the rising/falling stage of a clock signal will not perfectly occur at the ideal point. The short-term, random fluctuations in the signal phase which causes the signal deviated from its ideal points, is defined as the phase noise(PN). It is usually normalized to a 1 Hz bandwidth at some offset frequency away from the signal frequency and relative to the amplitude of the signal. In the time domain, jitter is the corresponding equivalent of PN and could intuitively illustrate the timing deviation. In general, jitter contains deterministic jitter and random jitter. Deterministic jitter is caused due to duty-cycle distortion and device mismatch. Random jitter is caused mainly due to random noise sources in the circuit such as MOSFET thermal noise, substrate noise and supply noise. From the measurement methods of a periodic clock signal, jitter could be divided into three types: edge-to-edge jitter (J_{ee}), cycle jitter (J_c), and cycle-to-cycle jitter (J_{cc}).

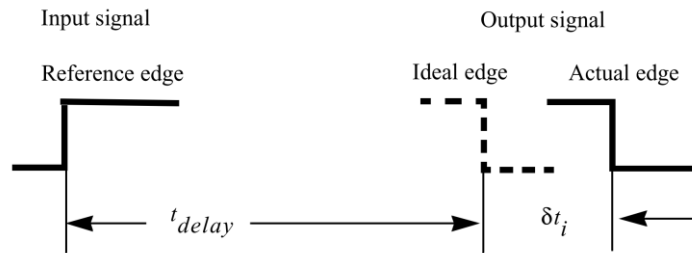*Figure 3.4 Edge-to-edge jitter*

The edge-to-edge jitter is characterized by the deviation of every actual single stage from its

ideal position (δt_i), as in Fig 3.4, and it is also one of the main performance evaluation factors of our design.

From the block diagram Fig 3.2, the noise sources in a DLL are mainly the phase noise of reference signal CLK_ref, limited detecting resolution of PD, current sources mismatch of CP and the delay stage mismatch of VCDL. The PN of the CLK_ref is unavoidable and depends only on the purity of the input signal since it is the original clock of the whole design. The limited detecting resolution of PD is shown in Fig 3.5.
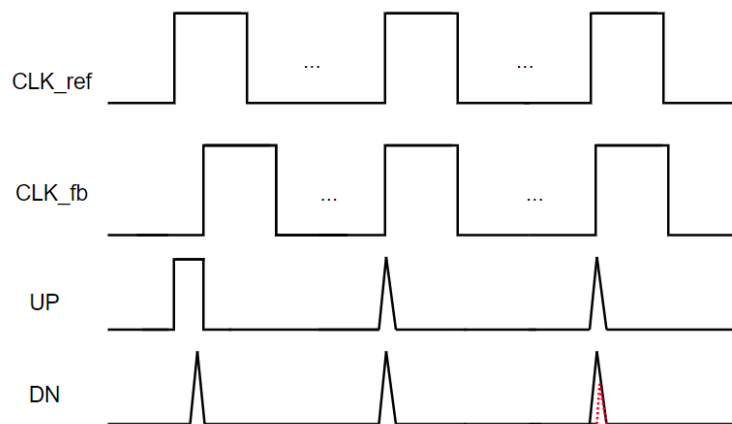

Figure 3.5 Output of PD under different phase difference

When the phase difference of CLK_ref and CLK_fb is huge (now the CLK_ref is ahead of CLK_fb), the UP signal has much more on-time than the DN signal which results in a large and positive value of the average voltage $V_{avg} = UP - DN$. If the two input clocks are aligned, both UP and DN give the same pulses which result in a zero average voltage. However, if the phase difference is extremely small, as the red-dot line indicated, the UP has a slightly longer on-time than the DN. The average voltage should be positive but has a small value. This small output voltage would be insufficient to open the charging switch of the CP, which means the control voltage will not be changed when the phase difference is smaller than a certain value $\Delta\varphi$. The minimum value $\Delta\varphi$ is the limited detecting phase resolution of the PD. Considering the parasitic capacitance, the output voltage of PD does not rise or fall with an extremely sharp edge, the actual on-time of UP or DN signal is even smaller which gives a larger $\Delta\varphi$.

As for the CP, the charging current and discharging current are normally from different current sources. If the two current sources charge or discharge the loop filter capacitor with the same current, the output control voltage will be like the black curve in Fig 3.6. When the UP signal equals the DN signal, the control voltage will keep at a constant value. However, both PMOS current sources and NMOS current sources are implemented in a CMOS circuit, which could not ensure a perfectly matched current source pair. The green dot curve indicates the variation of control voltage when the PMOS current source (I_UP) drives a larger current than the NMOS current source (I_DN), and the blue dot curve shows the opposite situation. With the negative feedback, the control voltage will finally oscillate at a subtle range which results in a rippled curve. The delay time is controlled by V_{ctrl} and will also keep varying.
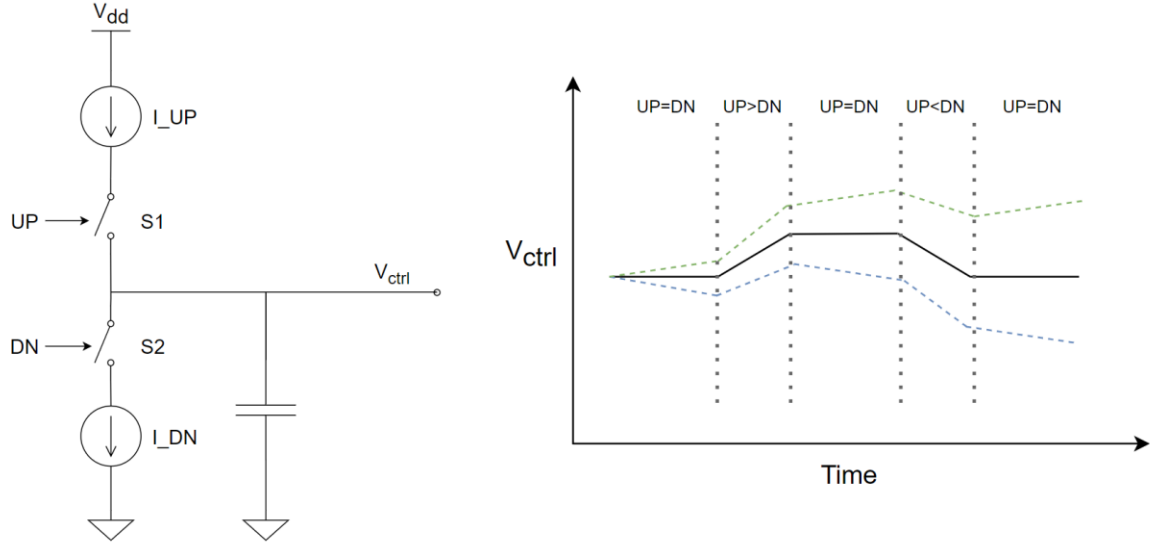
29

Figure 3.6 Current sources mismatch of CP

Except for the variation from the control voltage, inside the VCDL each delay stage suffers from the mismatch among them. The delay time is sensitive to the parasitic capacitance, so the surroundings of each delay stage should also be designed the same as each other. With dummy components, this problem could be mitigated. All these noise sources will result in jitter at the output clocks, and the requirement of our design is the edge-to-edge jitter should be smaller than $\Delta t/2 = 656.25us$. The following sections will be the design sections of the DLL.

3.3 Phase detector

The function of the PD is to generate two output signals, UP and DN, which represent the phase difference between CLK_ref and CLK_fb. Then these two signals will be sent to CP for the control of charging/discharging the LF capacitor. The characteristic curve of PD is shown in Fig 3.7.

For an ideal PD, the average voltage of the UP and DN signals should be linearly proportional to the phase difference through the whole detection range. Linearity is the first key performance factor to ensure the proper transition of the phase difference information from PD to CP. The slope of the characteristic curve is the gain of PD, K_{PD}, which indicates the sensitivity of PD to the phase change. The detection range is how large the phase difference between CLK_ref and CLK_fb could be properly recognized, and is expected to be as large as possible. In this figure, the detection range of the ideal curve is $[-2\pi, 2\pi]$. There are two blind zones near -2π and 2π for the red curve. During the blind zone, the PD still have an output but it does not reflect the actual phase difference of the input signals. The dead zone is related to the detecting resolution. During the dead zone, the phase difference is too small to be recognized. Some possible factors causing the dead zone could be the path delay of the circuit, charging the parasitic capacitance etc. In the dead zone, the PD output keeps zero and it is impossible to transfer information to CP. Besides, as mentioned above, even a small phase difference could be detected, the output voltage of PD should be adequate to enable the CP. In the design, the dead zone should be reduced as small as possible.
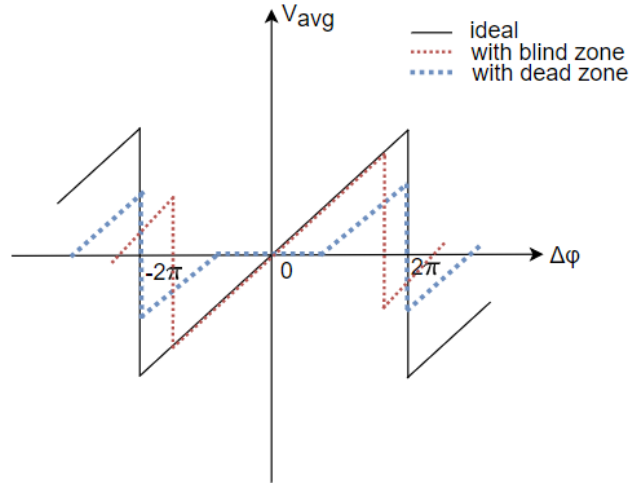
Figure 3.7 Characteristic curve of PD

Another problem in the PD design is the false locking or harmonic locking problem, which means the output CLK_fb is aligned to CLK_ref at a wrong phase difference value. As mentioned above, the CLK_fb could be aligned to CLK_ref with a phase difference $\Delta\varphi = n\pi, n = 0,1,2$ .... In our design, with the largest output phase as 180°, the $\Delta\varphi = 2\pi$ is an easy-achievable value as the phase difference when DLL is locked. The CLK_fb in our design would be an inverted signal of the 180° output clock. As in Fig 3.8, the proper locking situation is the rising edge 4 on CLK_fb aligned to edge 2 on CLK_ref. The lock-in state could be achieved when the edge 4 is aligned to edge 1 or edge 3, but it will result in a phase difference of $\Delta\varphi = 0$ or $\Delta\varphi = 4\pi$. The output clocks from VCDL will all reach the wrong delay time. The false locking must be avoided and it is related to the output delay range of the VCDL and the initial delay state. A start-up control circuit and a limited delay range could help to avoid false locking.
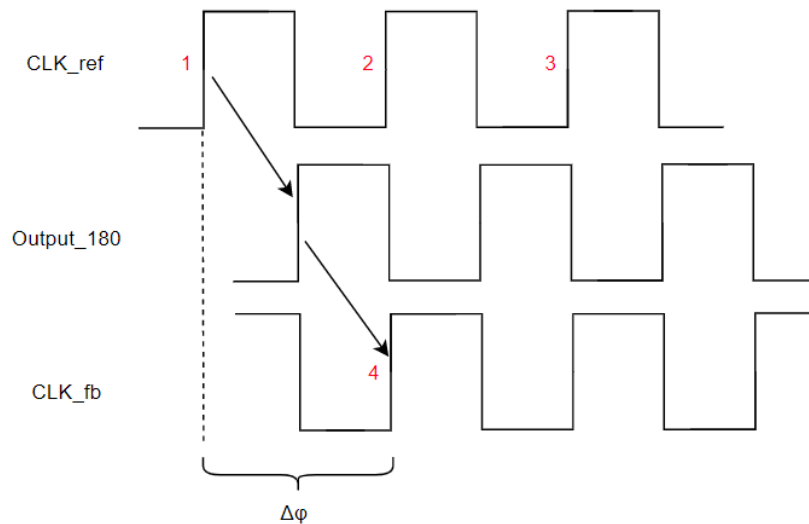

Figure 3.8 Phase locking diagram

In a conclusion, the design of PD should focus on the curve linearity, a large detection range, small dead zone and avoid the false locking problem.
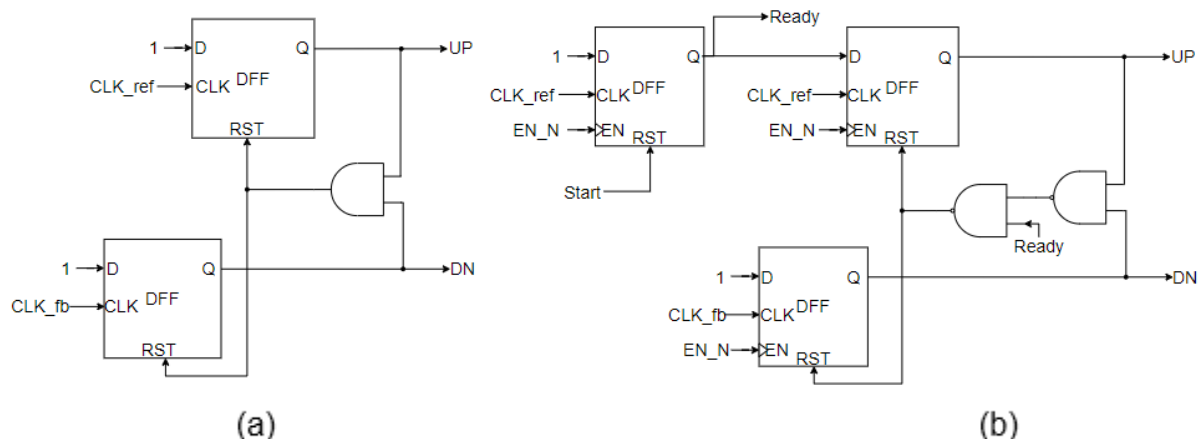
3.3.1 Start-controlled phase detector

Figure 3.9 Phase detector architecture
(a) Conventional edge-triggered PD (b) Start-controlled PD[24]

A widely-used topology for PD is the conventional edge-triggered one shown in Fig 3.9(a). The positive edge-triggered D flip flop(DFF) allows the UP or DN signal to switch to high when meeting the rising edges of the input clocks. Once the UP and DN are both high, the AND gate output becomes high and send the RST signal to drive both UP and DN signal to low. With the falling stages triggered by the same signal RST, the difference of rising edge between UP and DN indicates the phase difference of CLK_ref and CLK_fb. If the CLK_ref and CLK_fb are aligned, the UP and DN signal would both have a short pulse because of the gate delay from input of AND to its output. An appropriate design could utilize the short pulse to turn on the switch of CP, which avoid that the phase difference is too small to generate an adequate output voltage. Although the dead zone could be reduced in the edge-triggered PD, the conventional topology still suffers from the false locking problem. Due to the periodic property of a clock signal, the conventional PD could not identify the difference between $\Delta\varphi = 0$ and $\Delta\varphi = 2\pi$. Thus an extra DLL is added to implement a start-control function, along with two NAND gates instead of the AND gate.
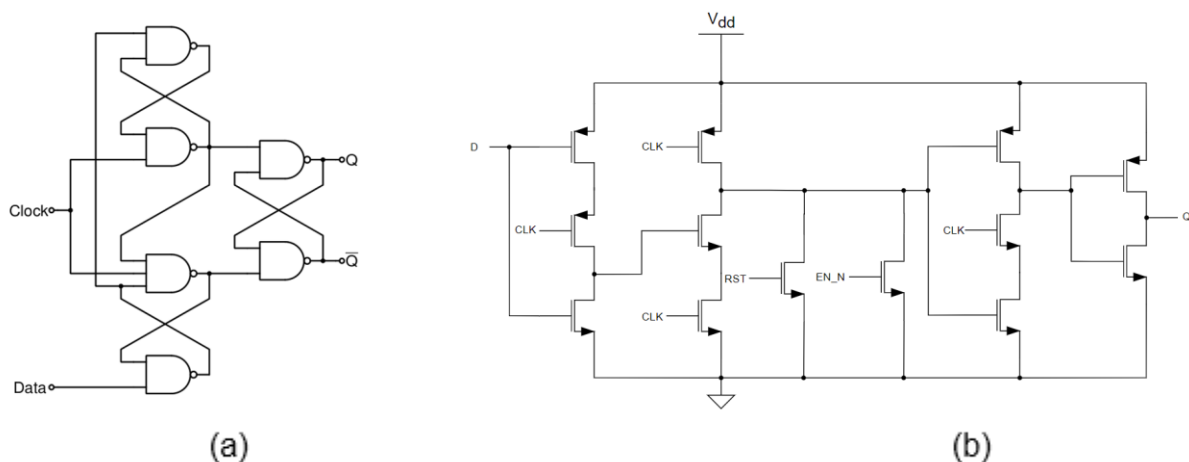

Figure 3.10 DFF topologies (a) Conventional digital DFF (b) TSPC DFF[24]

Before introducing the working principle of the start-control function, the DFF used in this

topology should be illustrated first. Instead of a conventional digital DFF composed of 6 NAND gates, a true single-phase clock (TSPC) DFF is implemented. The conventional structure would introduce a long gate delay from input to output which will degrade the detection resolution of PD. It will increase the dead zone and must be avoided. The TSPC DFF has a fast speed of operation either from input CLK to output Q or the RST to output Q, and the number of MOSFETs is also reduced. Through simulation, the delay from CLK to Q only has 70ps delay and RST to Q has a delay of 165ps.

A simulation with $\Delta\varphi = 74ns$ (1.76π) is shown in Fig 3.11(a) to explain how the start-control is working. In our design, the delay difference between CLK_ref and CLK_fb should be $\Delta\varphi = 84ns$ (2π) to have proper output clocks. The DN signal should have a $10ns$ on-time longer than the UP signal. If a conventional PD is implemented, the current situation would have an output that the UP signal has a $74ns$ longer than the DN. Because the first rising edge of UP will be triggered earlier than the DN signal. In order to have a proper output waveform, the DN signal should have a rising edge earlier than the UP which means the first rising edge of UP should be ignored. Before the start signal is set to high, no output is generated either in UP or DN. When the start signal is set to low, a ready signal will be generated at the next rising edge of CLK_ref, activating the DFF which controls the UP signal. Although the ready signal and UP both tracking the CLK_ref, the ready signal has a rising edge later than the CLK_ref through the DFF. The first rising edge of UP will always be one clock cycle later than the ready signal. Thus the first rising edge of UP is ignored and a proper output of UP and DN is ensured. When in Fig 3.11(b), the CLK_fb is later than the expected position, the UP will have a rising stage before the DN since the ready signal has been set to high at the second clock period.
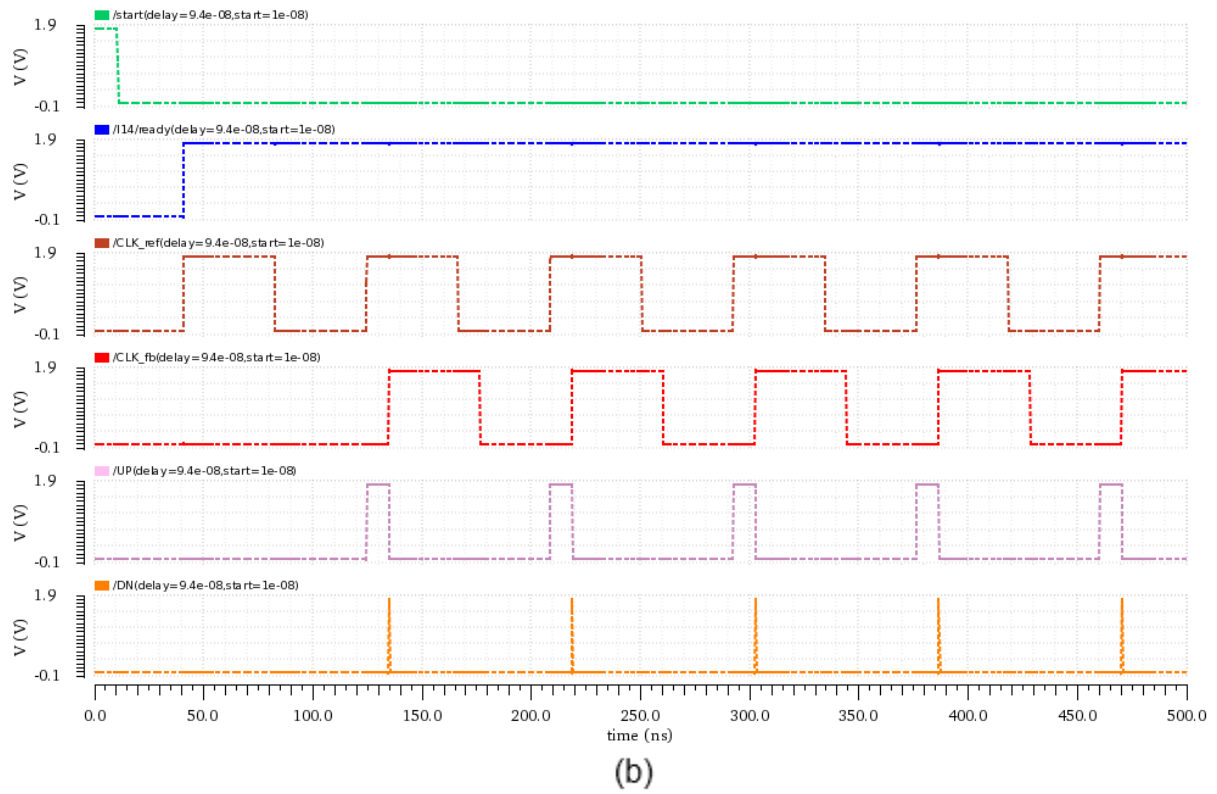

(a)

(b)
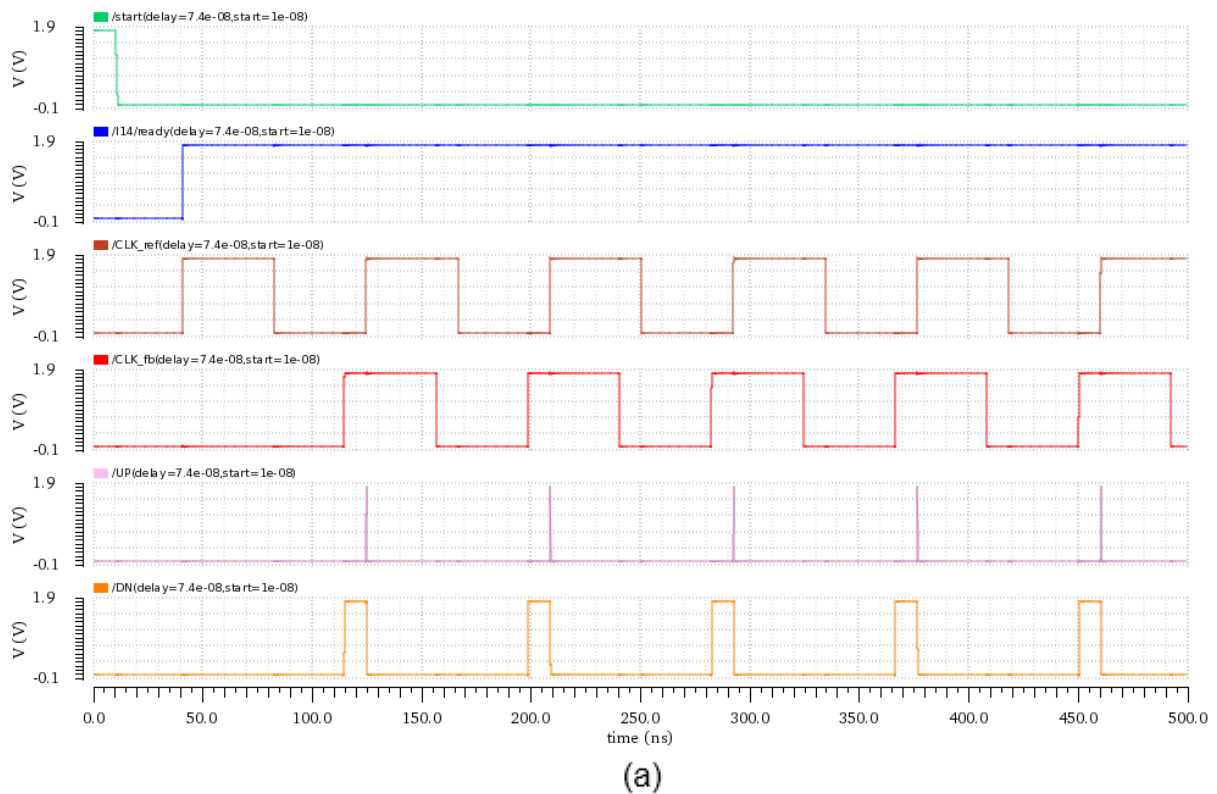
Figure 3.11 PD output simulation
(a) $\Delta\varphi = 74ns$ (1.76π) CLK_fb ahead of the expected position
(b) $\Delta\varphi = 94ns$ (2.24π) CLK_fb later than the expected position
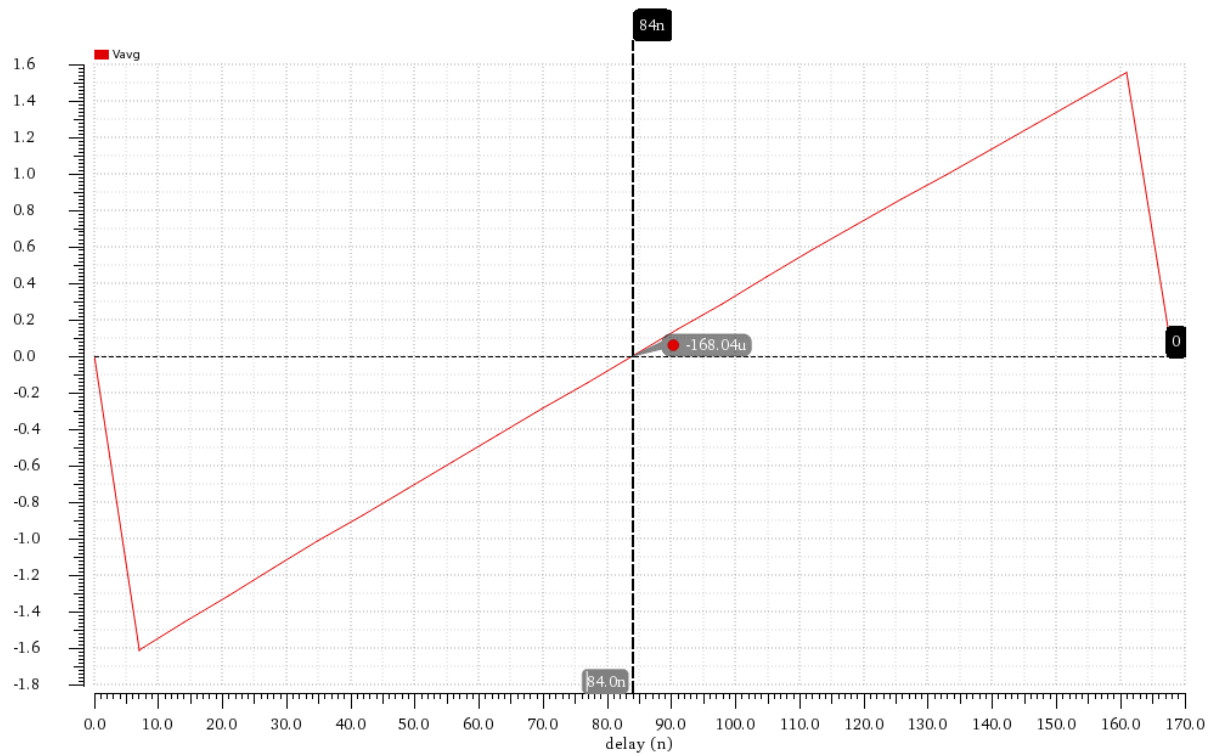

Figure 3.12 Characteristic curve of start-controlled DFF

The characteristic curve shows a good linearity in the $(0, 4\pi)$ detecting range which allows a

proper transition of the phase difference information from PD to CP. The non-zero value of the average voltage at $\Delta\varphi = 84ns\ (2\pi)$ is due to an error from input clocks which is shown below in Fig 3.13. The CLK_ref and CLK_fb have a 7ps error between them when the setting of the phase difference is 84ns, which leads to a bit of mismatch between UP and DN. It illustrates that even few-picoseconds phase difference could be detected, resulting in an extremely small dead zone. As mentioned above, the UP and DN will have a short pulse period during which the charge pump switch could be turned on. The time of this pulse contains two parts of delay, the NAND gate delay and the DFF delay. In the circuit design, the NAND gate delay (from UP/DN to RST) is expected to be as small as possible which could increase the operation speed. The balance of delay from NAND gate input to its output is also of great importance. The RST will be triggered either by UP or DN signal, which could have a different NAND gate delay with a conventional NAND gate topology. To avoid the error causing by NAND gate, a balanced topology in Fig 3.14 is implemented. Unlike the NAND gate delay, the DFF delay (from RST to UP/DN) should be kept as a relatively high value. Because the ready signal in the start-control function should be later than the rising edge of CLK_ref. If the DFF delay is short,(<80ps in simulation), the first rising edge of the UP signal could not be ignored to have a proper output. According to simulation, the NAND gate delay is 84.4ps and the DFF delay is 158ps. The power consumption of start-controlled PD is 20.3uW.
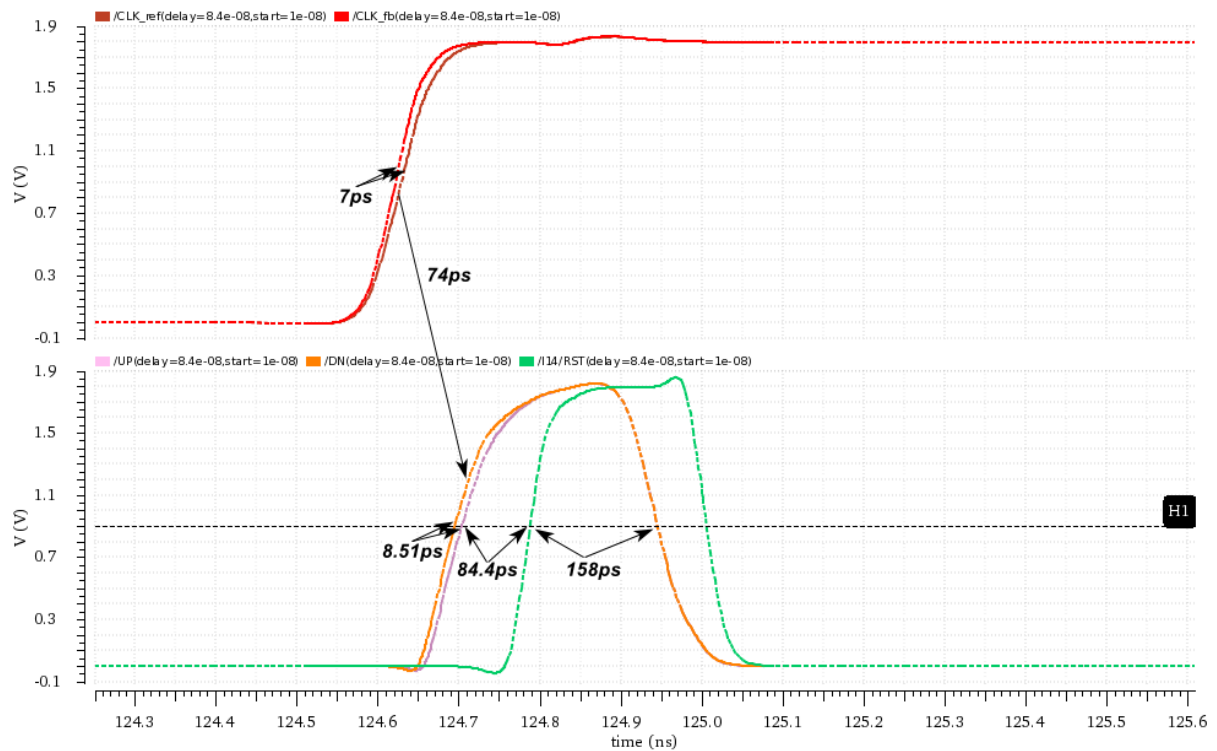

Figure 3.13 PD output with a phase difference $\Delta\varphi = 84ns\ (2\pi)$
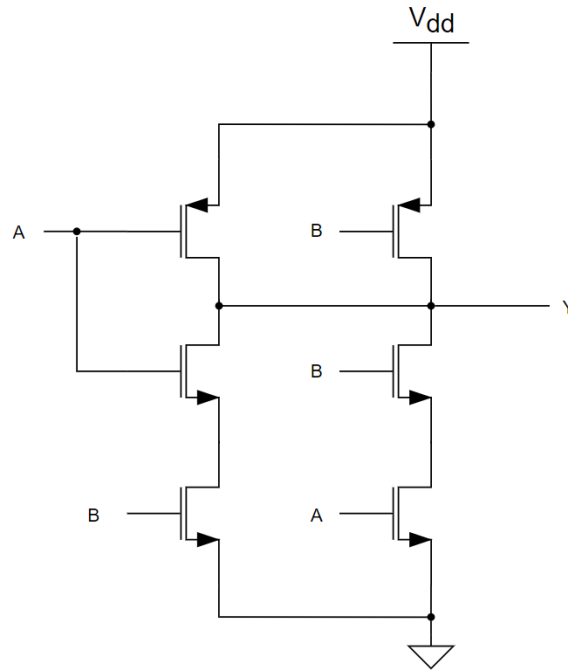
Figure 3.14 Balanced-delay NAND gate

3.4 Charge pump

The function of the CP is to convert the phase difference information to the control voltage on the LF capacitor. The working principle has been shown in Fig 3.6. Two switches responsible for the current sources I_{UP} and I_{DN} are controlled by UP and DN signals separately. The on-time difference between UP and DN would decide the direction of output current which is either charging or discharging the LF capacitor. The current value of either I_{UP} or I_{DN} affects the loop gain, bandwidth and lock-in time of DLL. When the CLK_fb and CLK_ref are aligned, the ideal output current should be zero. In the design of PD, the topology implemented in this design delivers a short pulse for both UP and DN when CLK_fb is aligned. During the period of this short pulse, the switches of two current sources in CP will both be opened. The requirement is more strict for PMOS current source to match the NMOS current source, also considering the effects from the two switches. The non-idealities in the design of CP are listed below.

3.4.1 Non-idealities in the Charge pump

The first one is the current source mismatch as mentioned above. The two switches are also implemented by PMOS and NMOS separately, which have different switching speed. During the turn-on time, the error from the current sources and switching speed would result in a variation on the control voltage V_{Ctrl} and affect the VCDL. Using the transmission gate to re-place the single MOS switch could help to minimize the error. In the PD design, the generation path for UP and DN signals are identical, so the error on the triggering time by either UP or DN signal is small.

Charge injection and clock feed-through are shown in Fig 3.15. When the MOSFET switches are turned on, there is a certain amount of charge Q_{ch} held in the channel inversion layer. The stored charge will inject into the drain and the source port when the MOSFET is turned off. Both drain ports of two switches are connected to the LF capacitor, so the injected charge will cause the variation in the control voltage. Another issue between the drain port and the node of the LF capacitor is the clock feed-through. The switching behaviour of the UP/DN signal will be coupled to the output node because of the gate-drain parasitic capacitors C_{GD}. A step error occurs on the control voltage during the transition of UP/DN. Reducing the size of the MOS switches could reduce the charge stored in the channel. To avoid the clock feed-through, the control switches should be placed away from the output node.
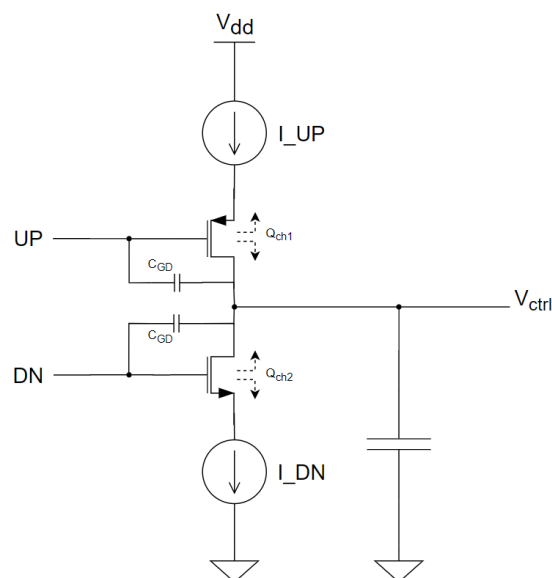
Figure 3.15 Charge injection and Clock feed-through

The charge sharing is also an issue that happens during the transition of UP/DN. As in Fig 3.16, the two current sources M1, M2 have the parasitic capacitors C_X and C_Y. When the UP/DN controlled switches are turned on, the charge on C_Y will flow out and the C_X will start to charge. The charge which should be sent to the output capacitor C_P are shared by the three capacitors. With a high-frequency signal, the glitches caused by charge sharing will be obvious on the output control voltage and result in the phase change in VCDL.

Figure 3.16 Charge sharing[25]

3.4.2 Design of Charge pump circuit

Figure 3.17 Concept of the charge pump in this design[26]

The block diagram of the proposed charge pump is shown in Fig 3.17. Unlike the conventional structure with only one pair of current sources, two extra stand-by current sources are added. The operation of this charge pump is similar to the conventional one. When the UP and DN are at the s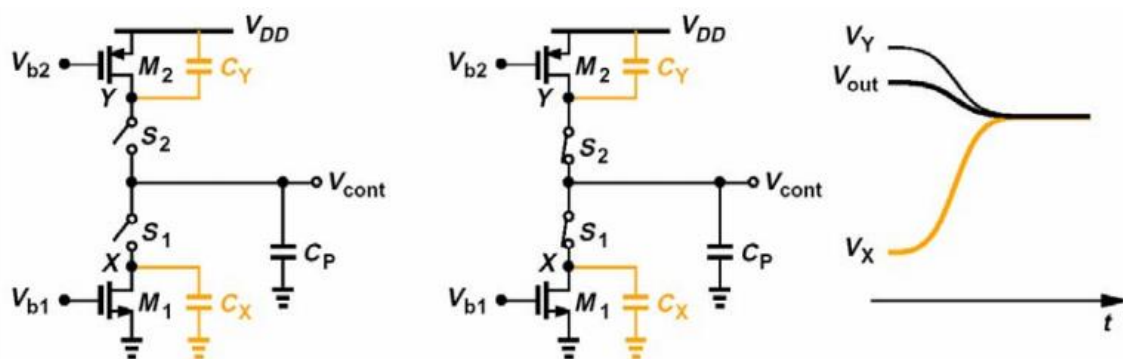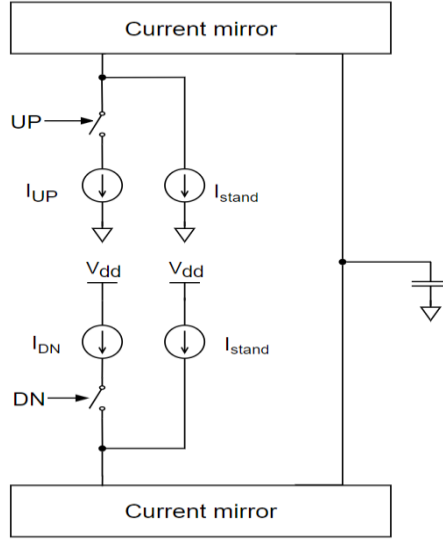ame state, either on or off, the total output current is kept as zero. Only when the UP and DN are in a different state, the output current will be generated. In this topology, the switches are not directly connected to the output node, which reduce the effect of clock feed-through. The current sources I_{UP}, I_{DN} and I_{stand} are copied from the same current source to achieve a better match among them. These two I_{stand} current sources are used to speed up the response of CP. Because there is always a current pass through the output node, any change in the current value will be quickly transferred to the output capacitor. Besides, an always-on path from VDD to GND eliminates the effect of charge sharing when UP/DN is switching. The detailed circuit of the CP and the standard current source are shown in Fig 3.18.

The transistors M1, M2, M11 and M12 are the current mirrors copying I_{UP}, I_{DN} and I_{stand} to the output node. When UP and DN are in the same state, the current on these four transistors are equal to I_{stand}. When the UP and DN are in a different state, the current changes to $I_{M1,M2} = I_{stand} + I_{DN}$, or $I_{M11,M12} = I_{stand} + I_{UP}$. The switches are implemented with two transmission gates to minimize the effect of different switching speed for NMOS and PMOS. Transistors M13, M14, M15, M16 and the resistor R1 compose a Widlar current source as a standard for I_{UP}, I_{DN} and I_{stand}. The current is given by:

$$I_{ref} = I_{M13} = \left(V_{GS13} - V_{GS14}\right)/R_1 \tag{3.3}$$

The reference current created by the Widlar current mirror is stable and only depends on the gate-source voltages of MOSFET and the value of resistor R1. The main problem of this current reference is the multiple operating points. Except for the value in Eq 3.3, the Widlar current source has another stable operating point at zero $I_{M13} = I_{M14} = 0$. It could happen because

the current of M14 is copied from M15 through M16, while the current of M15 is the same as M13 which is generated from M14 an R1. A start-up circuit is required to break the stable state at $I_{M13} = I_{M14} = 0$ by injecting current into this loop. Transistors M17, M18 and resistor R2 form the start-up circuit. When the loop of M13 to M16 are locked in a zero-current state, the gate-source voltage of M13 is zero which forces the source terminal of M17 to zero as well. The gate-source potential of M17 is not zero now and turns on M17 to inject current into the loop of M13 to M16. With the current injection, the gate-source voltage of M13 gradually increases while the gate-source voltage of M17 is reduced. When the loop reaches the stable point $I_{ref} = (V_{GS13} - V_{GS14})/R$, the transistor M17 is totally off and does no effect on the loop.
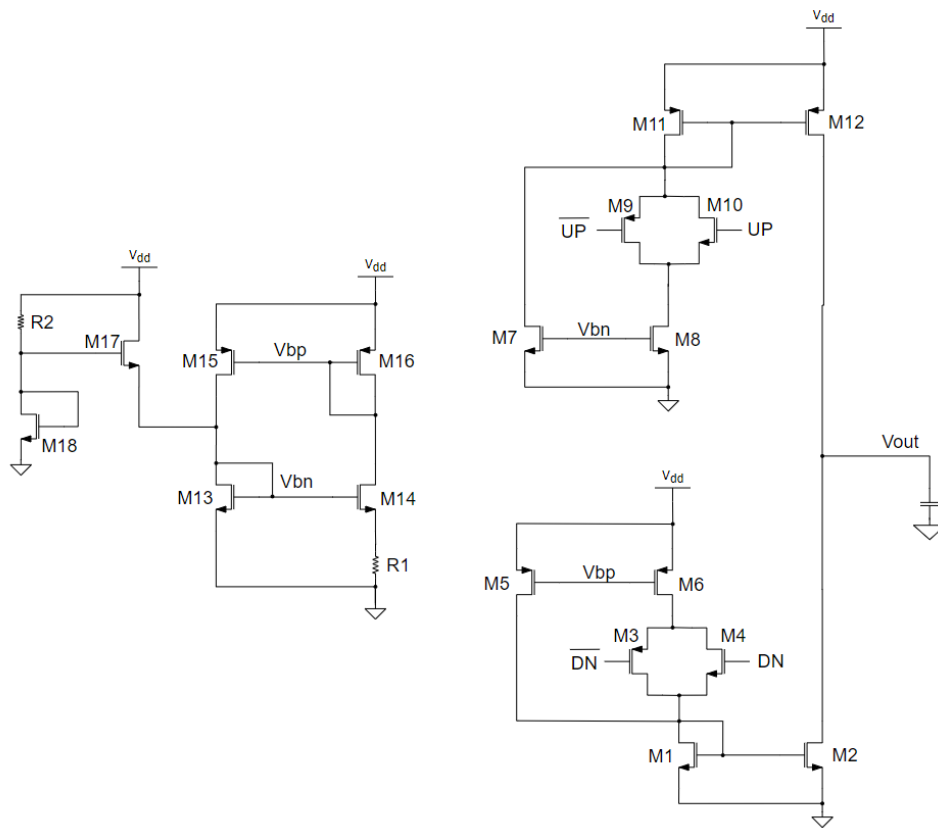

Figure 3.18 Charge pump in this design

The performance of the CP is shown in Fig 3.19. The input signals UP and DN are from the PD with a phase difference from $0\ ns$ to $168\ ns$ $(0 - 4\pi)$. It can be seen that the output voltage V_{ctrl} could reach a stable value at a high speed, most of the curves smaller than 0.5 us. Because the simulation is conducted only with PD and CP, the CP is always charging or discharging due to the PD output. When V_{ctrl} is from 0.5V to 1.5V, there is a relatively flat period in the curve which indicates the UP and DN are in the same state. With the increase or decrease of the output voltage, the curve is not flat during the same state for UP and DN, which is caused by the unbalance of current sources. The curves on the top-half plane indicate the situation that UP>DN. In the PD design, the first rising edge of UP is later than the DN signal, which makes all the curves for UP>DN are triggered later. The weird green curve is a delay difference(4π) out of the detecting range, which gives an incorrect response.
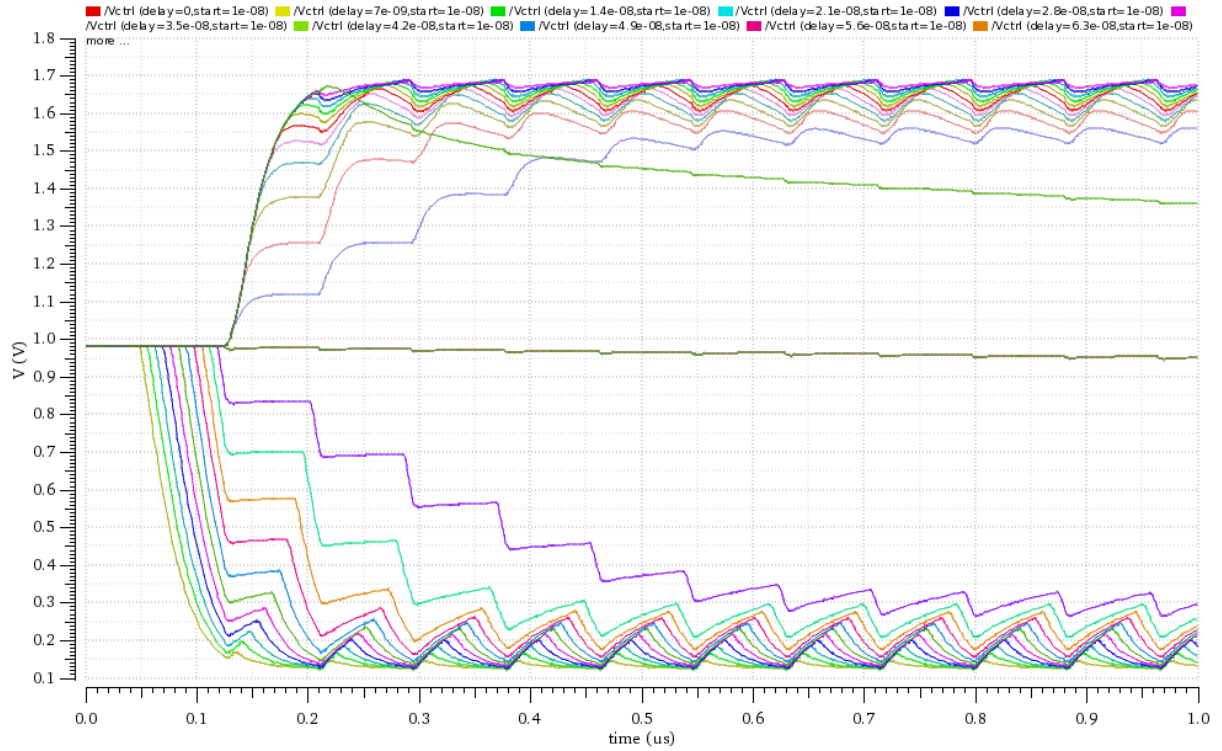
Figure 3.19 CP simulation result, $\Delta\varphi = 0$ to 4π

A DC simulation of the output voltage has been down to analyse the matching of current sources. The output voltage is swept from 0 to 1.8V, and the source terminal current of M2 and M12 are tested. If the two switches for UP and DN are turned off, the simulation result is Fig 3.20. Only the standby current sources I_{stand} are mirrored to the output node. These two current sources show a good matching from 0.5V to 1.5V, which is related to the flat period in Fig 3.19. When the output voltage is out of this good-matching range, the standby current sources are unbalanced. Thus the flat period of the output voltage does not exist outside the range from 0.5V to 1.5V. However, in Fig 3.19, the curve in the middle indicating the lock-in state also suffers a subtle shift even the output voltage is around 1V. As mentioned above, the output of PD, UP and DN would have an equally short pulse when the system is in the lock-in state. The switches for I_{UP} and I_{DN} are turned on during the short pulse. The simulation with turned-on switches is shown in Fig 3.21. Because now the current mirror copies both currents from I_{UP} and I_{stand}, the matching performance is degraded. Although the pulse period is short, after several cycles of operation, the output voltage gradually drifts a bit from the original value. The benefit is that when the output voltage is far from the balance point, such as at 1.8V, the attenuated value of I_{UP} will increase the speed of CP to reach the balance point. For the total performance of CP and DLL, if the DLL is in the lock-in state, more than 99% of the time both UP and DN switches are off which indicates the matching of two I_{stand} sources shown in Fig 3.20. The poor matching situation in Fig 3.21 only happens in the short pulse period during which both switches are turned on. Thus the current sources matching for this CP is generally good.

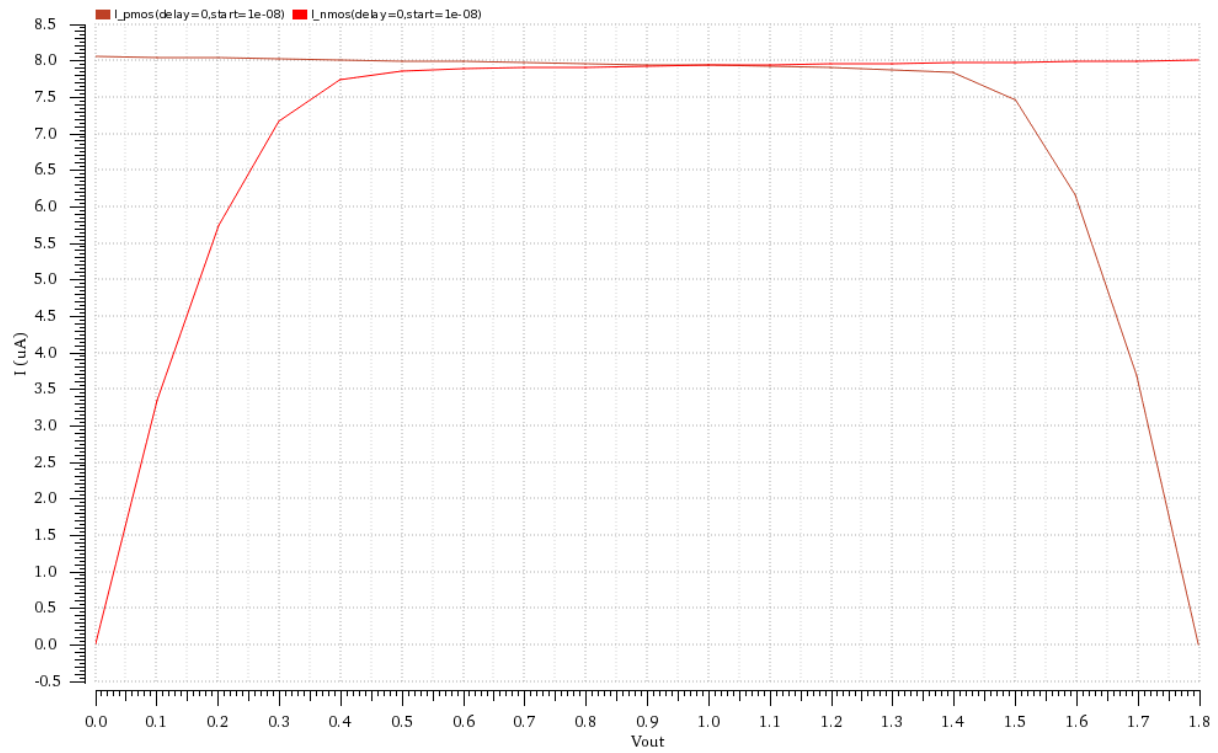The total power consumption of the CP is 87uW under 1.8V supply.

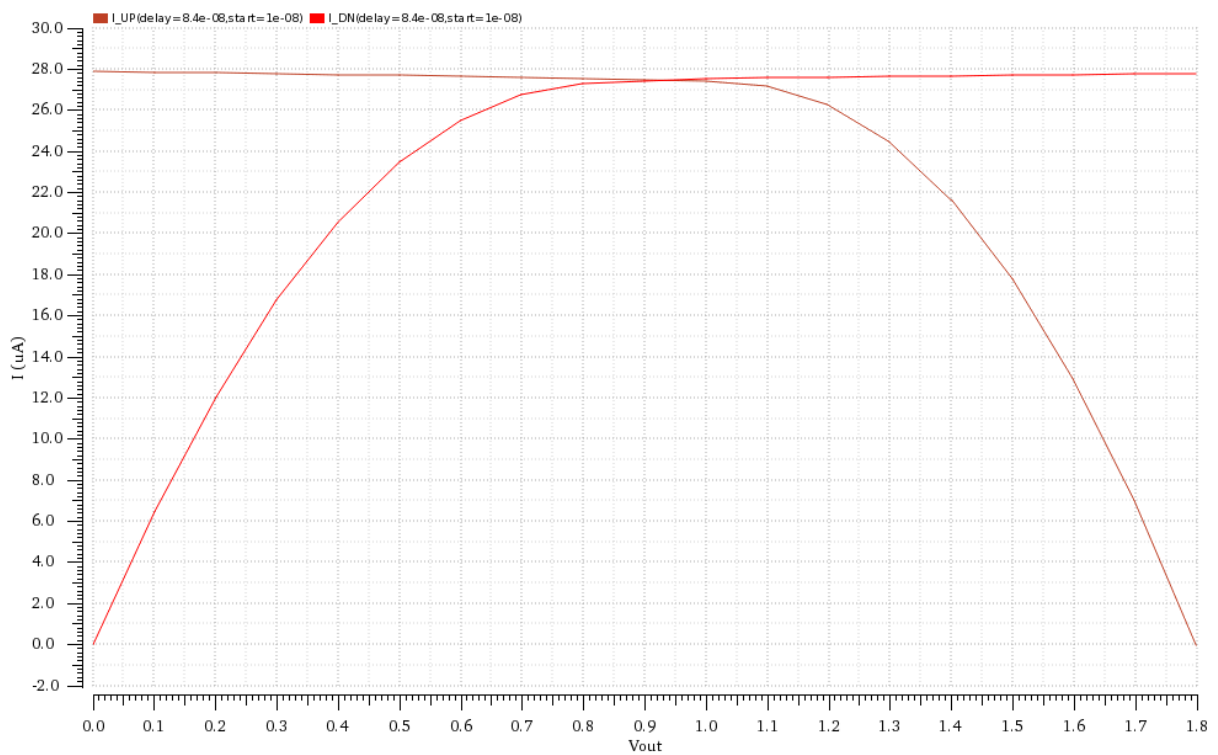*Figure 3.20 Simulation of I_{stand}*

Figure 3.21 Simulation of $I_{stand} + I_{UP/DN}$

41

3.5 Voltage-controlled delay line

The VCDL is of great importance in a DLL since the delay signals are all directly generated by the VCDL. The function of VCDL is to modify the delay time of each stage inside VCDL based on the control voltage from the CP. Unlike the VCO in the PLL, a VCDL itself is an open-loop configuration and thereby has no oscillation and less jitter. With analog control, the VCDL is typically a chain of cascaded delay elements. Basically, all the delay elements are designed to be identical. The number of stages is normally decided due to the number of output in the application. The main purpose of our DLL is to generate four reference clocks, thus four stages of delay elements are required. The output clocks range from $0°$ to $180°$ with $45°$ as one step. Therefore, the delay time of each delay element should be:

$$\Delta t_{delay} = \frac{(T/2)}{4} = 10.5ns \qquad (3.4)$$

Delay element is the core of a VCDL. Two typical technique for designing delay elements are: shunt capacitor delay element and current-starved delay element. Their topologies and the delay range are shown in Fig 3.22. Both of these two types of delay elements are based on the inverter. However, the main contributor to delay time in the shunt capacitor method is transistor M2, acting as a capacitor. The current charging or discharging M2 is controlled by transistor M1 whose gate voltage is the control voltage V_{ctrl}. The benefit of the shunt capacitor topology is the linear delay transfer curve in a whole range of V_{ctrl}. However, the delay time of the shunt capacitor delay element is small, normally less than 1ns under a 1.8V supply. To further increase the delay time, large devices are required which consumes more power and area. Besides, the shunt capacitor approach also suffers from a narrow tuning range of the delay which limits the speed of DLL to reach the lock-in state.
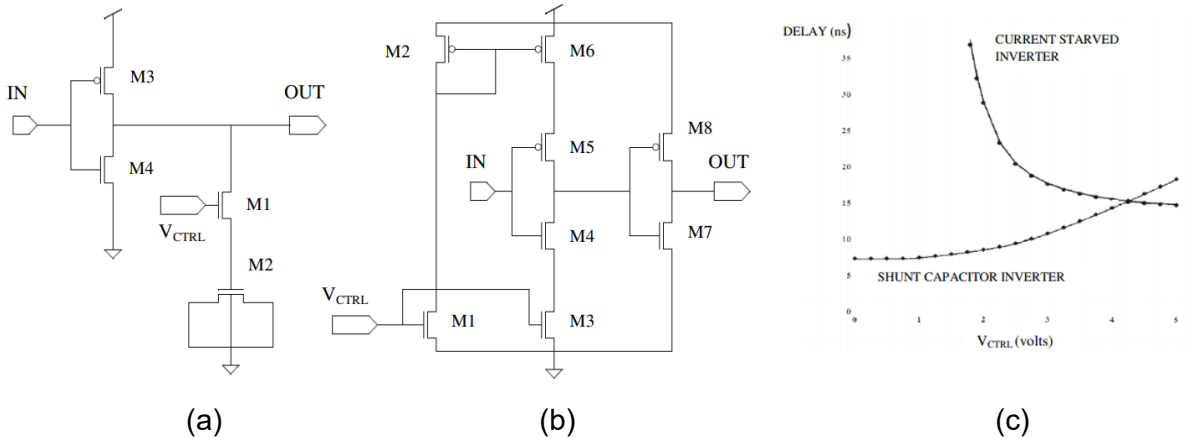

(a)   (b)   (c)

Figure 3.22 Typical delay elements and their delay range[23]
(a) Shunt capacitor (b) Current-starved (c) delay range

The current-starved delay element is composed of two inverters. The delay time is dependent on the current charging or discharging the parasitic capacitance between the two inverters. The first inverter containing M4 and M5 are connected to two current sources, M3 and M6. The output current of this inverter is limited by the current sources, unlike the second inverter containing M7 and M8, which could drive as much as current from the supply. Both current sources are controlled by V_{ctrl}, either directly or through the current mirror. The current-starved topology achieves a wider delay tuning range compared to the shunt capacitor. However, the

control voltage in the current-starved one should be higher than the saturation voltage for M1 and M3 because the current source should be kept in the saturation region. When the V_{ctrl} is smaller than the saturation voltage, the VCDL does not have any output and the DLL will not work. Thus an initial value should be set for V_{ctrl} to keep the whole device out of the shutdown situation. In this design, the current-starved topology is preferred to generate the 10.5ns delay time.
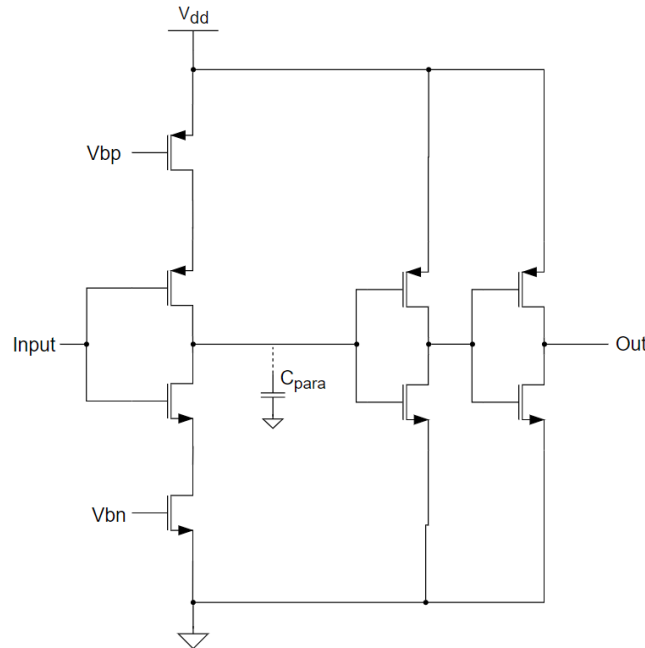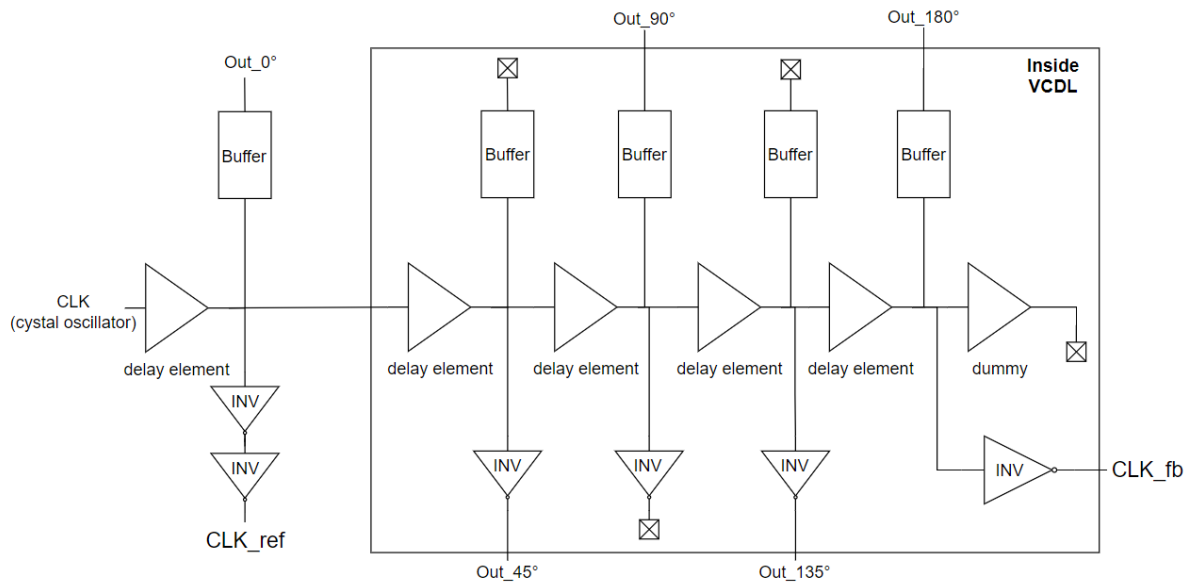

Figure 3.23 Delay element in this work

Figure 3.24 The VCDL structure

The architecture of the delay element and the VCDL in our design are shown in Fig 3.23 and Fig 3.24. An extra inverter is added to a normal current-starved delay element. The extra inverter is to compensate for the duty-cycle variation caused by the second inverter stage in the

delay element. The variation is affected by the process corners and degrades the precision of the output clocks and the jitter performance. However, the extra inverter inside the delay element will invert the signal, which requires an outside inverter(INV) to achieve the final output signals. Thus the $45°$ and $135°$ clocks are the outputs of the inverter, while the $0°$, $90°$ and $180°$ clocks are the outputs of the buffer which does not invert its input signal. Since the delay time in VCDL is highly sensitive to the node capacitance, extra INVs, buffers and the dummy delay element are all added to achieve the same node-capacitance for each delay stage. Besides, the PD in our design only compares the rising edges of CLK_fb and CLK_ref, which ignores the effect caused by the falling edges. Using a signal-inverted delay element, the error from losing falling-edge control is reduced. The delay time of both rising edge and falling edge are simulated in Fig 3.25.

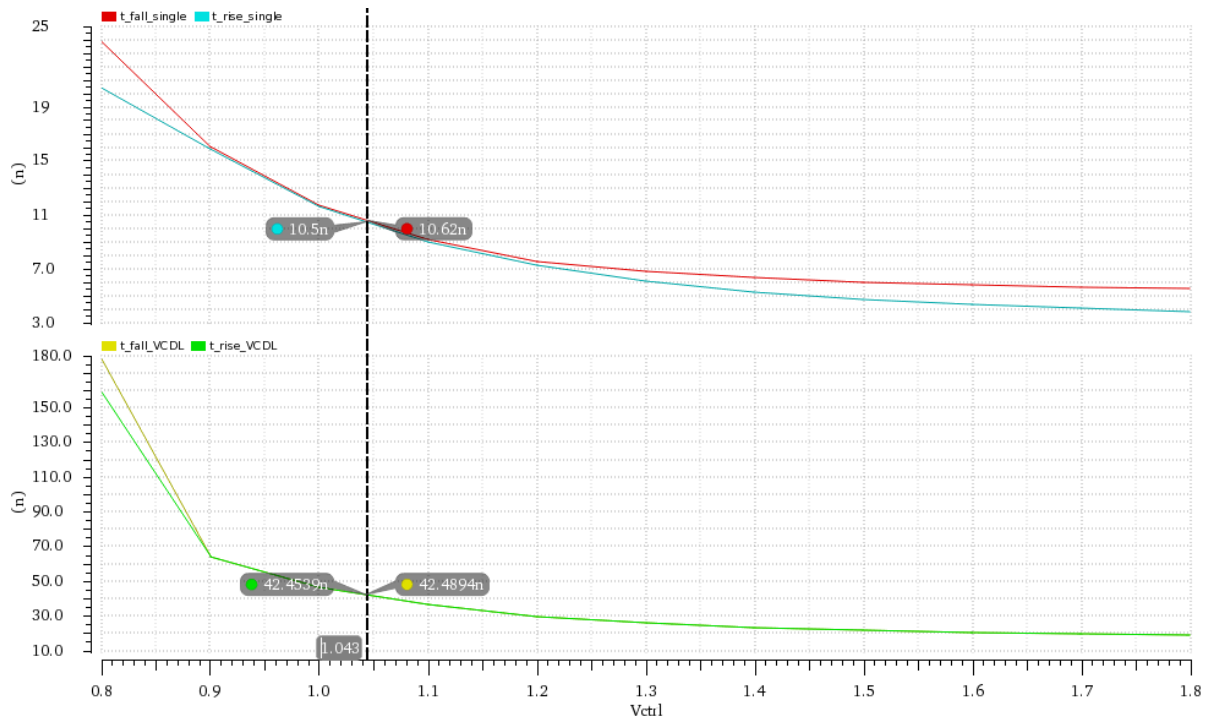*Figure 3.25 Delay time of the single delay stage and the whole VCDL*

The V_{ctrl} could vary from 0.8V to 1.8V, according to a 10ns to 180ns delay range for the whole VCDL. When the $V_{ctrl} = 1.043V$, the delay time meets the requirement and both rising edge and falling edge have a good match. Simulation under different corners is conducted in the whole DLL in the next section. The total VCDL consumes a power of 53.5uW.

3.6 DLL performance
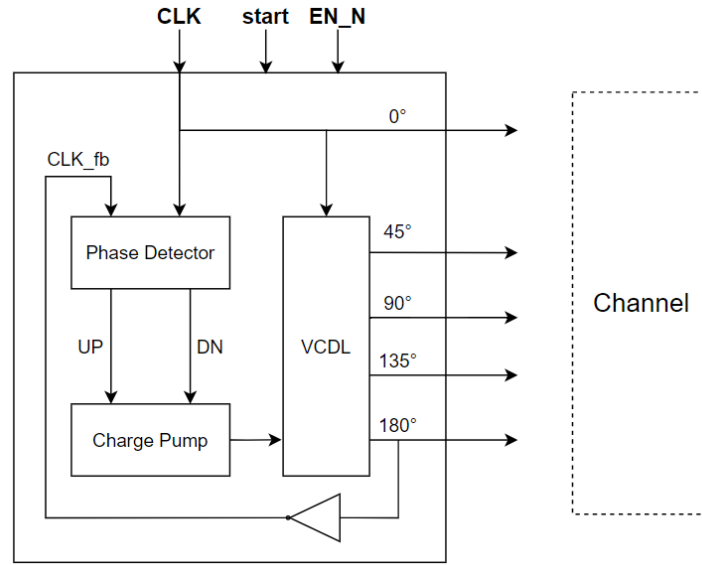
Figure 3.26 The whole architecture of DLL

In this chapter, the DLL providing the reference clocks for the transmitter channel are designed. Including the original clock, in total five clocks ranging from $0°$ to $180°$ with a step of $45°$ are generated. The block diagram of the whole architecture of this DLL is shown in Fig 3.26. The operation simulation is shown in Fig 3.27.
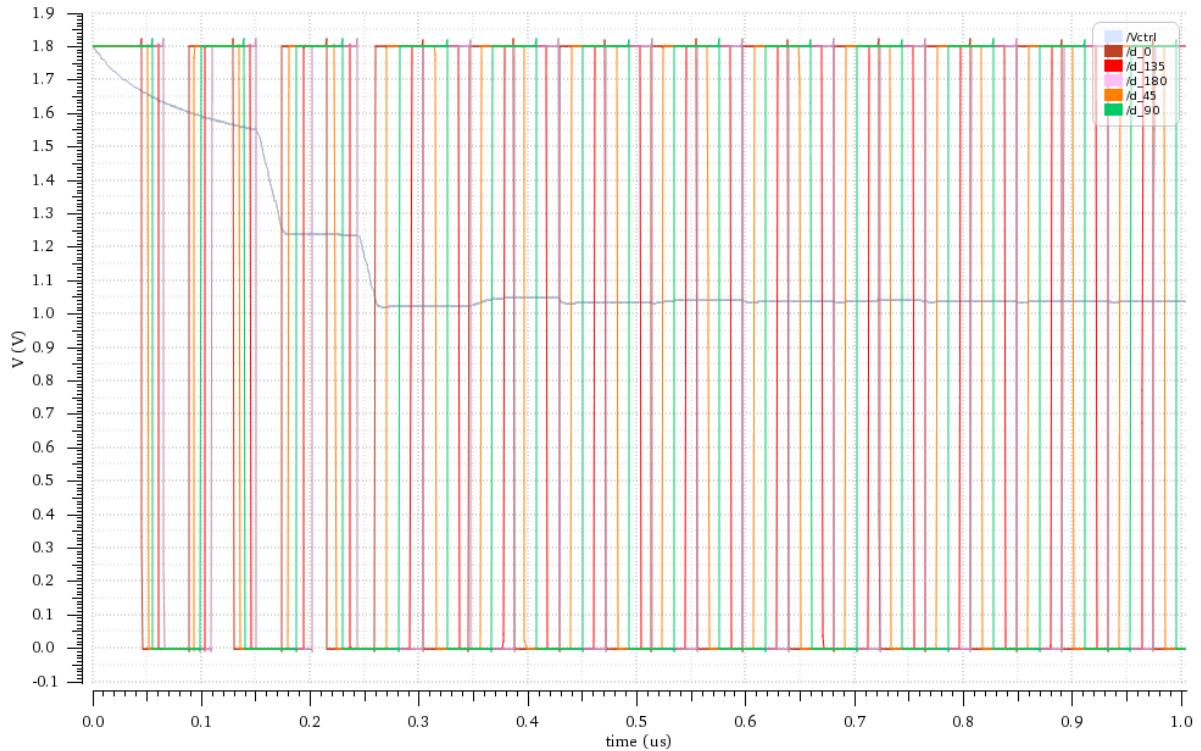
Figure 3.27 Operation of DLL

The DLL reaches the lock-in state in less than 0.5us (<6 cycles). When it is locked, four

reference clocks are generated and the control voltage becomes stable. Compared with Fig 3.19 where the V_{ctrl} suffers from drifting, the negative feedback of DLL retrieve the V_{ctrl} back to the balanced value. The power consumption of the total DLL and all the components are listed in Table 3.1.

Table 3.1 Power consumption of DLL


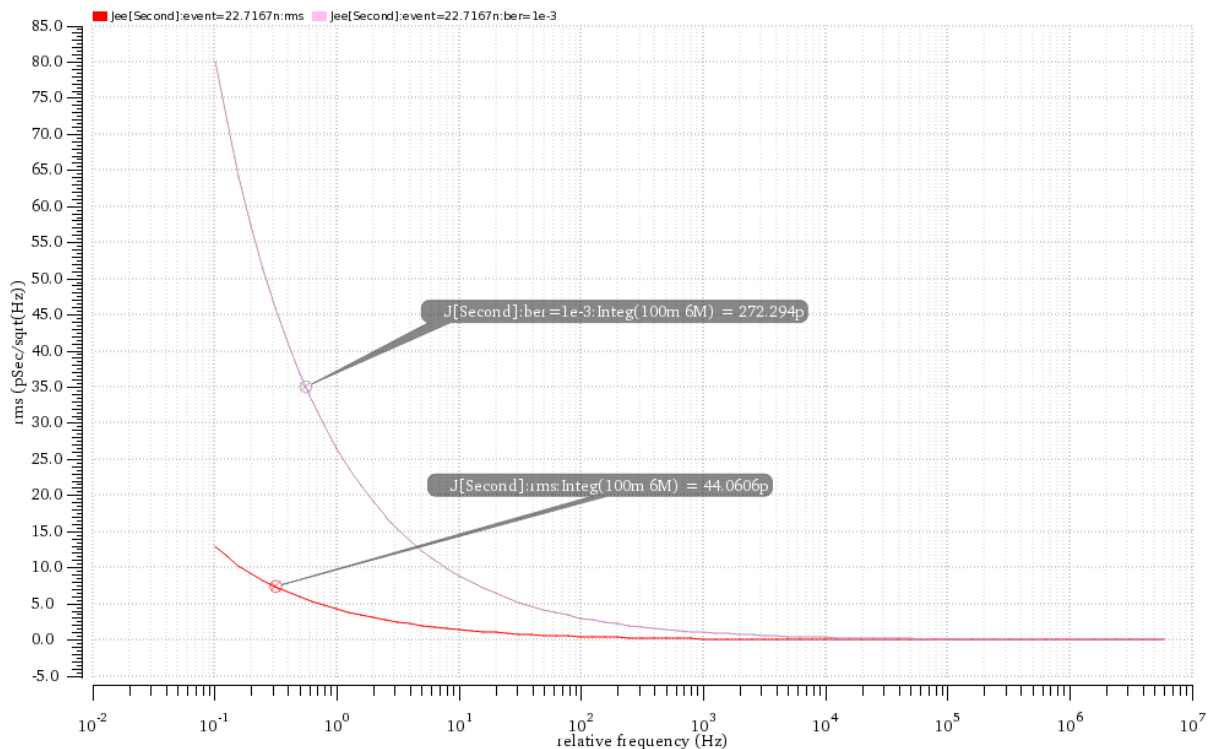|	Current	Power
Phase detector	11.29uA	20.3uW
Charge pump	48.3uA	87uW
VCDL	29.74uA	53.5uW
Biasing and other auxiliary components	11.87uA	21.4uW
Total DLL	101.2uA	0.18mW
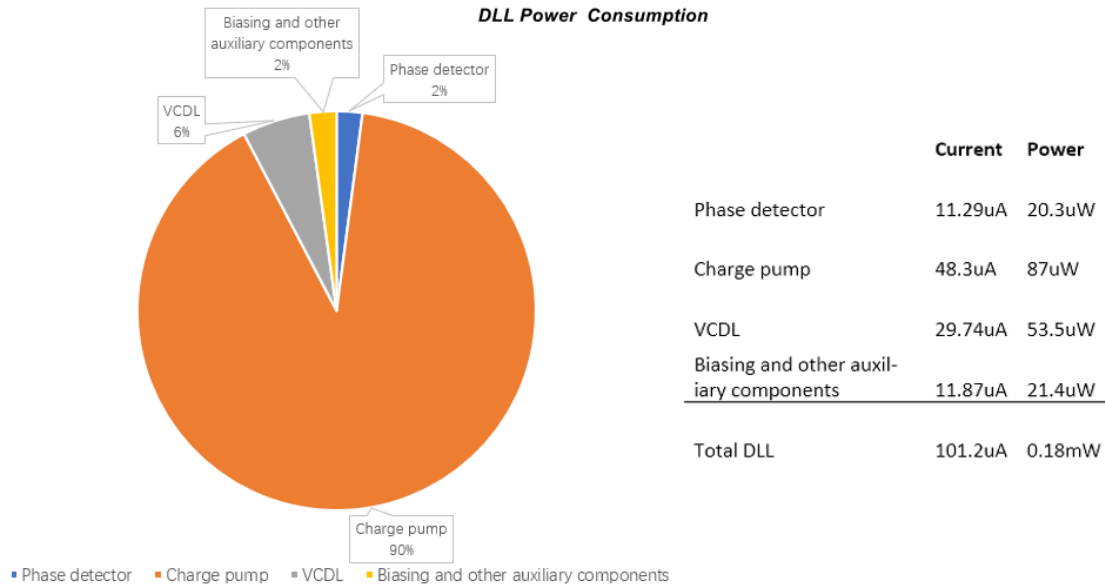
Figure 3.28 Edge-to-edge jitter of DLL

The peak-to-peak value of the jitter is $J_{ee} = 272.294ps < (LSB/2) = 656.25ps$ which meets the requirement. The corner simulations are also conducted, both the delay time variation and the duty-cycle variation.
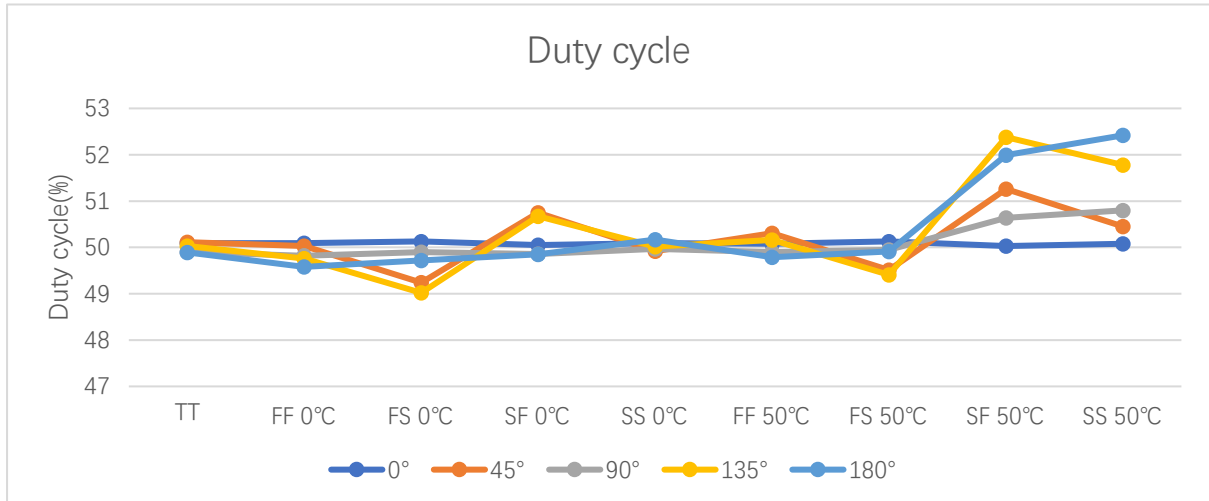
Figure 3.29 Duty-cycle variation

Each output clock is simulated under all the corners at both $0℃$ and $50℃$. Except for the SF and SS corner at $50℃$, the duty-cycle variation is smaller than 1% in other situations. The delay time variation has a similar trend as the duty-cycle variation. A plot for the control voltage under different corners indicates the problem. The change in the control voltage might be caused by a mismatch in CP at $50℃$. Although the control voltage has a big ripple, the function of the whole DLL is still valid and keeps retrieving the V_{ctrl} back, as shown in Fig 3.31.

Figure 3.30 Delay time variation

Figure 3.31 Control voltage under all corners

Chapter 4 Fine delay generation

After the reference clock signals are provided by the DLL, they are sent to each channel where two adjacent phased clocks are selected to generate a finer delay. According to the MATLAB simulation in chapter 2, the resolution of the minimum delay for our proposed system is $1.3125ns$. More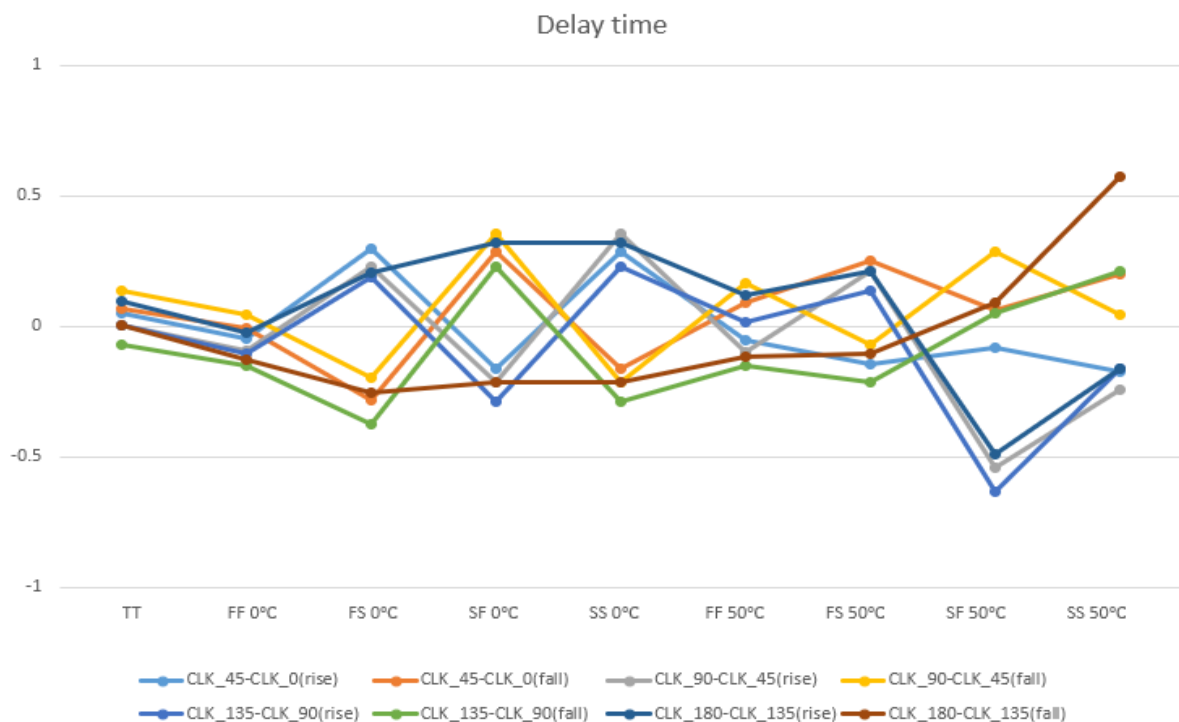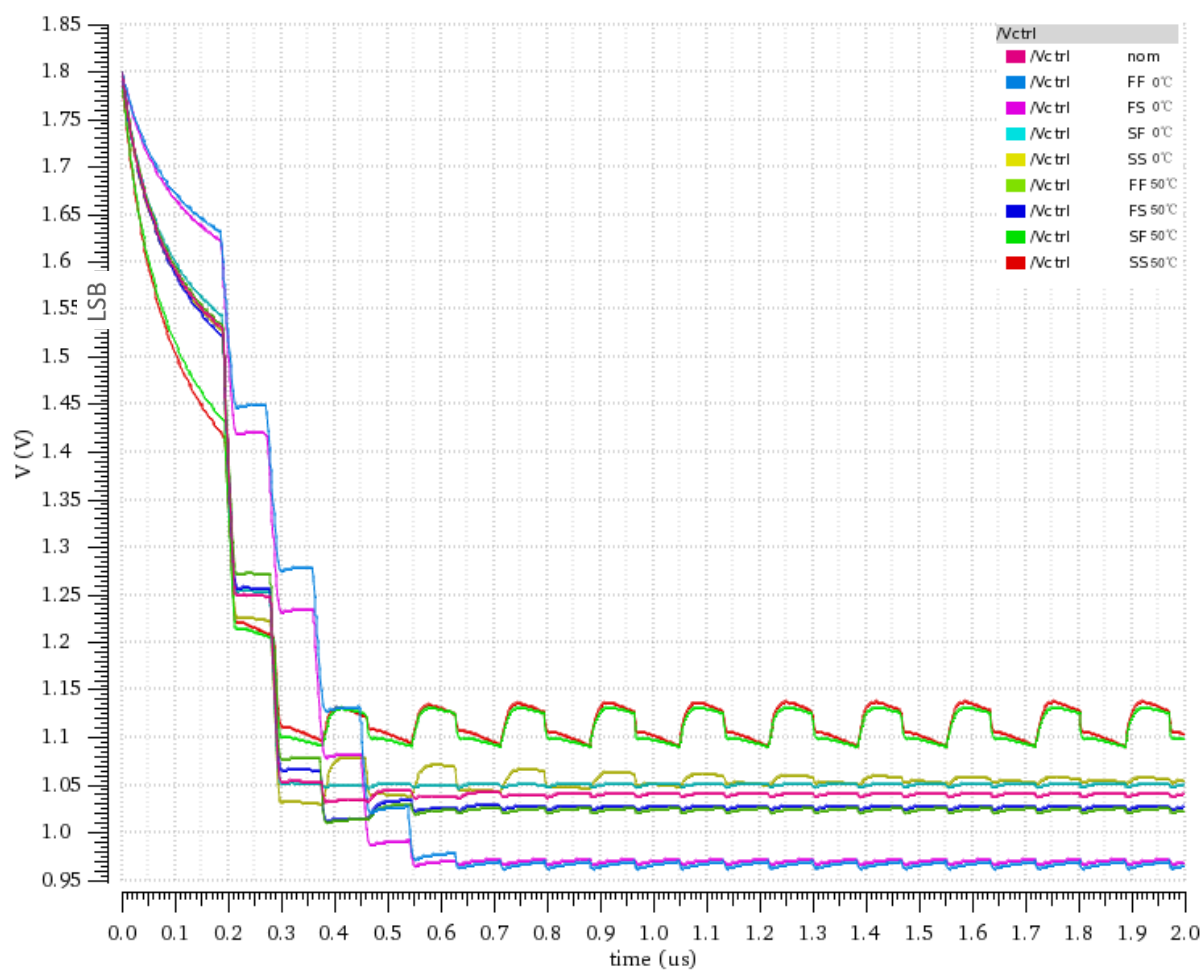over, this fine delay is generated inside one channel, where the power and area are limited. The design of the precise delay device is quite challenging with these requirements. Instead of using delay elements, a phase interpolator(PI) is implemented in this design to achieve the precise delay generation. The arrangement of the fine delay generation for the VNS will be introduced. Then the design of the PI and the thermometer decoder for PI are explained.

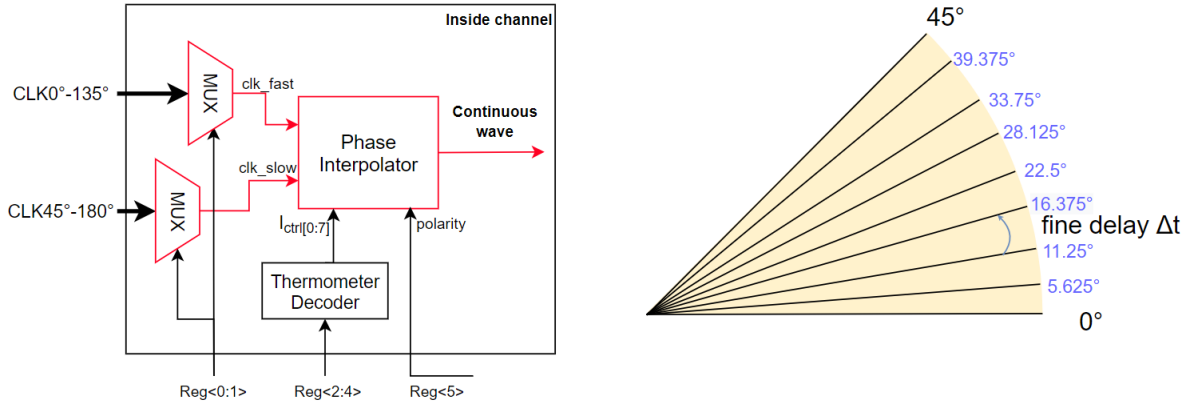## 4.1 Fine delay generation for the VNS

Figure 4.1 Block diagram of the fine delay generation section

As mentioned in Chapter1, the continuous wave is required for the neuromodulation because it could provide a high focal pressure to stimulate the vagus nerve. In a 2D array, the time delay among different channels to get the expected spatial resolution is only limited in one time period $[0, 1T)$ for the continuous wave. The accuracy of spatial resolution depends on the delay-time resolution which is $\Delta t = 1.3125ns(5.625°)$ after quantization. Thus the fine delay generation is the core circuit of a VNS transmitter channel. The channel area is limited therefore a delay circuit like DLL is not suitable to be implemented inside the channel. Without a negative feedback loop, the error in the delay elements is hard to control. Therefore, a phase interpolator(PI) is implemented instead of using delay elements. PI receives two input clocks of the same frequency with phase φ_α and φ_β respectively, and generates a clock output whose phase is the weighted summation of the two input phases. The weighted summation could be easily controlled by a digital input. With a polarity control to invert the input signals, the PI could cover the whole range of delay from $0°$ to $360°$.

In Fig 4.1, the fine delay generation block in our work contains two 4-to-1 multiplexers(MUXs), a PI and a thermometer decoder. Two adjacent reference clocks(clk_fast and clk_slow) with $45°$ phase difference are selected by the MUXs as input signals for PI. Two register bits

Reg<0:1> are responsible for the input signal selection. According to the register setting, the PI could generate 8 levels of output with a phase delay ranging from $0°$ to $39.375°$ compared to the clk_fast. The polarity control is also set by the register and will be set to high when the output delay is in the range from $180°$ to $360°$. The design of MUXs and register are included in the next chapter, and this chapter will focus on the design of the PI and the thermometer decoder.

4.2 Phase interpolator

Although both delay elements and the PI could generate a delayed signal, the working principle between them is quite different. In chapter 3, either the shunt capacitor delay element or the current-starved delay element could really delay the input signal to a certain value. Thus the larger delay time would require larger devices for the delay elements. However, the PI interpolates a phase/delay level between two clock signal with the same frequency but different phases without adding any delay stage. It allows the PI to generate a delayed signal with smaller devices and thus less area consumption, but the PI must have two input signals. More input signals would increase the burden of routing from the DLL to each channel, especially for a large phased array. Therefore a polarity control inside the PI is beneficial to achieve a full-phase range of output. A typical current weighted PI [27] is shown in Fig 4.2. There are two input signal paths and each one has an integer N number of unit cells. Each unit cell has a pair of input MOSFETs, two control switches and a common current source. The value for the current sources are I_{out1} and I_{out2} which have the same magnitude but different phases. To control the output phase, the total current of the PI is kept stable while the number of unit cells connected to different input signal paths is changed by enabling or disabling the control switches I_{ctl}.

The total output current is:

$$I_{out} = n \cdot I_{out1} + (N-n)I_{out2}, \text{ with } n = 0 : N \tag{4.1}$$

With N unit cells, the PI could interpolate N-1 number of phase levels in-between two input clocks and the amount of total output levels is N+1.
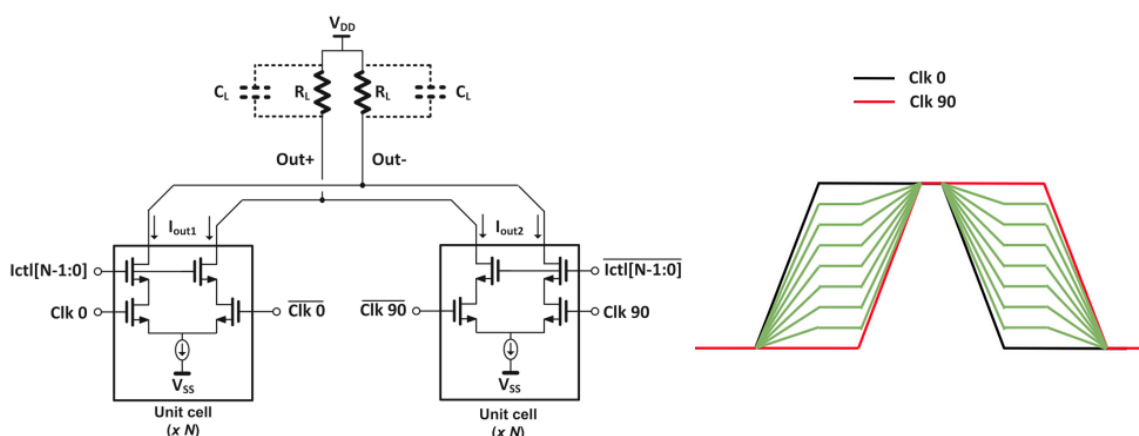

Figure 4.2 Typical current weighted PI and the expected output voltage[27]

As shown in Fig 4.2, if N=8, 7 green curves are interpolated between the input CLK_0 and CLK_90. The slope change of the green curves could intuitively illustrate the different amount of connected unit cells for two input paths. In real operation, the output nodes are connected

to a large capacitor and the transition is smoother than the green curves in Fig 4.2.

The linearity of phase change is the most essential performance factor for the PI. The current sources in the current-weighted PI should be kept identical because any mismatch would directly translate to a phase error at the output. The switching behaviour of input clock signals would cause oscillation of the output voltage. A large output capacitor could help to mitigate this problem but the drawback is the reduced output swing. The output voltage is proportional to $e^{-t/\tau}$ with $\tau = R_{out}C_{out}$. The output capacitor would enlarge the time constant and reduce the magnitude of the output voltage. An inverter could be added as a buffer at the output node to retrieve the waveform back to a square wave. There are some other approaches to improve the linearity such as a better control or a feedback loop, but they are not suitable for the area-limited channel in our design.
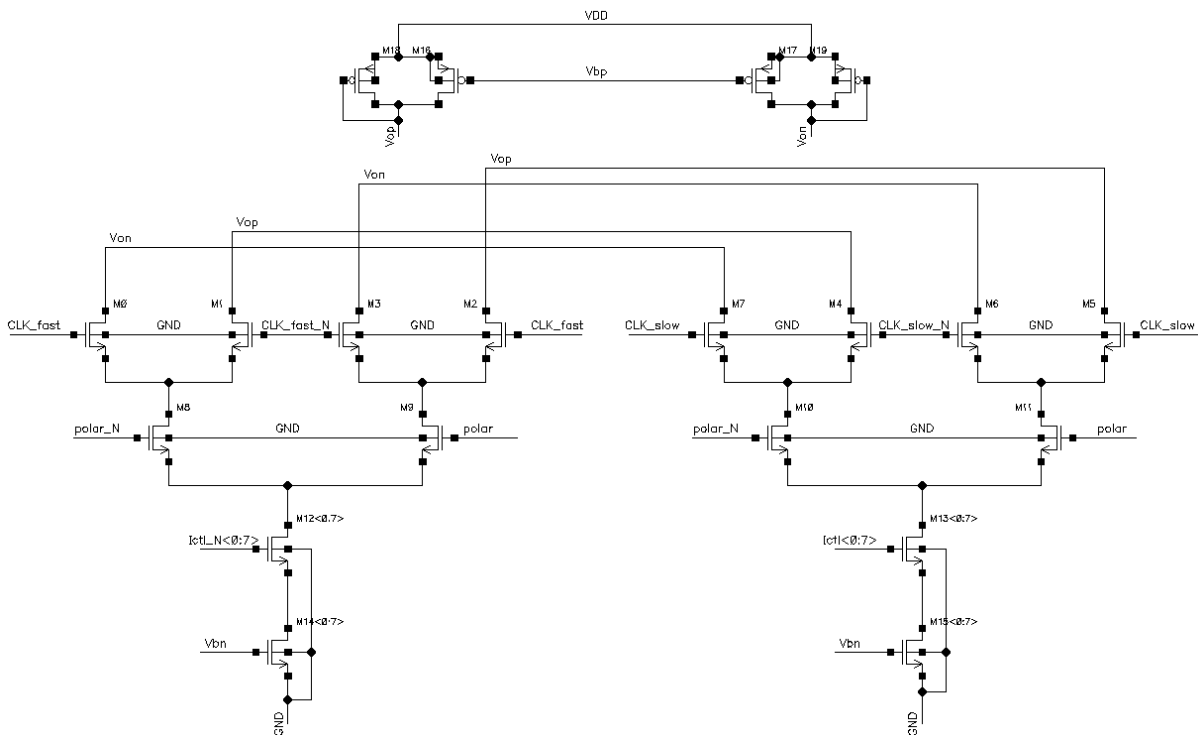

Figure 4.3 PI in this work

The PI in this design has been shown in Fig 4.3. Two PI unit are included in this topology, and each unit has two input signal pairs and eight current source branches. With a polarity control, two pairs of input signal could share the same current source branch[18]. Only the current source and its switch are duplicated, which reduces a huge amount of MOSFETs. With a resolution of $\Delta t = 1.3125ns(5.625°)$, 7 extra levels should be interpolated into two input clocks. When the control bits of the I_ctrl switches are set to either all low or all high, the output phase of PI is the same as CLK_fast or CLK_slow, which makes the total phase levels of PI output to be 9. To represent the 9 phase levels, 8 current sources are required for each PI unit. The 9 phase levels and the control bits are illustrated in Table 4.1. However, the last phase level with all bits set to high does not need to be conducted in the real operation. Because when the input clocks are shifted to the next state ($0°$ and $45°$ to $45°$ and $90°$), the control bit 0000 0000 represents the output phase the same as 1111 1111 in the previous state. Thus, in the control of PI, the Ictrl<7> could always be set to low and only Ictrl<0> to Ictrl<6> is generated

from the thermometer decoder. This setting could reduce the register bits for thermometer decoder and therefore increase the data loading speed for the whole system.

Table 4.1 Control bits of current sources and related phase variation

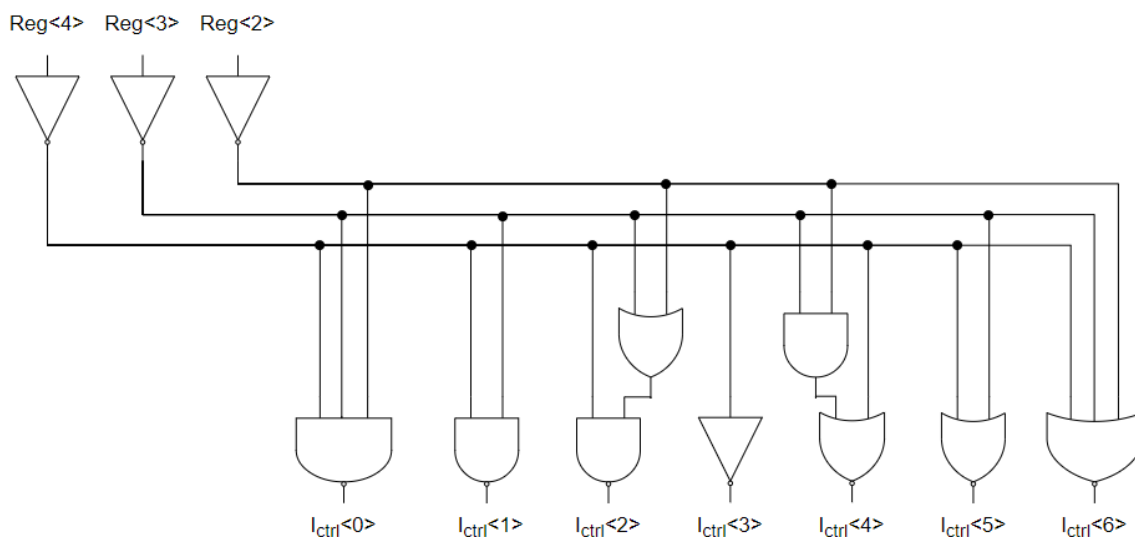| Ictrl<7>	Ictrl<6>	Ictrl<5>	Ictrl<4>	Ictrl<3>	Ictrl<2>	Ictrl<1>	Ictrl<0>	Phase variation(degree)
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	1	5.626
0	0	0	0	0	0	1	1	11.25
0	0	0	0	0	1	1	1	16.375
0	0	0	0	1	1	1	1	22.5
0	0	0	1	1	1	1	1	28.125
0	0	1	1	1	1	1	1	33.75
0	1	1	1	1	1	1	1	39.375
1	1	1	1	1	1	1	1	45

4.3 Thermometer Decoder

Figure 4.4 3-to-7 Thermometer decoder

Because the output levels are reduced, 3 bits are adequate for the decoder to generate all the output control bits for the current source switches. For each channel, the delay time is determined in the computer and will be sent to the register. The control bits for Ictrl<0:6> will not change during the operation unless the focal spot is changed and require a new delay time setting. Thus the decoder does not transfer any error or switching behaviour to the PI. The structure of the thermometer decoder is shown in Fig 4.4. All the digital gates are customized from the transistor level, so the topology, power and area might be further improved if the Verilog synthesis approach is conducted.

4.4 Performance

The simulation for all the output phase from 0° to 360° is shown in Fig 4.5(a)&(b). In Fig 4.5(c) is the output curve before adding the inverter. Due to the large output capacitor, the swing of the output voltage is reduced to a range from 0.3V to 1.6V. However, the linearity is set to best at a level of 0.9V where is the measurement point for phase delay. The switching decision point of the inverter is also set around this value thus the final output has good linearity. Besides, since the PI is a differential structure, the negative output could be directly used as a delay range from 180° to 360°, which is the same for the continuous wave in the stimulation mode. However, the imaging function might require a different delay setting so the improvement would require future verification in the data and clock control design.
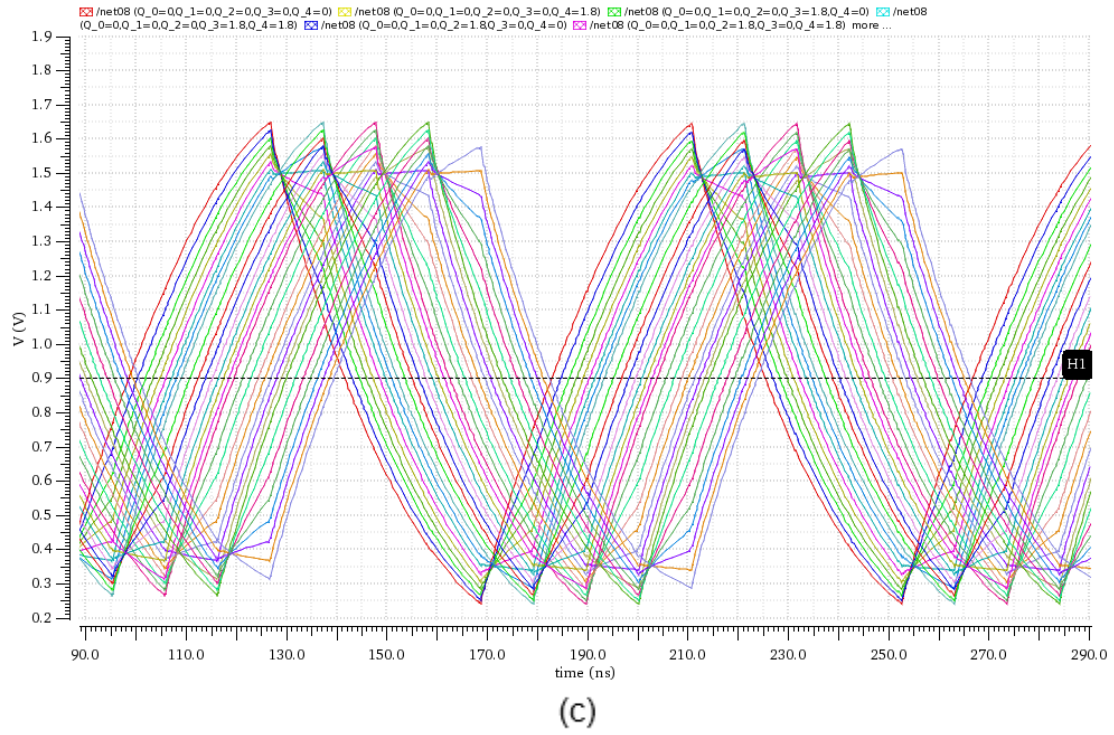

(a)

(b)

(c)
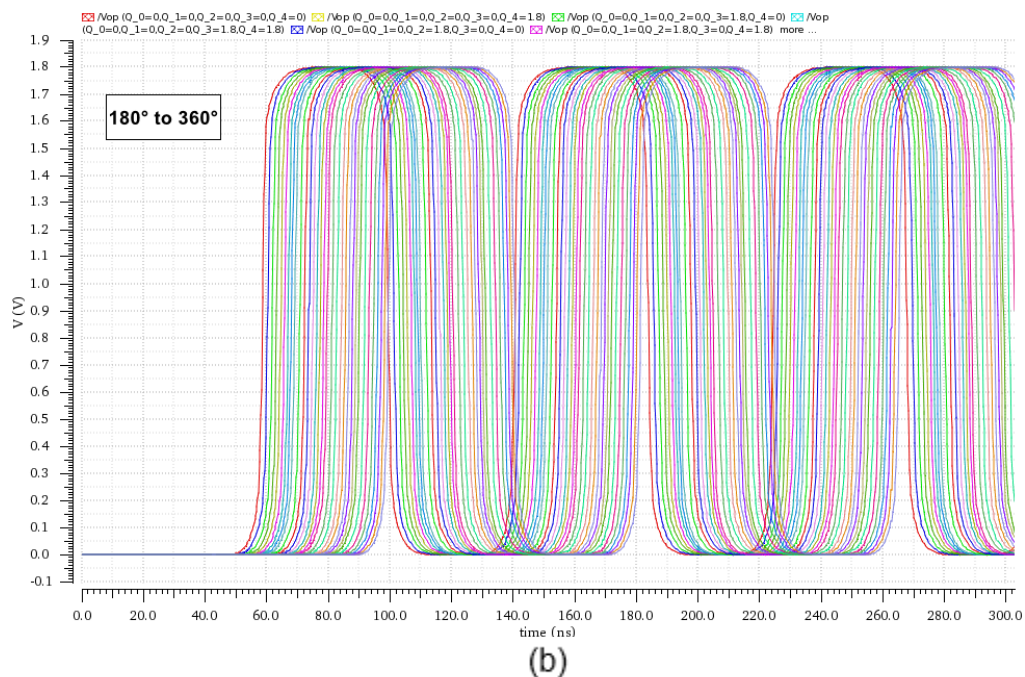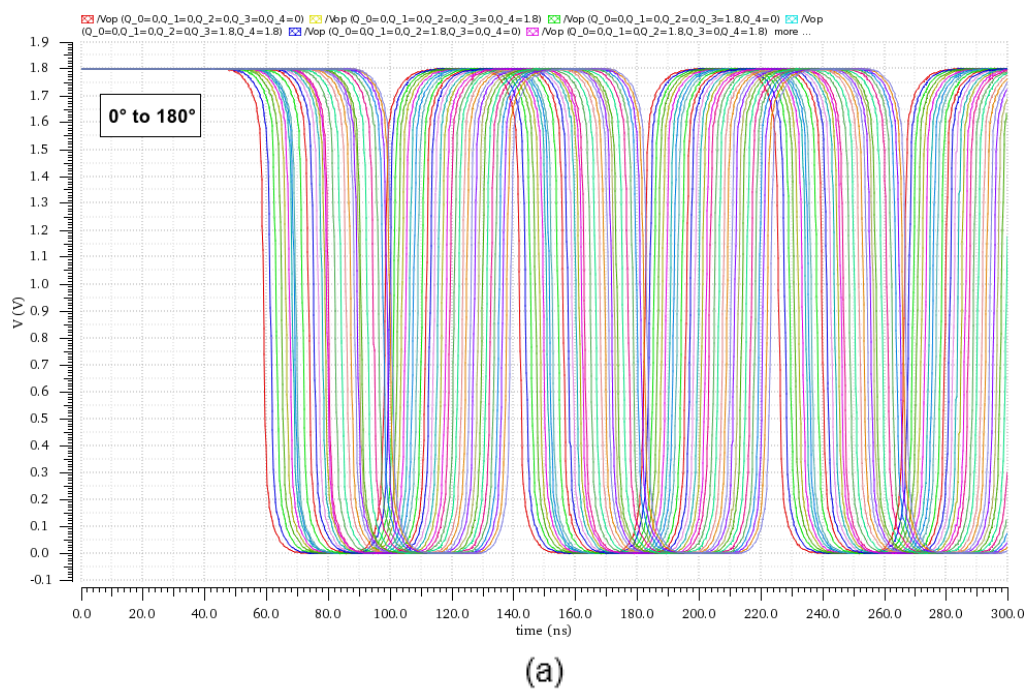
Figure 4.5 Simulation of PI output voltage
(a) Output voltage of PI from 0° to 180° (b) Output voltage of PI from 180° to 360°
(c) Output voltage of PI before the inverter
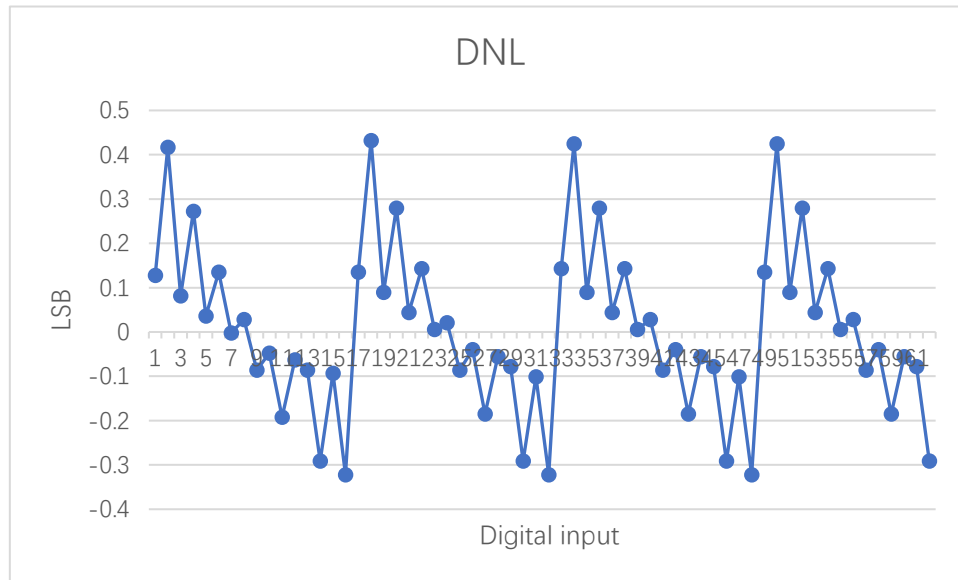
Figure 4.6 DNL of the PI

Both rising edges and falling edges are measured to evaluate the linearity of the PI. The differential nonlinearity(DNL) is shown in Fig 4.6. The periodicity in the DNL is due to the phase levels are repeated when the input signals are switched to the next state. The power consumption of the fine delay generation block is shown below. The biasing of the PI could be extracted to the outside-channel area as a sharing circuit for all the channels.

Table 4.2 Power consumption of the fine delay generation block

	Current	Power
Phase interpolator	150.4uA	0.27mW
Thermometer decoder	157pA	0.28nW
MUXs	43nA	77.4nW
Biasing	95.47uA	0.17mW
Inside channel	150.4uA	0.27mW

Chapter 5 Pulse wave mode function Design

Since the position of the vagus nerve is different from individuals to individuals, and it might move a bit from the original location during the stimulation, a real-time ultrasound imaging function is required. In many works[13] [28] [29], the PW mode function is achieved by digital devices considering the small area and power consumption. This design also uses a fully digital block, but the final output is based on its embedded stimulation function and without using a large counter or comparator.

The general way of generating digital PW mode function will be introduced first. Then the concept of using the continuous wave to trigger output is illustrated, followed by the double DFF architecture of the digital imaging block in this design. Each device inside this digital block and its simulation results will be demonstrated at last.

Since all the devices in this chapter are digital, the MUXs and the register inside the channel are also included in this chapter.

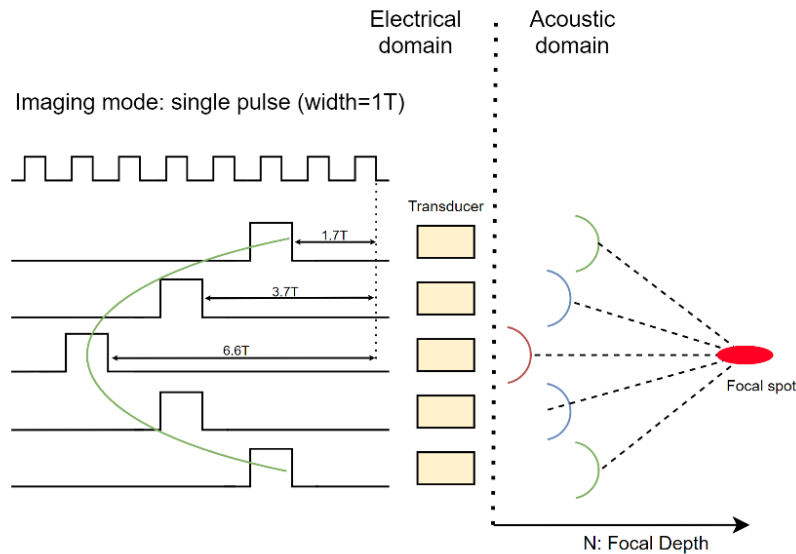## 5.1 TX pulse wave mode using digital circuit

Figure 5.1 Pulsed-wave for ultrasound imaging

As shown in Fig 5.1, for focused ultrasound imaging, a pulsed-wave needs a long-range of delays with precise resolution. In each channel, the pulse-wave delay time should be exactly in the form of $a \cdot T + b \cdot \Delta t$ where $a, b = 0,1,2 \ldots$. To achieve this, a high-frequency clock and large-bit counter and comparator are normally used, such as in Fig 5.2. An outside-channel counter is counting the delay and the trigger time is determined by the 10-bit comparator. The delay time resolution is determined by the counter frequency.

For small size transmitter array and short channel delay time, it is convenient to use this topology. However, when the number of elements increases, the delay time for channels at the

edge of a 2D phased-array increases rapidly, which requires large digital devices. The clock frequency is also increased if the resolution needs to be improved (1GHz for 1ns resolution). To support these digital devices, the size of the required registers will also be increased.

In our proposed system, the delay resolution is $\Delta t = 1.3125ns$, and the maximum delay among all the channels is $delay_{max} = 22 \cdot T - 7 \cdot \Delta t$. Since the counter is working in the binary system, to achieve the desired resolution and the maximum delay, a $760MHz$ clock and an 11-bit counter under this frequency are required if using the traditional digital counter scheme. This high-frequency clock should be added externally and it will increase the switching power dramatically. The large counter also squeezes the channel area for other components. Both of them are not suitable for a wearable device.

An alternative option instead of the high-frequency clock and large counter is inspired by the fine delay block. In chapter 4, a continuous wave with the required delay resolution is already generated from the phase interpolator, which could be used to trigger the pulsed-wave output. The main issue is the control of triggering the output to a long-range of delays.
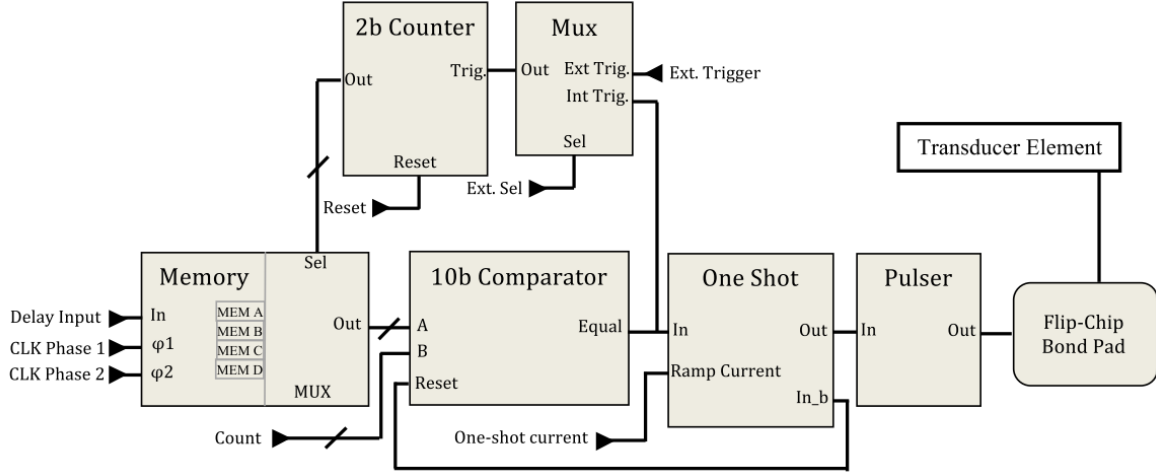

Figure 5.2 Digital Tx circuit for ultrasound imaging[13]

5.2 Continuous wave triggered output

5.2.1 Direct approach

As shown in Fig 5.3, the counter is working at the ultrasound frequency($12MHz$) with a counting unit equals to one time period($T = 84ns$). Compared to the counter working on $760MHz$, this counter only has to count 2^5 times to cover the maximum delay time $delay_{max} = 22 \cdot T - 7 \cdot \Delta t$. This reduces the size of the counter and comparator to 5 bit. After the comparator giving a high output when the counting number matches the pre-set delay value, the PI output with proper delay resolution will trigger the final pulsed-wave output. Thus the delay $a \cdot T + b \cdot \Delta t$ is achieved separately by the counter and PI. The EN_N signal for the counter is utilized to freely control the time when users need to conduct the imaging function. The EN signal for the comparator is the inverted signal of EN_N and it can avoid the always-high output when the register value is 0 and the counter is not activated.
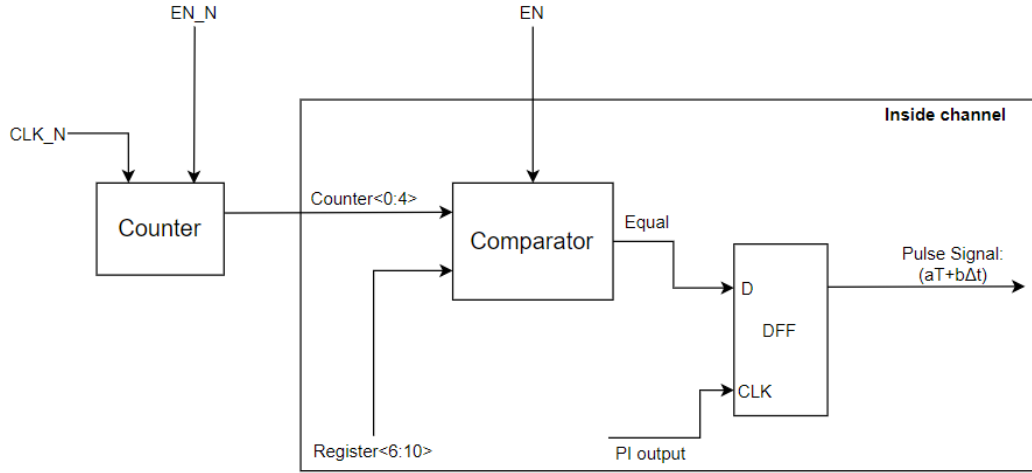
Figure 5.3 PI output trigger Pulsed-wave

But this approach requires extremely accurate control of the Equal signal and PI output signal. As shown in Fig 5.4, the two red arrows indicate when two types of PI output signal have to be to properly trigger the final output. The black wave is the PI output with a delay of zero which is the minimum delay for the phase interpolator. The red wave is the PI output with a delay of 2π-LSB which is the maximum delay. Two colors of the final output illustrate the proper triggering wave for these two PI output signals.
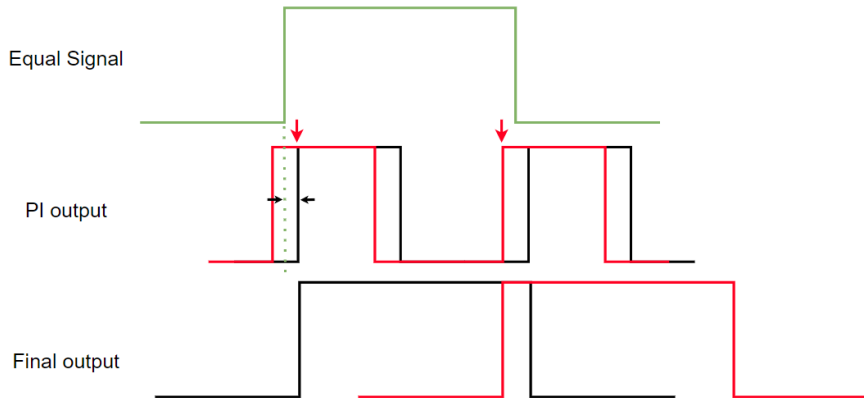

Figure 5.4 Triggering time for PI output and Equal signal

Due to the working principle of a rising-edge triggered DFF, to have the proper Final output signal, the two transition edges of the Equal signal should be ahead of the zero-delayed PI output while later than the 2π-LSB delayed PI output. The error range is indicated using black arrows, which is $\pm$LSB/2. Considering the delay error accumulated from the counter and comparator, and the different path which the PI output passes through, this approach requires an accurate delay control unit to arrange the zero-delayed PI output and the Equal signal inside the error range, which is hard to implement.
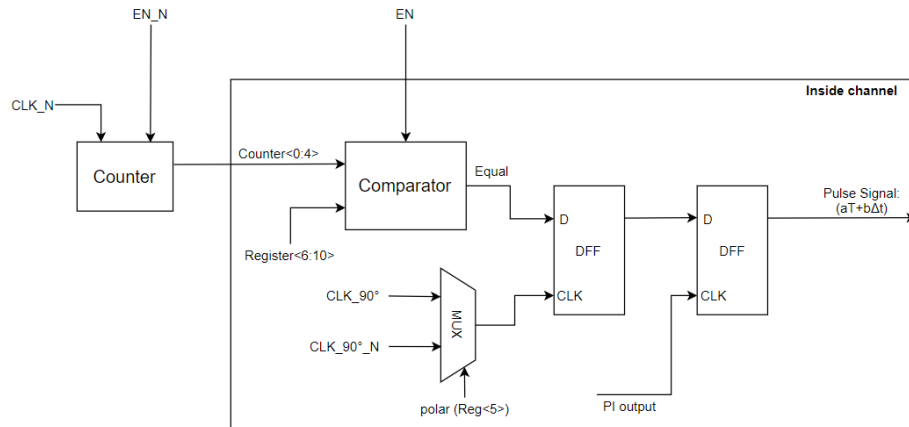
5.2.2 Double DFF approach

Figure 5.5 the double DFF approach

To release the strict requirement on transition edges of the Equal signal, an extra DFF is added, along with a MUX. This double DFF structure creates two different Equal signals from the original one. As shown in Fig 5.6, the timing diagram illustrates these two Equal signals. In the phase interpolator design, the reg<5> bit is used to determine the polarity of the PI output wave, which also indicates that the output phase range of the phase interpolator is from $0°$ to $180°$ or from $180°$ to $360°$. Depending on the polarity, the CLK_$90°$ from the DLL or CLK_ $90°$_N will generate a one-period pulse from the first DFF, which are the Equal with $90°$ and the Equal with $90°$_N. The PI output from 0 to π will trigger the second DFF under the Equal with $90°$_N, while the PI output from π to 2π will trigger under the Equal with $90°$. Now the PI output triggering edge is in the middle area of the Equal signal instead of the edge, which increases the error range from $\pm$LSB/2 to T/4. This approach releases the design requirements on the counter and the comparator. The error on the transition edges does not transfer to the final output.
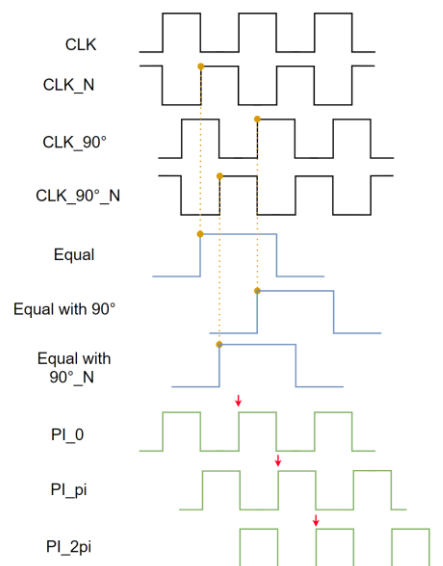

Figure 5.6 Timing diagram of the double DFF approach

5.3 Counter design

Since the error from the counter does not transfer to the final pulsed-wave output, the counter could use a simple asynchronous topology. To save the area, all DFF in the counter uses a TSPC structure. The whole counter structure is in Fig 5.7.
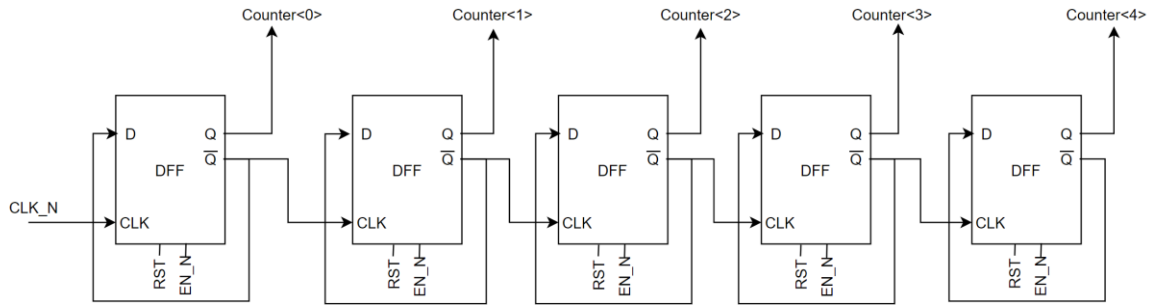

Figure 5.7 The 5-bit asynchronous counter

From Fig 5.8, the logic function of this counter could work properly under every corner and temperature $0°$ and $50°$. Although the transition edges accumulate delay from the first DFF to the last, and these delay values changes according to PVT variations, this delay error will not affect the final output.

A 5-bit counter could count from 0 to 31, but due to the control signal EN_N, the first output code after the counter is enabled is 00001. To have a full counting range, the output code 00001 is recognized as number 0 for the comparator, while the code 00000 which will appear after the code 11111, is recognized as number 31. This adjustment is for the control signal setting in the future application.

(a)

60
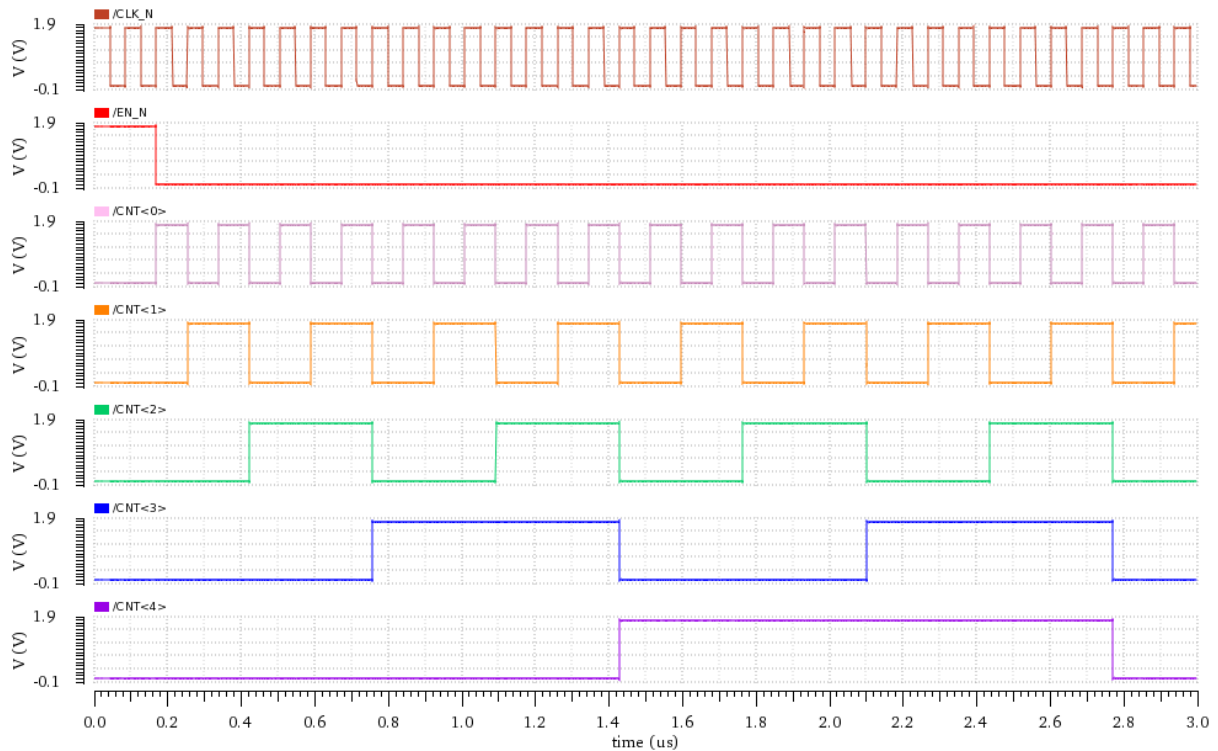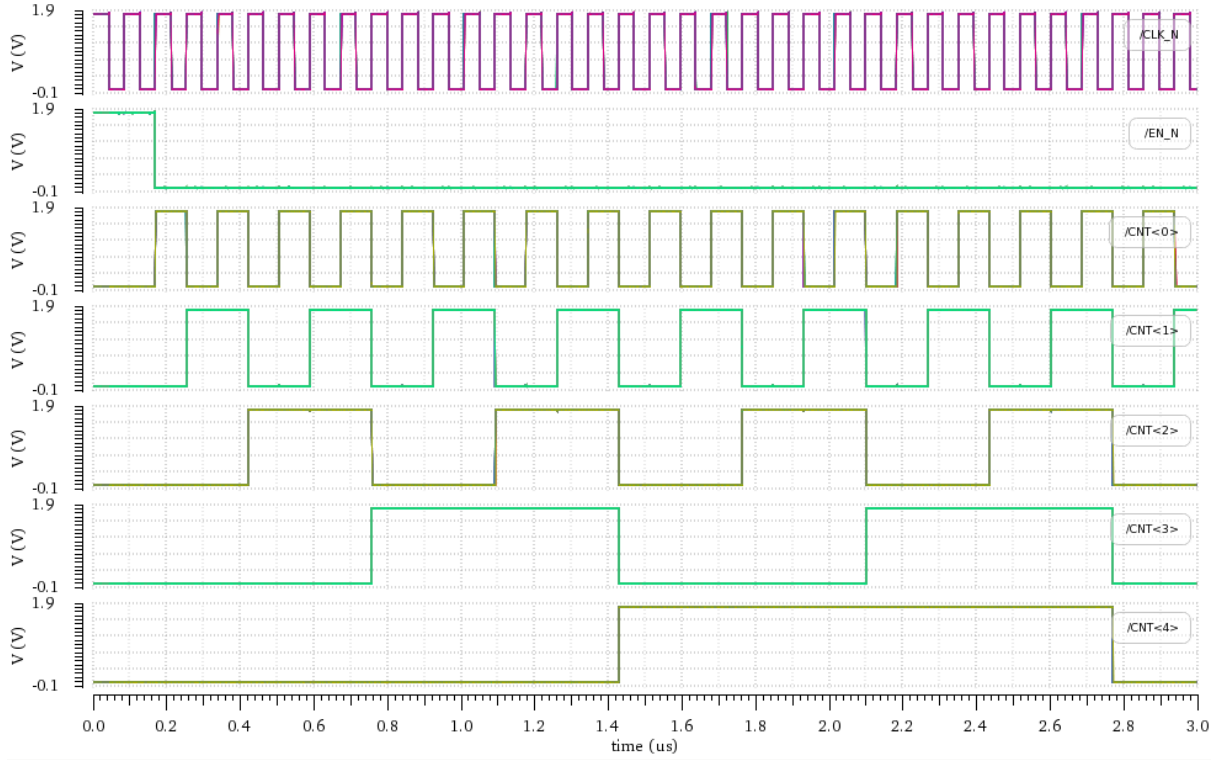
(b)

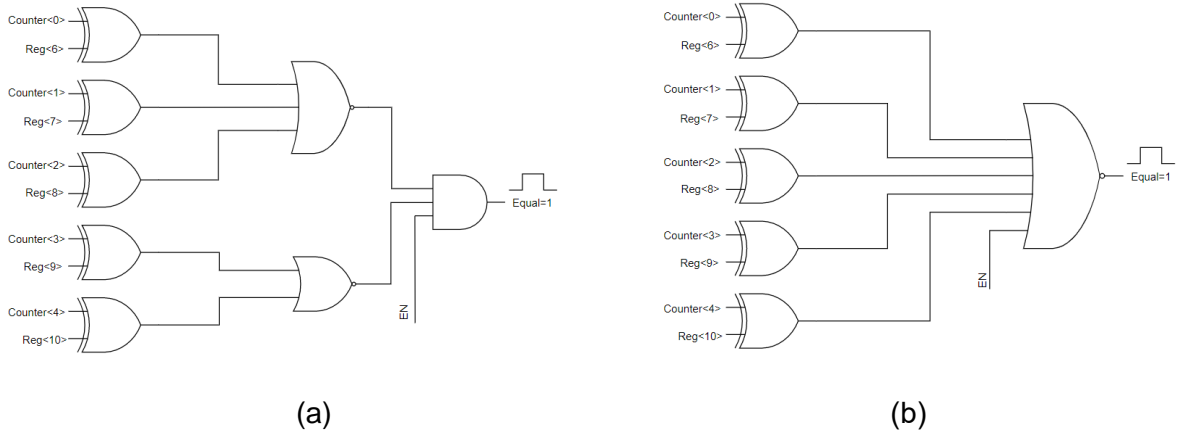*Figure 5.8 The transient simulation of the counter*
(a)TT corner at $27°C$ (b)All corners at $0°C$ and $50°C$

5.4 Comparator design

(a)   (b)

Figure 5.9 Comparator topology (a)Three stages (b)large fan-in

The comparator is set as the same bit size as the counter. The output of this comparator only becomes high when all the counter output bits are equal to the register bits. Compared to the dynamic comparator in [29], although the dynamic logic reduces many areas, the output waveform is always high after the matching point is reached, which is not desired for our double DFF topology. Normally, a clock signal is needed for the dynamic logic and it will increase the design complexity to get the output waveform incorporating with the PI output signal we used. Thus for the comparator, a static logic is preferred.

61

In Fig 5.9, two static logic topologies are shown. The large fan-in topology uses less number of transistors, but this 6-input NOR gate suffers from a much longer delay. Pseudo-NMOS technique could be applied but the static current increases and thus the power. If using differential cascade voltage switch logic (DCVSL) to solve the static current problem, the number of transistors is doubled, and the two compensated arrangement of transistors makes the design more complicated. The three-stages topology is much preferred here.

To reduce the area consumption in the three-stages topology, the five XOR gates are designed using the transmission gate. The architecture is shown in Fig 5.10. Considering the extra inverter needed for complementary inputs, using the transmission gate still reduces 4 transistors for each XOR(a normal CMOS logic XOR contains 12 transistors), which is 20 transistors in total for the whole comparator.

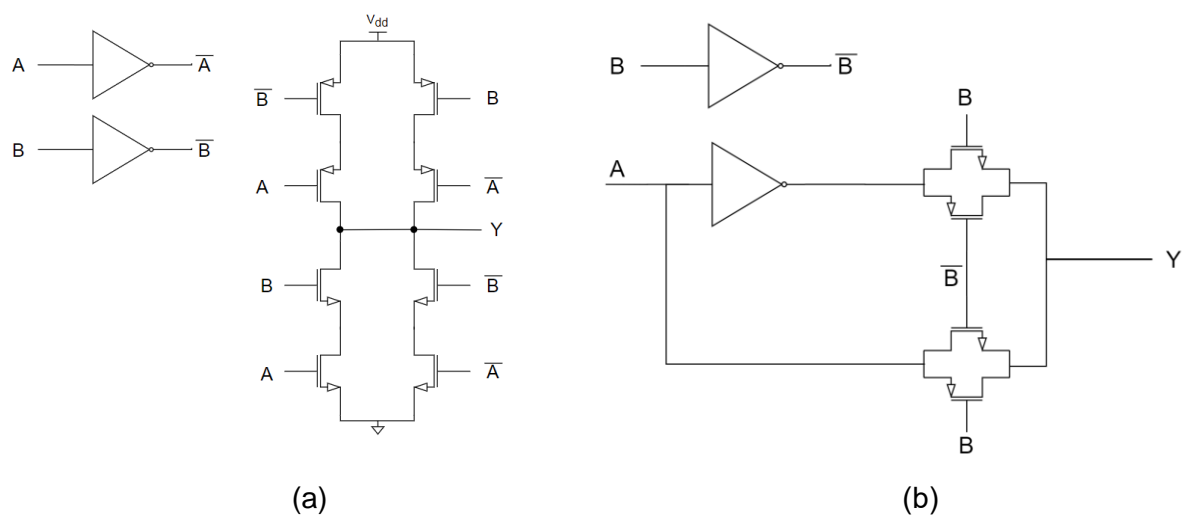(a)   (b)

Figure 5.10 XOR gate (a)CMOS logic (b)Transmission gate

Considering the counter is asynchronous, the output delay is accumulated from the least output bit Count<0> to the largest one, Count<4>. However, in the NOR gate inside the comparator, the PMOS which is the furthermost to the output suffers the largest delay. This PMOS could be utilized to compensate for the delay difference of the counter output. Hence, the Count<0> is set at the input port with the longest delay path and the Count<4> is set at the least one.

The comparator is tested with different input codes: 00001,00010,00100,01000,10000,00000. These codes representing 0,1,3,7,15 and 31 times of one clock period. Because each Count<i> is capable to trigger the comparator output to high, these codes are chosen to represent all different situations. The simulation results are in Fig 5.11. All these codes are tested in TT at $27°C$ and all the other corners(FF, FS, SF, SS) at $0°C$ and $50°C$.

Although some spikes might happen when the counter bits are switching, this will not cause an error output. Because the Equal signal will then pass through one DFF with either CLK_90 or CLK_90_N. The position of rising edges of these two clock signals is at least T/4 away from the spikes.

(a)

(b)

(c)
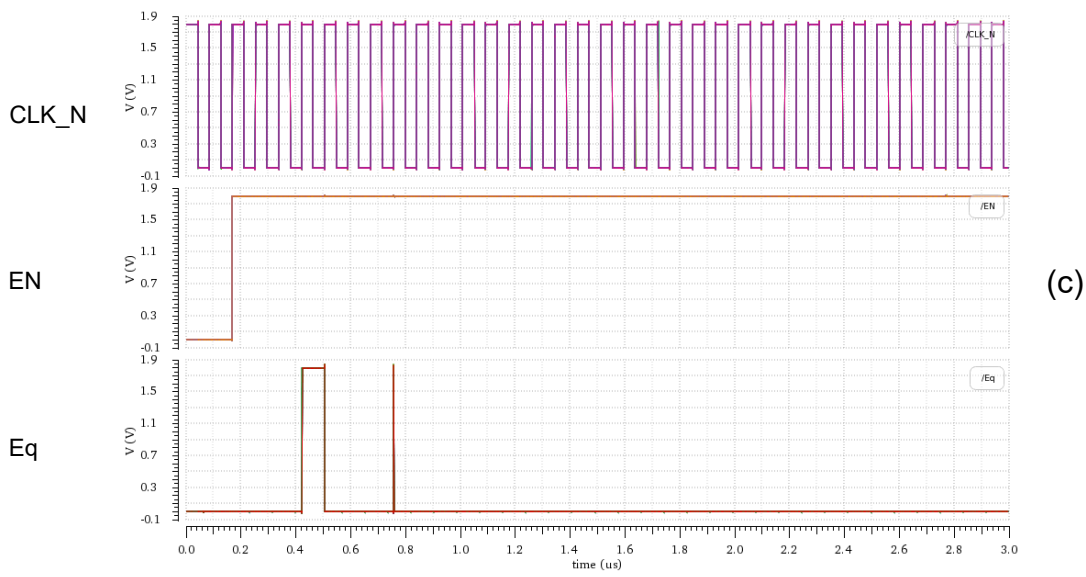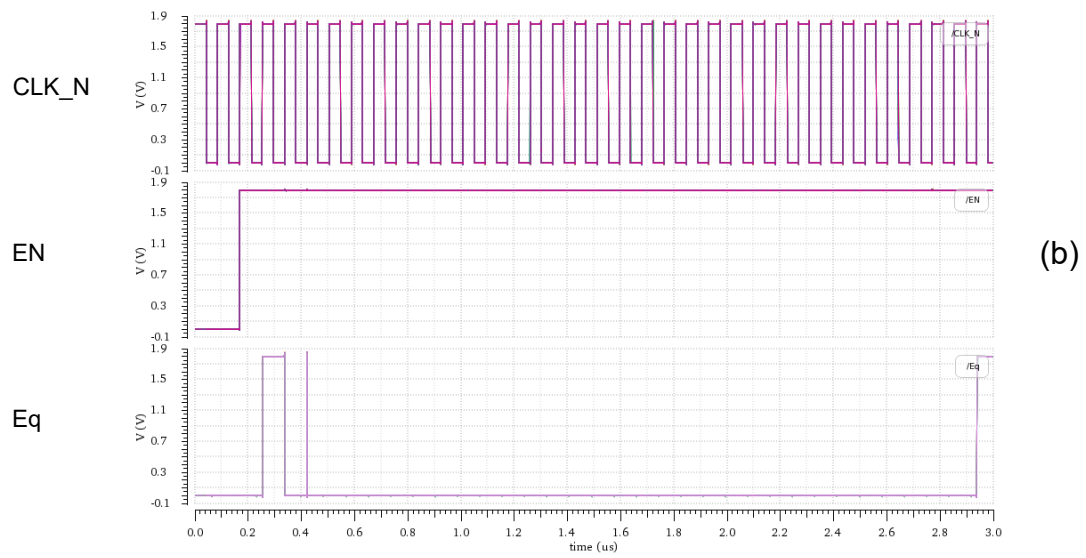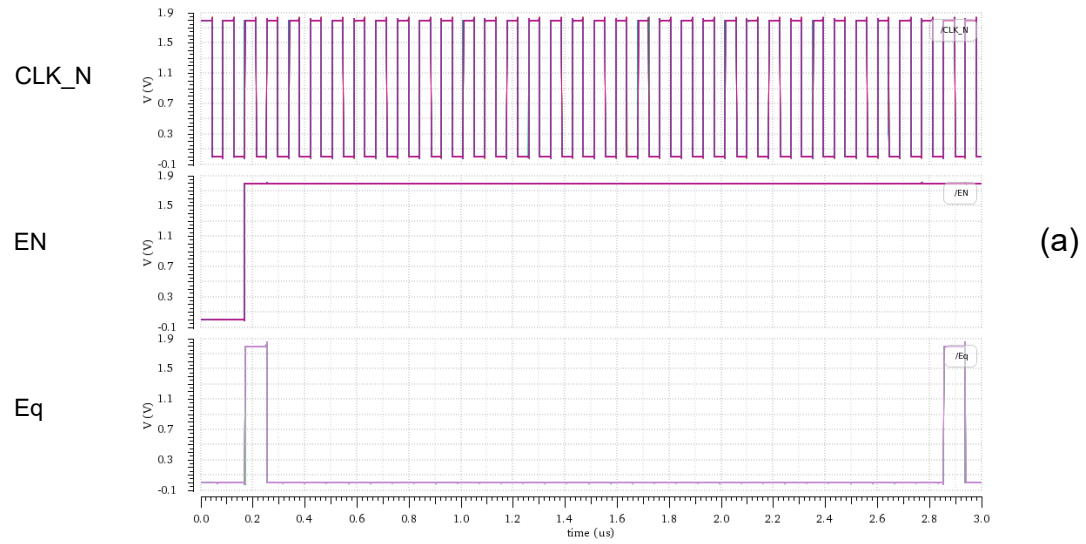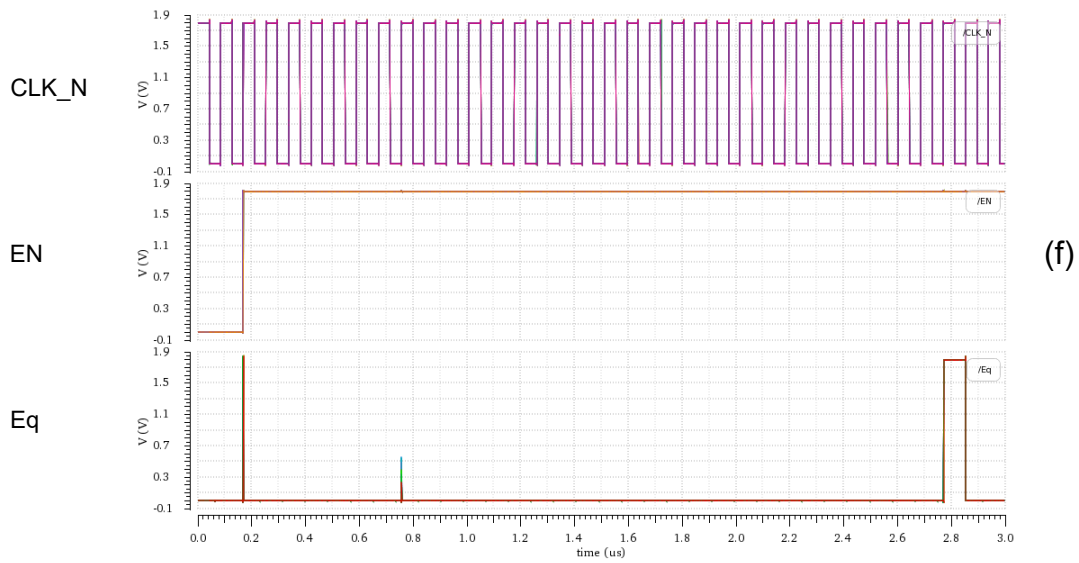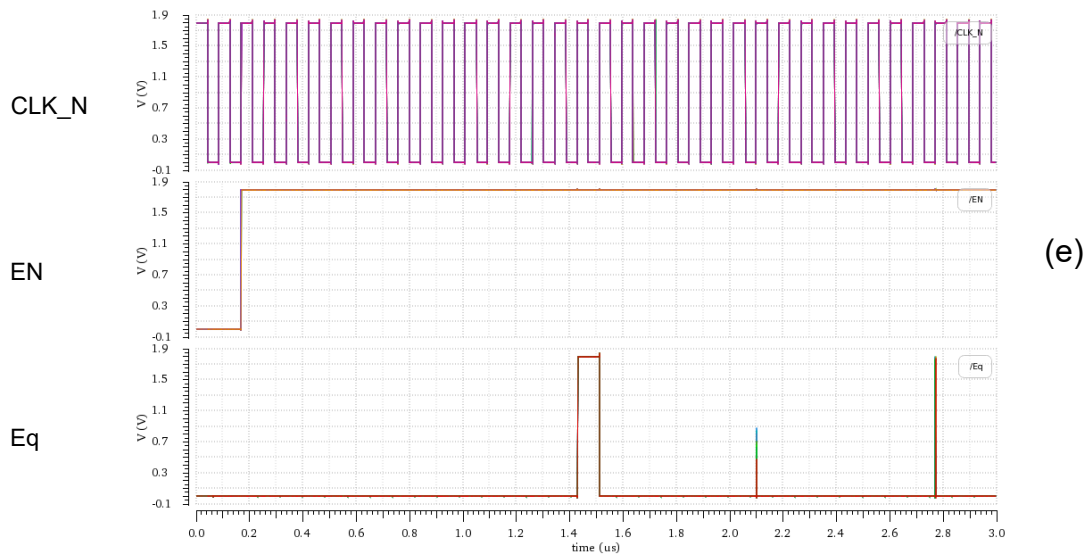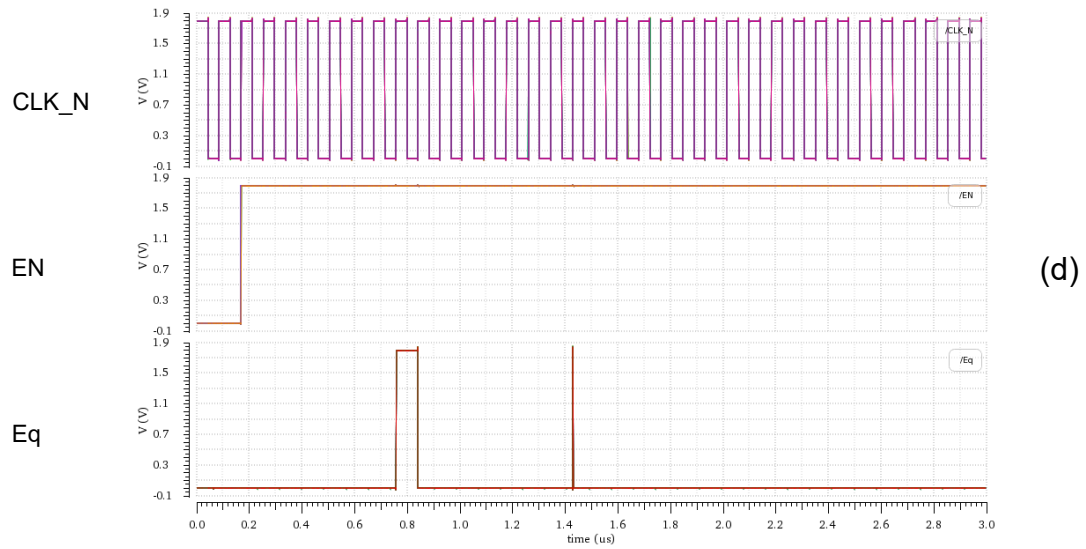
Figure 5.11 Comparator simulation results in different corners and temp
(a)00001 (b)00010 (c)00100 (d)01000 (e)10000 (f)00000

5.5 MUX design

There are two kinds of MUXs inside each channel, one is 2-to-1 and the other is 4-to-1. The 4-to-1 MUX is used to select the input clocks for the phase interpolator. The 2-to-1 MUX has been used in two positions, one is in Fig 5.5 for the clock signal in the double DFF structure. The other 2-to-1 MUX is in the transducer interface, where it selects either the continuous wave or the pulsed wave is sent to the transducer.

To save area, the MUX is designed using the transmission gate. 8 transistors are reduced for a 2-to-1 MUX. The 4-to-1 MUX is composed of three 2-to-1 MUX, thus 24 transistors are reduced.
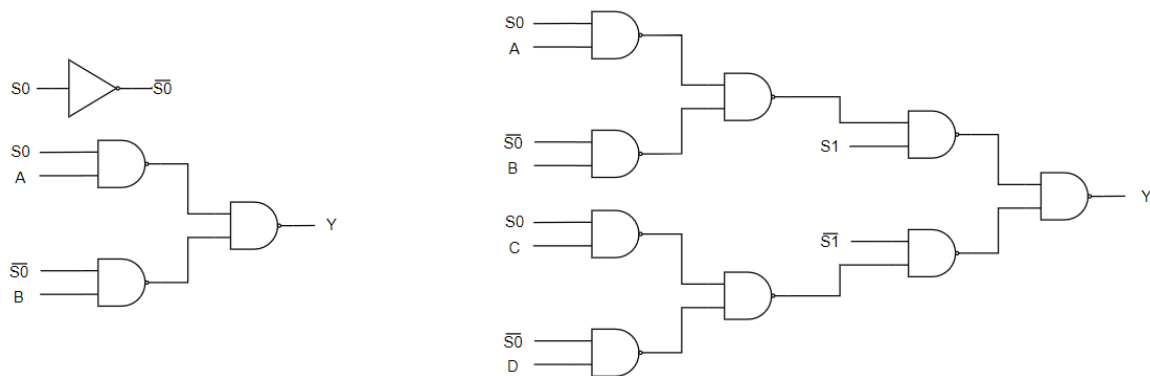
Figure 5.12 MUX using CMOS logic gate

Figure 5.13 MUX using Transmission gate

5.6 Register design

Depending on the operations, registers could be divided into four types: Serial-In Serial-Out (SISO), Serial-In Parallel-Out (SIPO), Parallel-In Serial-Out (PISO), and Parallel-In Parallel-Out (PIPO). But the register cell is similar for all these four types, normally is a DFF. In this section, the design of a transmission gate based register cell [29] is demonstrated, and two types of operation (SISO and PIPO) are shown to prove that the register cell could be applied in different registers.
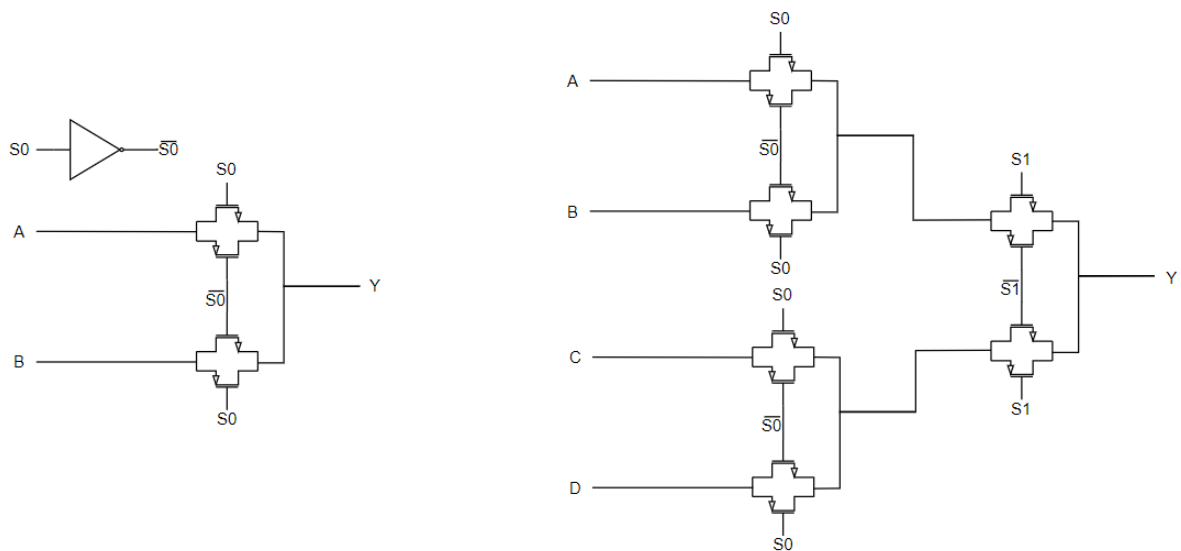
As shown in Fig 5.14, the register cell is composed of two CMOS transmission gates and three inverters. The CLK_1 controls the data loading, while the Load signal decides whether the output will be updated or not. In the PIPO mode, all register cells have an independent data input line so the loading time is extremely short. While in the SISO mode, each register cell is connected one by one, and only one data line is required, but the loading time is several times of the PIPO structure. For future applications, the operation mode will be decided by the control system. In this work, all the simulations using the register are based on the PIPO structure.
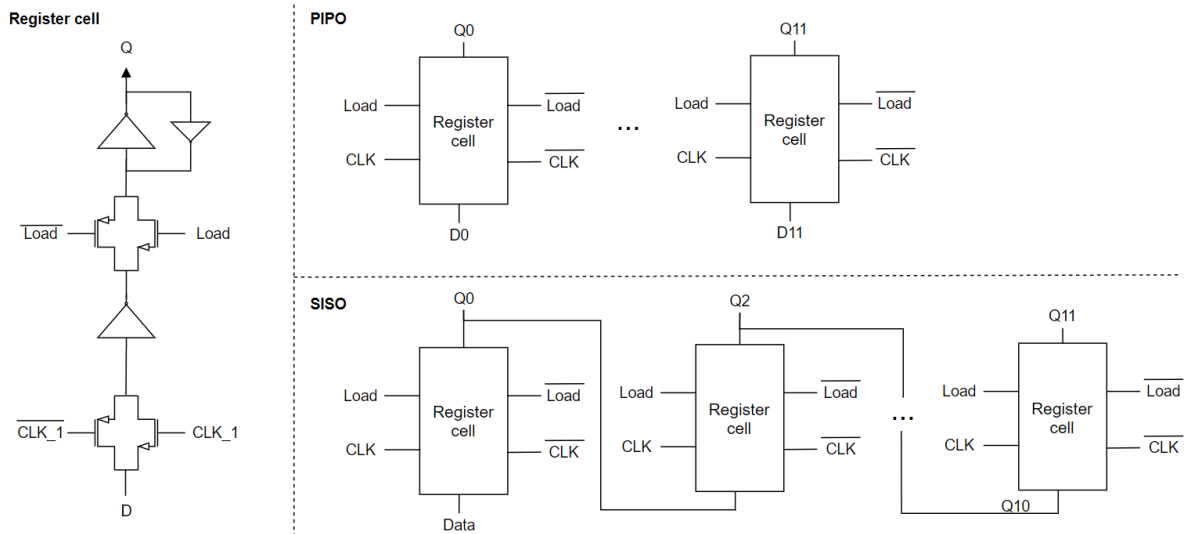
Figure 5.14 Register cell and two operation modes

In Fig 5.15, these two operation modes are simulated. In PIPO, the output is synchronized to the data when the Load signal is high, the final value of the output depends on the data value when the Load signal is turned off. In SISO, the data is load into each register one by one with the cycle of the clock signal. The 1001 data appears at the third register cell output after three clock periods.
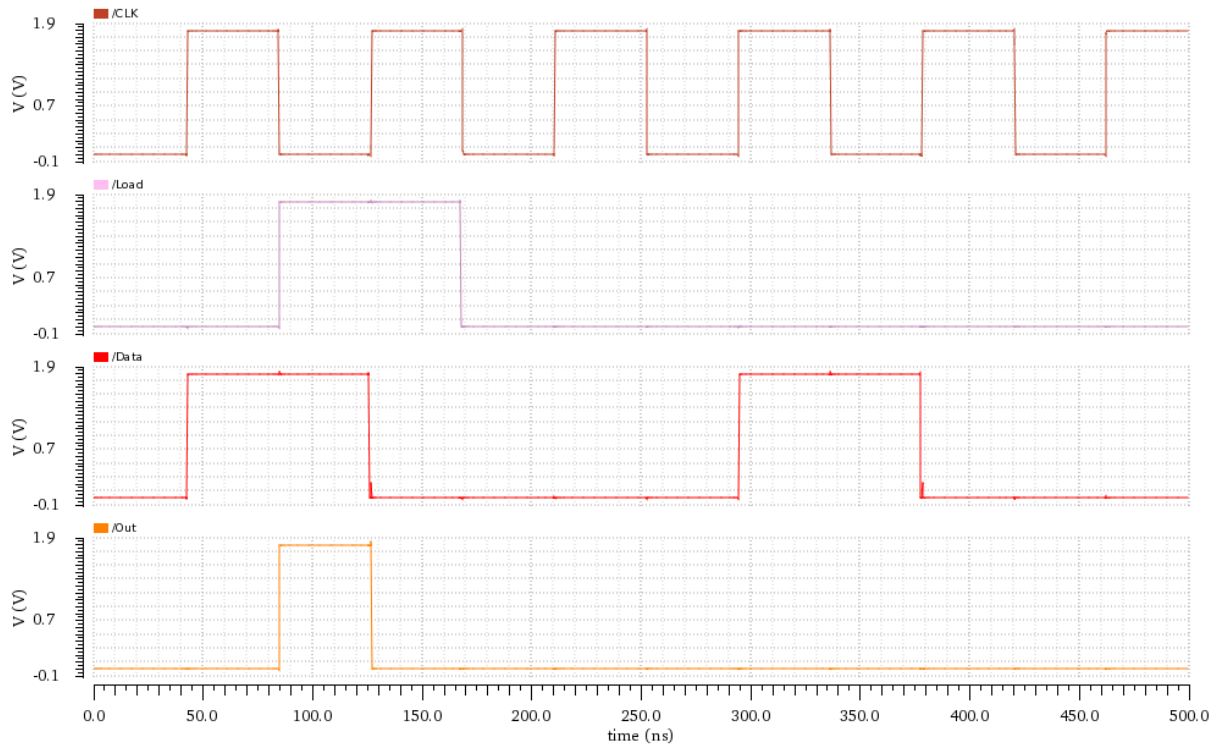
(a)
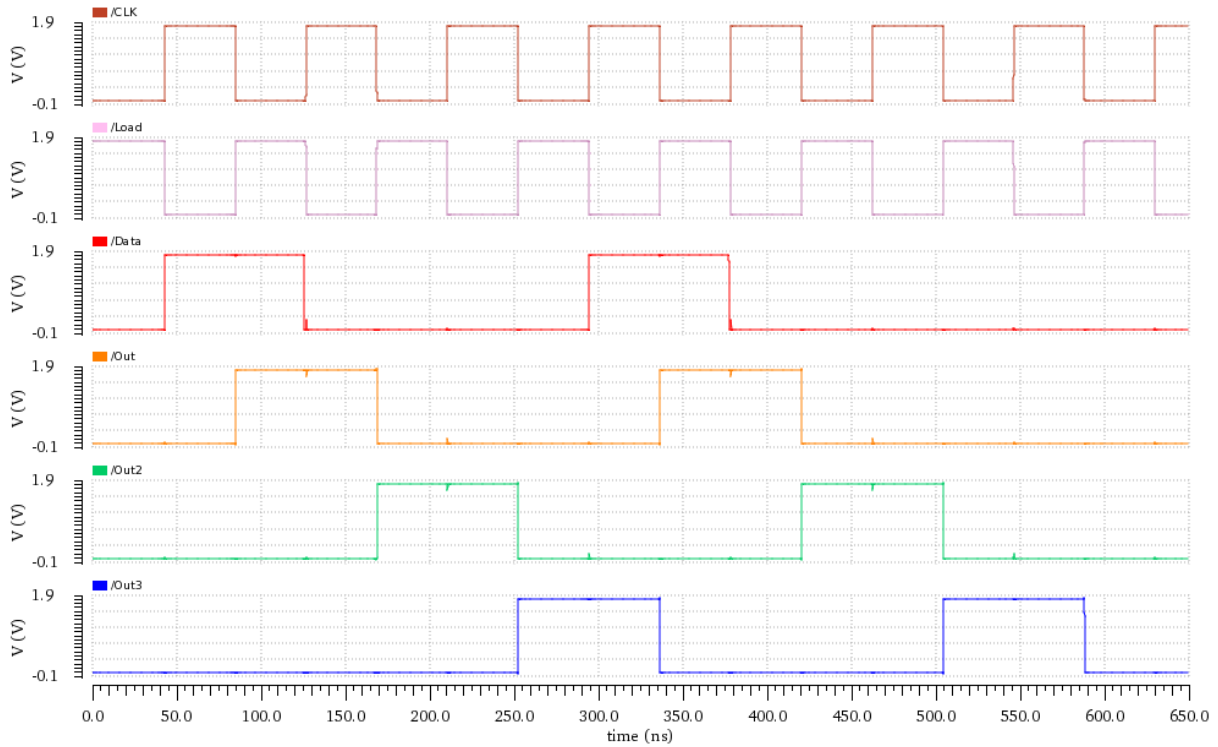
(b)

Figure 5.15 Two operation modes (a)PIPO (b)SISO

5.7 Performance

The simulation for the whole US imaging function of this work is presented in Fig 5.16. The delay value is randomly chosen: 00101(which represents 4 because the code for 0 is set as 00001) for the comparator and 20 times of LSB for the PI output. After then EN_N is off, the

67

counter starts counting and the Eq signal goes to high after four clock periods. Then the first DFF generates a new Eq signal according to the CLK_90_N. The PI_output triggers the second DFF to have the final output. The second DFF is the only device that affects the final output accuracy, and it is the same for all the channels. So the pulsed-wave error among different channels only depends on the PI_output which comes from the phase interpolator.
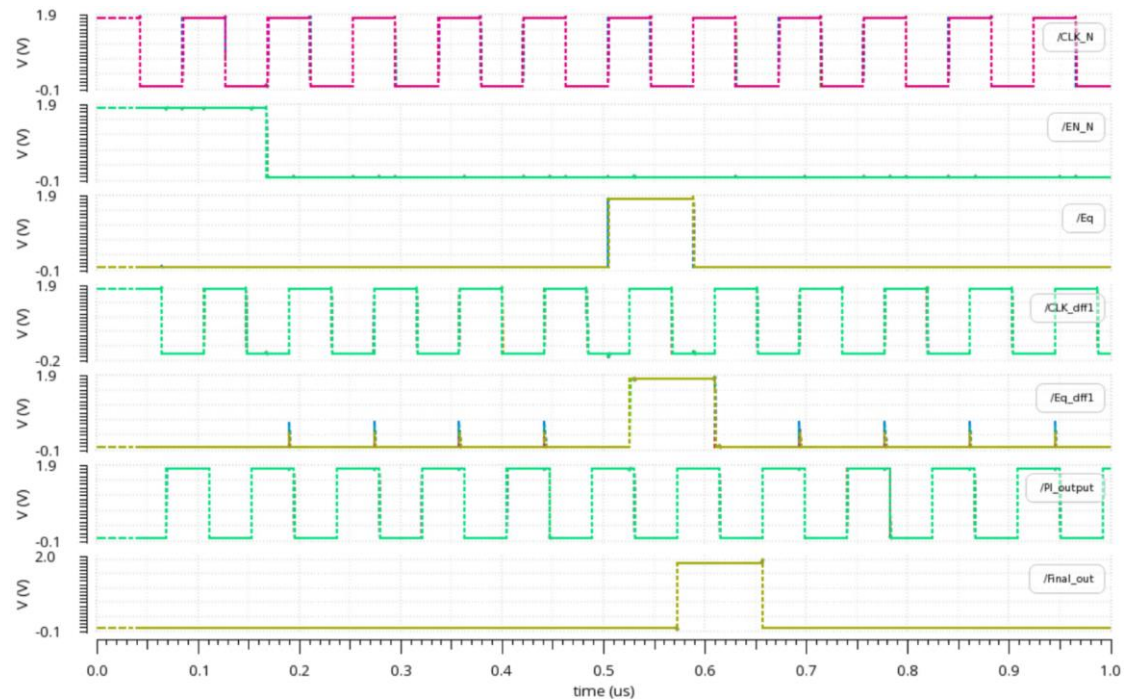
Figure 5.16 All corners simulation for the imaging function

Table 5.1 Power consumption of the devices in the pulsed-wave path

	Current	Power
Counter	389.4nA	0.7uW
Comparator	161.2nA	0.29uW
1st DFF	179.5nA	0.32uW
2nd DFF	176.6nA	0.32uW
MUX	57.33pA	10.3nW
Clock buffer	14.26nA	25.7nW
Register	73.46nA	0.13uW
Inside channel (Except Counter)	605.1nA	1.1uW
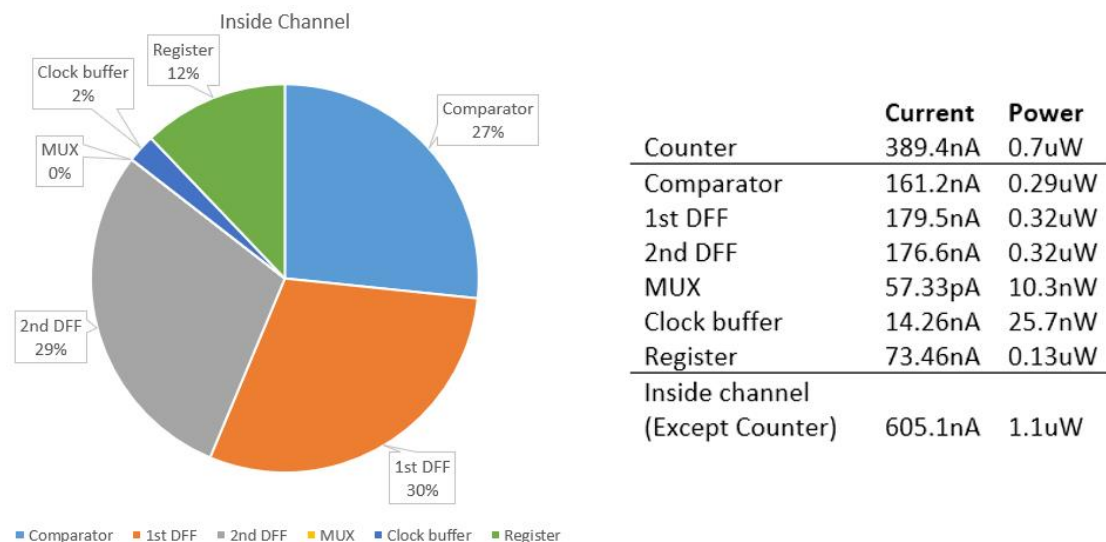
Table 5.1 illustrates the power consumption of all the devices in the pulsed wave path. Except for the counter, the other devices are inside the channel and are shown with a pie chart. From Fig 5.16, the DFF output has periodic spikes even its value is zero, so a two-inverter based clock buffer is added after the second DFF to have a clean final output.

Chapter 6 Transducer interface

Following the two modes of waveforms are generated, these low voltage (0-1.8V) wave has to be shifted to a higher voltage (30V according to the specification) to drive the transducer. This chapter introduces these circuits interfacing with the transducer. First, the devices included in the transducer interface block will be explained, followed by the design of the level shifter and HV driver.

6.1 Devices required in the transducer interface
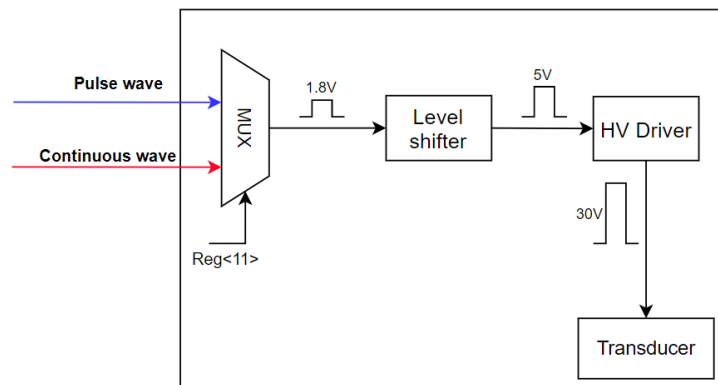
Figure 6.1 block diagram of the transducer interface

In Fig 6.1, the high voltage devices we used is 36V MOSFET in the $TSMC\ 0.18\mu m\ HVBCD$ library, which requires a 5V gate-source voltage. Thus the input wave has to be first shifted to the 5V domain by a level shifter. Since the 36V MOSFET occupy a large area, the design should consider using as few transistors as possible. Fig 6.2 illustrates the area of a minimum-sized 36V NMOS occupies compared to 5V and 1.8V NMOS.
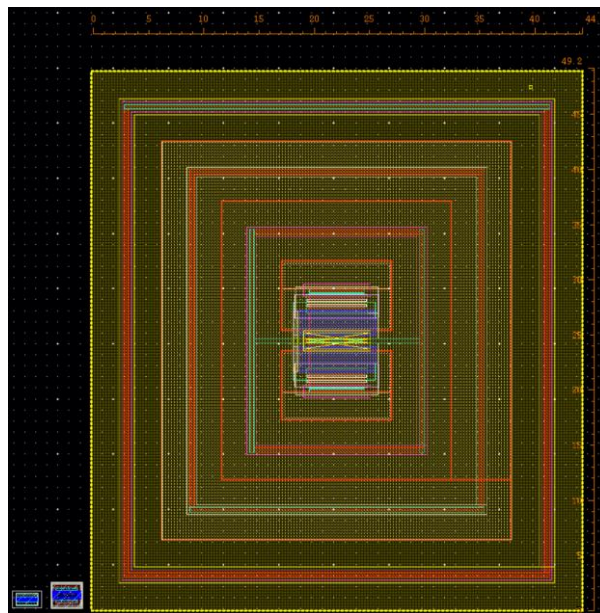
Figure 6.2 1.8V, 5V, and 36V NMOS

To match with the design of the receiving circuit, the transducer is chosen to use the same

material, Pb(Mg$_{1/3}$Nb$_{2/3}$)O$_3$-PbTiO$_3$ (PMN-PT). Compared to lead zirconate titanate(PZT) used in previous work[18], PMN-PT is preferred due to its higher piezoelectric coefficients and coupling coefficient. Two commercial products of piezoelectric materials are listed in Table 6.1. The higher d_{33} and k_{33} means the PMN-PT could convert electrical energy to mechanical placement more efficiently than PZT.

Table 6.1 piezoelectric properties[30]

	symbol	PZT(Navy type II)	PMN-PT
Piezoelectric coefficient (Coulomb/Newton)	d_{33}	390×10^{-12}	1285×10^{-12}
Coupling coefficient	k_{33}	0.72	0.89
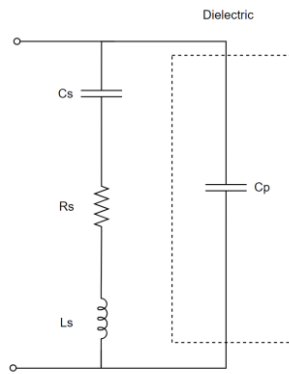
Figure 6.3 Electrical model for piezoelectric transducer[31]

The PMN-PT, like the other piezoelectric materials, could be modelled using a lumped-element model. This model is composed of two parts, one is an RLC branch representing the mechanical property of the material. The other one is a shunting capacitor which describes the dielectric property. Each element in this model could be calculated precisely using the equations below [32].

$$C_p = \frac{WL}{t} \varepsilon_{33}^T \left(1 - k_{31}^2\right) \tag{6.1}$$

$$R_s = \frac{|Z|_{\min}}{\sqrt{1 - \omega_s^2 C_p^2 |Z|_{\min}^2}} \tag{6.2}$$

$$L_s = \frac{\rho L t}{8\omega} \left(\frac{s_{11}^E}{d_{31}}\right)^2 \tag{6.3}$$

$$C_s = \frac{8L\omega d_{31}^2}{\pi^2 t s_{11}^E} \tag{6.4}$$

Here, W and L are the dimensions of the transducer, with t as its thickness. ε_{33}^T is the permittivity under constant stress, ε_{11}^E is the elastic compliance under a constant electric field. ω is resonance angular frequency, ρ is the mass density. d_{31} and k_{31} are piezoelectric coefficient and coupling coefficient, which would be d_{33} and k_{33} in our case. These values should be obtained from the impedance spectrum by experiments. But in the design, several assumptions are made for our simulation.

As for the transmitter design, the transducer model is recognized as a load impedance. Without a real product to test, a commercial product PMN-PT from [33] is taken for the calculation. First is the shunting capacitor. As discussed in chapter 2, the size of each transducer element is set to be $90\,\mathrm{um}\times90\,\mathrm{um}$. The operating/resonance frequency is $12\,\mathrm{MHz}$. With these specifications and parameters from the commercial product, the shunting capacitor could be calculated as $1.5\,\mathrm{pF}$. As for the RLC branch, it is dominated by the resistor and its impedance is normally larger than $30\,\mathrm{k\Omega}$ [32]. Finally, the load impedance is dominated by the shunting capacitor and the value is $1.5\,\mathrm{pF}$. These assumptions are made currently for circuit design. Further experiments will be conducted to get the exact parameters when the real product of the transducer material is ready to use.

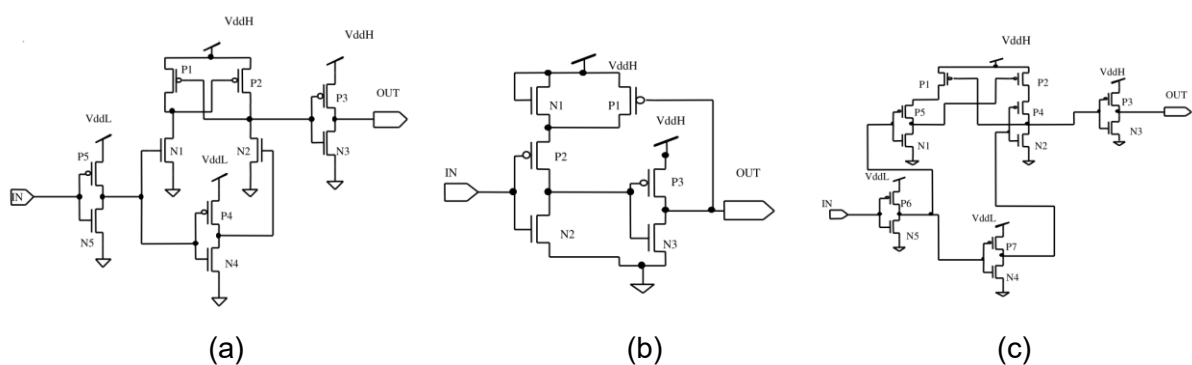## 6.2 Level shifter design

Figure 6.4 Three types of Level shifter[34]:
(a)Conventional (b)Single supply (c)Contention mitigated

Since the HV transistors require 5V input control, firstly the low voltage 1.8V signals need to be shifted to 5V. Three types of level shifters are shown in Fig 6.4. The first one is the conventional level shifter. The top two PMOS, P1 P2, forms a latch. When the input is high, N1 is off and N2 is on, which pulls downs the gate voltage of P1. Due to the positive feedback, the gate voltage of P2 is pulled up to VddH, and the gate voltage of P1 is pulled down to VSS, which gives a high output after the inverter. This structure requires two supply voltages and strong NMOS for N1 and N2, since they have to overcome the PMOS latch. Another disadvantage is the contention problem, which is the different current driving capabilities of NMOS and PMOS. Delay variations and more power consumption are caused by contention[35].

As for the single supply topology, it uses fewer transistors and only needs one supply voltage. When the input is low, P2 is open. With the diode-connected N1, the input port of the inverter is pulled up. The output goes to VSS and forces the half-latch P1 to further pull up the inverter input port to VddH. However, the biggest disadvantage of the single supply topology is leakage power. In our case, VddH is 5V and the VddL is 1.8V. When the input is high, P2 is expected to be off. But its threshold voltage is 1.2V, while the gate-source voltage of P2 is larger than 2V even the diode-connected N1 creates a lower "Vdd" for P2. So N1, P2, and N2 form a direct path from VddH to VSS while the input is high, consumes more static current. A simulation result more intuitively shows the static current driven by P2, in Fig 6.5.
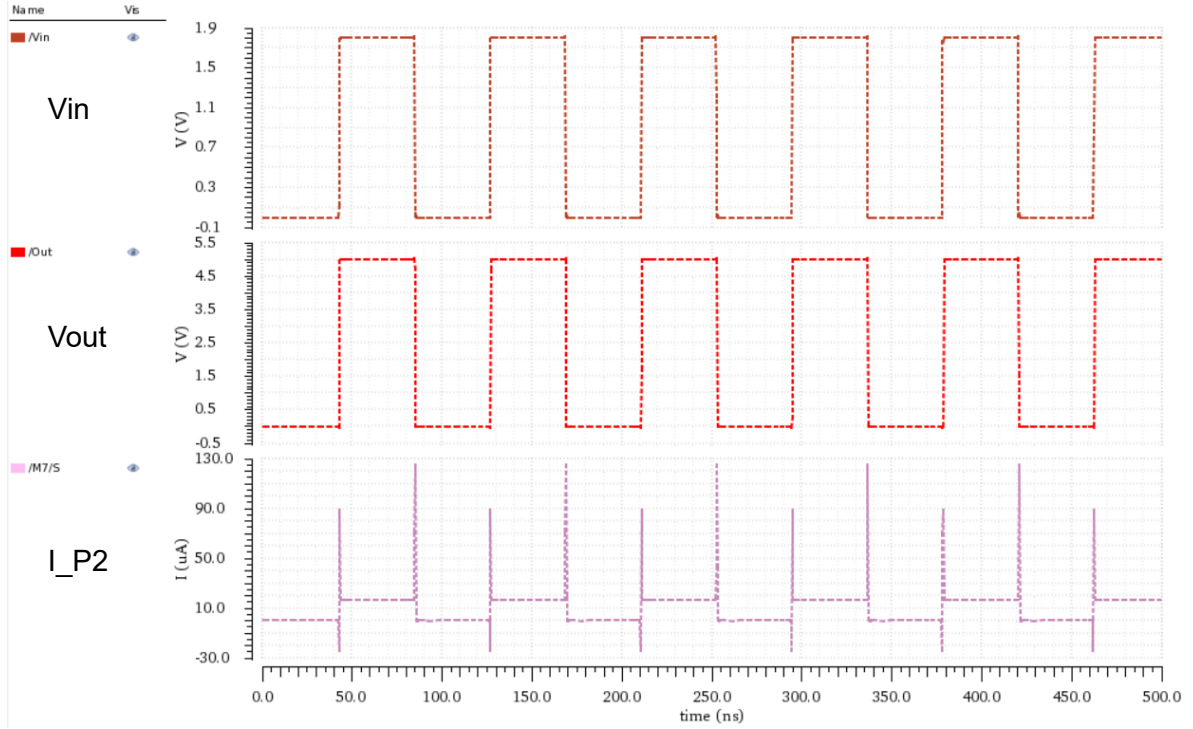
*Figure 6.5 P2 current curve in single supply topology*

Contention mitigated level shifter is modified based on the conventional level shifter. By introducing two extra PMOS above the N1 and N2, two quasi-inverters are formed and the logical values on the source terminals of NMOS are established faster than that in the conventional level shifter[36]. When the input of N1 is high, it pulls down the gate voltage of P2. While the P4 is opened by the inverted input signal, which along with P2 pulls up the source terminal of N2. Now the P1 is off and leaves no static current to P5 even the input of P5 is a 1.8V low voltage. The source terminals of P1 and P2 do not directly connect to the NMOS, thus the NMOS does not drive the terminal voltage from the high voltage source, which mitigates the large dimension requirements for NMOS. Besides, the inverter before the output node could act as a buffer or a 5V transducer driver if the transducer does not require a 30V high voltage. With these benefits, the contention mitigated level shifter is applied in this work.

6.3 High voltage driver

Normally an HV driver is also using a level shifter to convert the input signal from the low voltage domain to high voltage to have a better performance[28, 37]. In Fig 6.6(a), a pulse-triggered level shifter is introduced, but the left branch resistor always draws static current from the 60V supply. In Fig 6.6(b), a whole level shifter is under high voltage supply, which means six HV MOSFETs are required. This structure directly transfers the signal from the 1.8V domain to the 32V domain, which would use a 1.8V gate-source voltage controlled HV MOSFET. As for the Fig 6.6(c), it only requires two HV MOSFETs and could allow any arbitrary high pulse to pass through. However, an extra high voltage pulse generator is needed, which will occupy more area inside our channel.
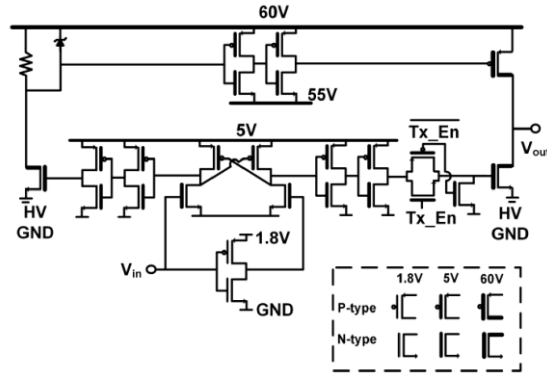
(a)
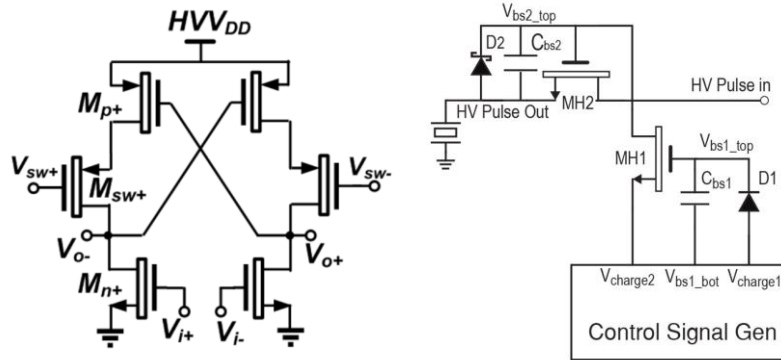
(b)   (c)

Figure 6.6 Three types of High voltage driver (a)[28] (b)[37] (c)[38]

Recalls the larger area occupied by one HV NMOS, from a conservative perspective, it would be more implementable using only two HV MOSFETs. But the control signal on a floating voltage for the high side PMOS is a problem. AC coupling could help shift a 0-5V signal to a higher level. In Fig 6.7, with a diode that limits the gate voltage of MHV2, no extra 25V voltage is needed.
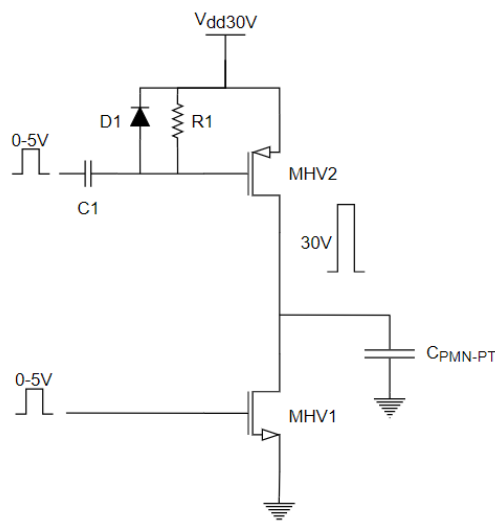
Figure 6.7 AC coupling level shifter

The diode D1 has a 5V break-down voltage. The resistor R1 and C1 form a high pass filter,

73

the cut-off frequency is below 1MHz, leaving a safe region for the input signal operating at 12MHz. The AC simulation of the RC filter is shown in Fig 6.8.
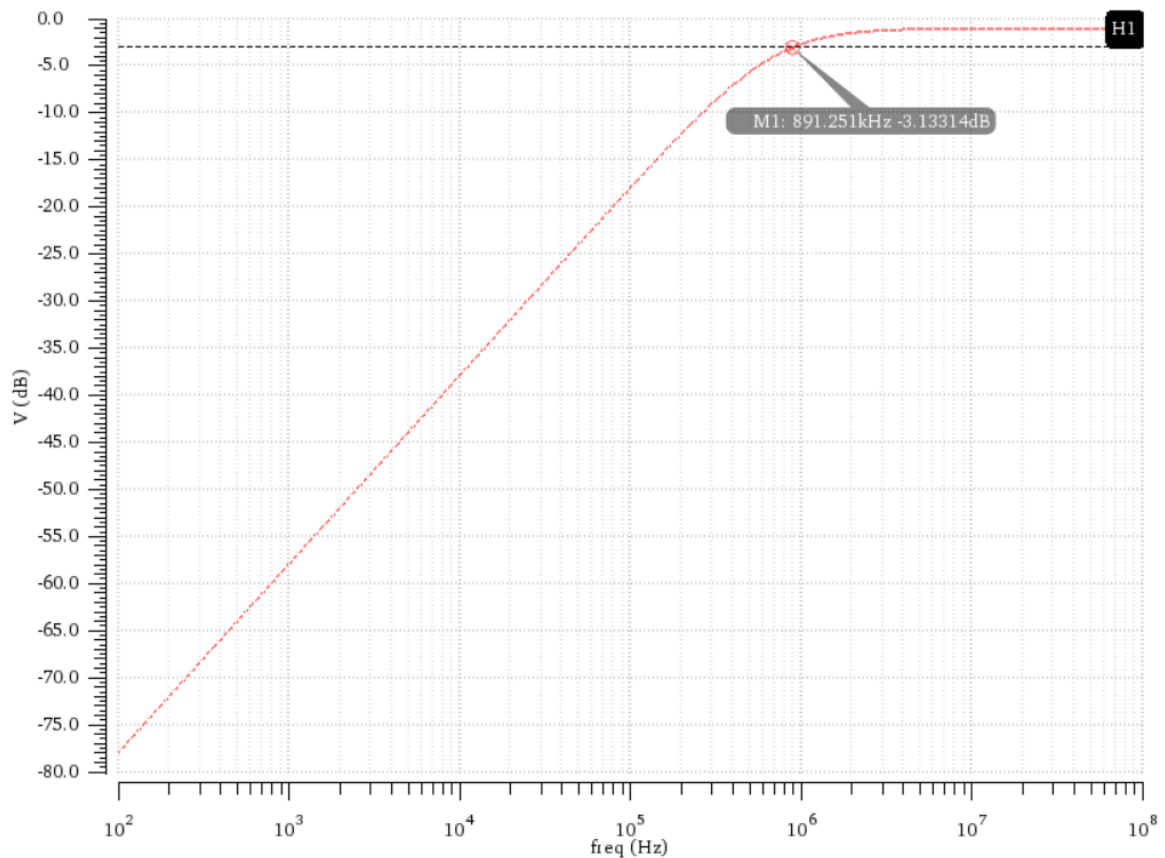
Figure 6.8 RC high pass filter ac simulation

6.4 Performance

The whole transducer interface block is simulated with two different input waveforms. Fig 6.9 is the continuous wave and Fig 6.10 is the pulsed-wave. These figures illustrate the procedure of transferring the 1.8V signal to 5V then to 30V. For each channel, the path from 1.8V to 30V is the same, so the main consideration for the output wave is the duty cycle. The 30V output has been tested under all corners and two temperature conditions. From Table 6.2, the maximum duty cycle variation is $\pm 2.3\%$.

Table 6.2 Duty cycle of 30V output wave under different conditions

Corner & Temp	TT, 27℃	FF, 0℃	**FS, 0℃**	SF, 0℃	SS, 0℃	FF, 50℃	FS, 50℃	**SF, 50℃**	SS, 50℃
Duty cycle	49.74%	49.59%	**47.68%**	51.49%	49.68%	49.90%	48.03%	**52.09%**	50.30%

Figure 6.9 Continuous wave mode

Figure 6.10 Pulsed-wave mode

75

Figure 6.11 All corners under 0℃ and 50℃ for the whole driver

As for power consumption, the 30V driver is power-hungry. To save the average power it consumes, several approaches are available. First is the on-time for the transducer, the value in Table 6.3 is based on an always-activated input. For the continuous wave mode, if the on-time of the transmitter is less than 5.3%, then the average power of eac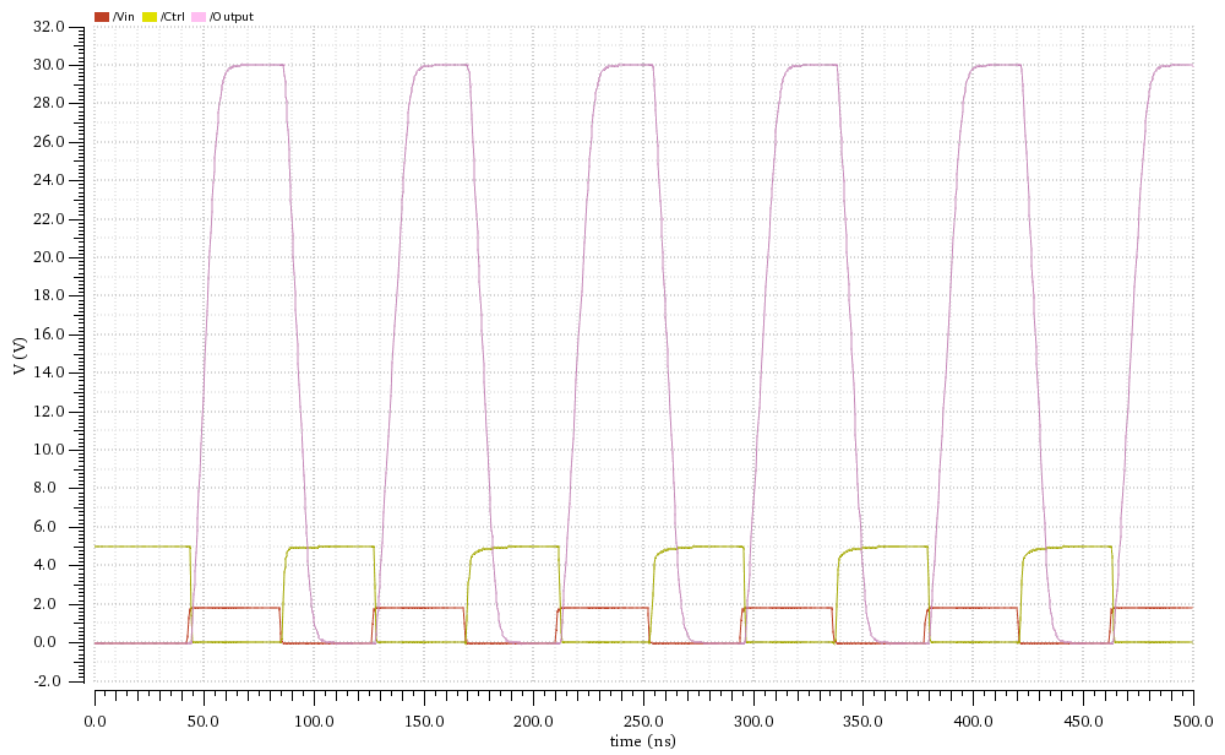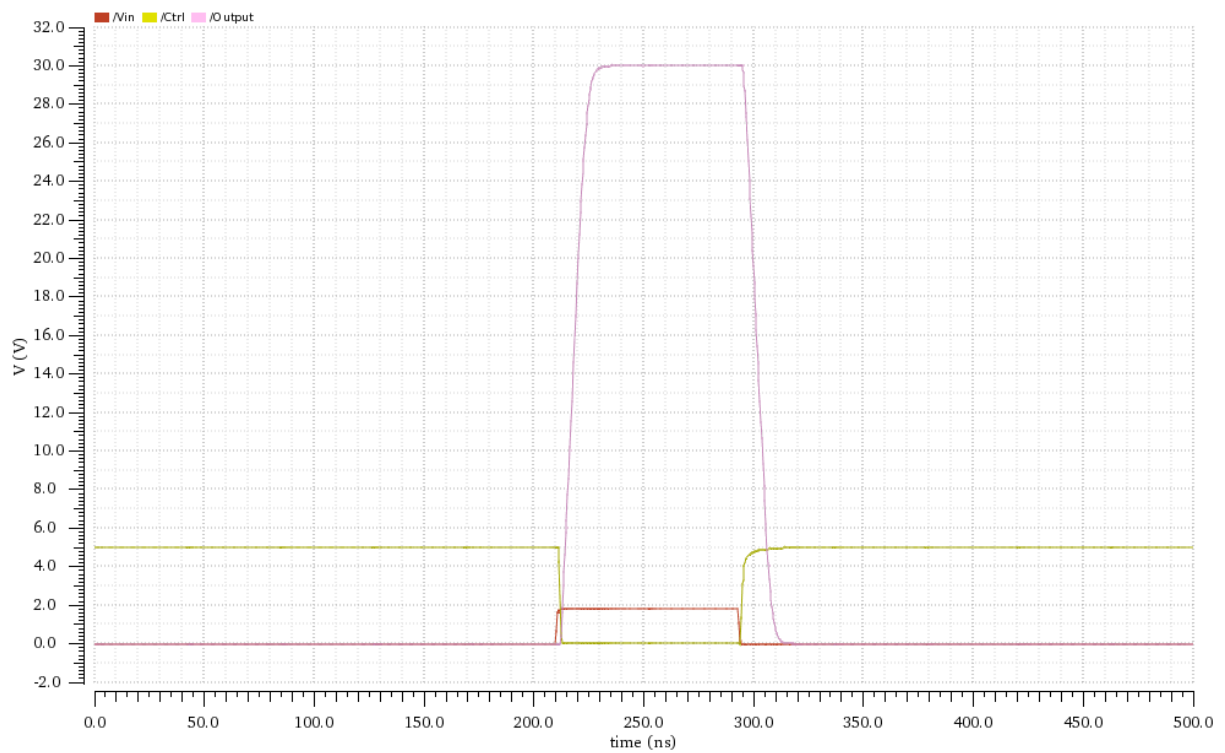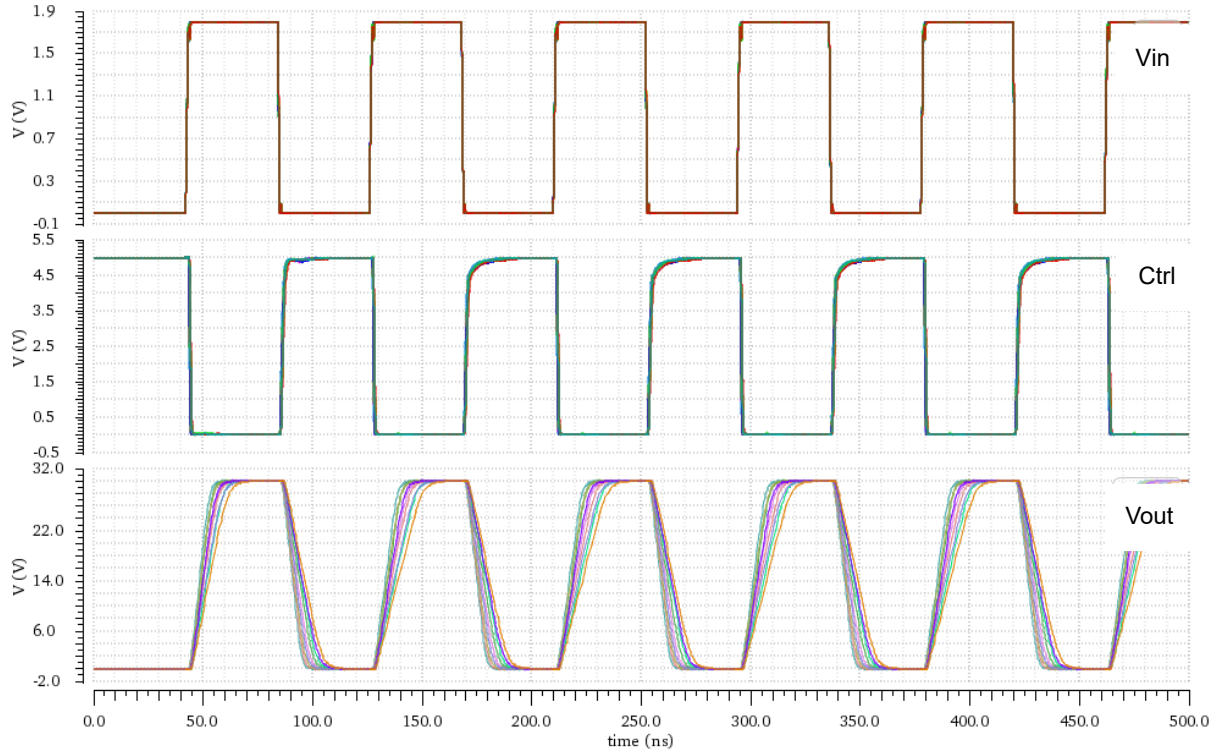h channel is less than 1mW. The second approach is based on the sparsity technique. As chapter 2 introduced, randomly shut down some of the channels would not affect the resolution, only the focal pressure is reduced. In our design, a 65x65 phased array, if only one channel is activated among every 19 channels, the average power will also be smaller than 1mW. There are still over 200 channels are activated, which is enough for imaging operation. The last approach is to avoid using 30V MOSFETs. In previous work[18], a 5V output wave under 10MHz gives a 100kPa pressure. With a good matching layer for the transducers, the 12MHz output wave in this design might be adequate with a 5V driver or a less-than-30V driver for the ultrasound stimulation.

Table 6.3 Power consumption for each device in the Transducer interface

(a) Continuous wave			
	VDD	Current	Power
30V Driver	30V	632uA	18.9mW
5V Level shifter	5V	18.75uA	93.75uW
MUX	1.8V	4.44nA	8nW
Input inverter	1.8V	370nA	0.67uW
Total(with 30V Driver)			18.9mW
Total(without 30V Driver)			94.4uW

(b) Pulsed wave			
	VDD	Current	Power
30V Driver	30V	105.5uA	3.165mW
5V Level shifter	5V	4.62uA	23.1uW
MUX	1.8V	1.87nA	3.37nW
Input inverter	1.8V	71.3nA	0.13uW
Total(with 30V Driver)			3.19mW
Total(without 30V Driver)			23.2uW

Chapter 7 Conclusions

7.1 The whole design of the transmitter
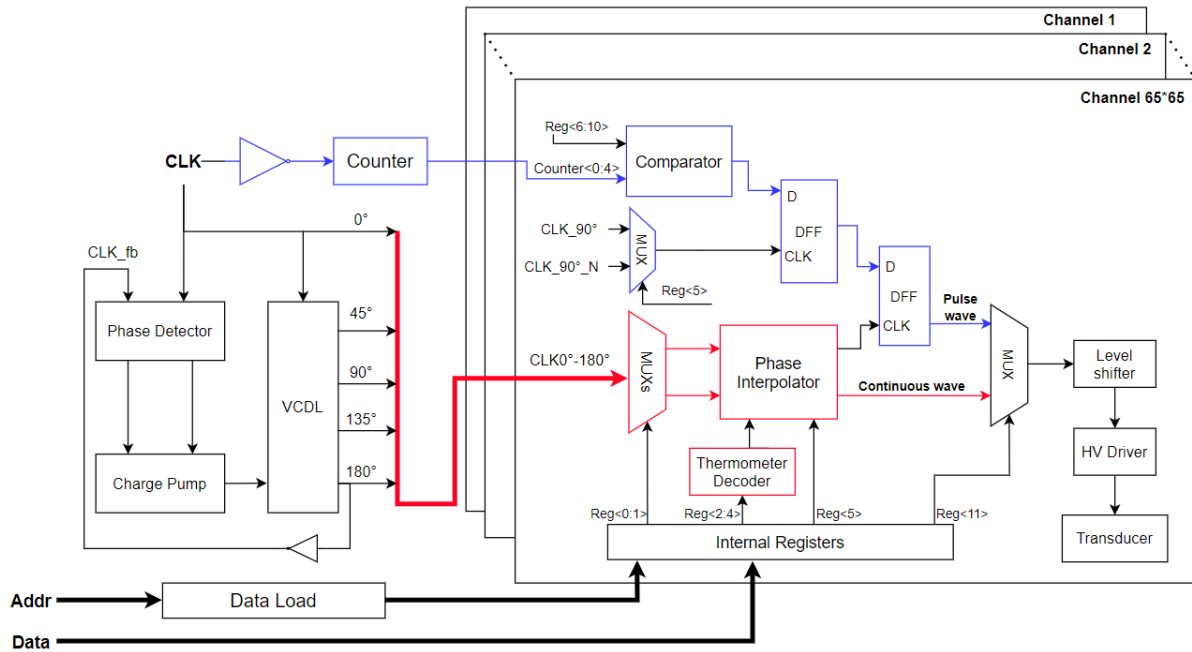
Figure 7.1 Complete block diagram of the whole design

The whole circuit design of this ultrasound transmitter is shown in Fig 7.1. The shared sections are the DLL, the global counter, and data loading line. The circuits inside channels are the same for each channel. The circuit design could be divided into four blocks.

The first block is the coarse delay generation which is a DLL, containing a phase detector, a charge pump, and a VCDL. Five reference clocks from $0°$ to $180°$ with a step of $45°$ are generated by this block.

The second block is the red outlined fine delay generation path. This block is mainly composed of a phase interpolator and its supporting devices(MUXs and the Thermometer Decoder).

The third block is the blue outlined PW mode function path. This block is responsible for a single pulsed-wave generation.

The last block is the transducer interface containing the level shifter and the high-voltage driver.

7.2 Thesis contribution

My contribution to this thesis work includes:
- Literature study on the vagus nerve, mechanism and devices for VNS, and phased array beamforming

- MATLAB simulation for 2D-array VNS system, and the array sparsity simulation
- DLL circuit design
- Channel circuit arrangement and design (for both stimulation and imaging function)
- A new approach in the digital circuit for pulsed-wave generation
- Level shifter and High-voltage driver design

7.3 Future work

The circuit design should be processed to the layout design and verification. After that, the post-layout simulation results could be extracted to the MATLAB simulation model to have a full array function simulation. PCB implementation and test are expected to be done in the future. This work is focused on the transmitter design, which is only part of the whole ultrasound VNS system. In the future, the transmitter channel needs to be implemented together with the receiver channel in the total array. The control system and transducer array fabrication are also required. Besides, the arrangement of transmitter channels and receiver channels using the sparsity is of great potential to achieve a power-efficient and high focusing performance system.

The other future improvements to this design are listed below:
- The charge pump could use a differential topology to improve the match of PMOS current source and NMOS current source. Now the current sources only have a good match in a limited voltage range. Out of the matching area, the control voltage keeps changing even the DLL is locked in.
- A duty-cycle recovery circuit and an extra loop to tune the VCDL could be implemented to have a better duty-cycle and jitter performance. The DLL in this design only compares the rising edge of the reference signal and VCDL output signal. The error on the falling edge should also be processed since the duty-cycle will be affected.
- The fine delay generation could consider a pipelined structure to reduce the number of current sources, which could reduce the power consumption. The control bit for polarity control might also be deleted if the imaging function could using the differential output signal.

Bibliography

[1]   MEpedia. "Vagus nerve." MEAction.

[2]   H. Leung, "The Physiological and Psychological Effects of Electrical Vagus Nerve Stimulation in Patients with Refractory Epilepsy," PhD, Faculty of Medicine, Dentistry, and Health Sheffield Institute for Translational Neuroscience, The University of Sheffield, 2017.

[3]   S. Breit, A. Kupferberg, G. Rogler, and G. Hasler, "Vagus Nerve as Modulator of the Brain–Gut Axis in Psychiatric and Inflammatory Disorders," *Frontiers in Psychiatry,* vol. 9, 2018, doi: 10.3389/fpsyt.2018.00044.

[4]   S. Eljamel, "Mechanism of Action and Overview of Vagus Nerve Stimulation Technology," in *Neurostimulation*, 2013, pp. 109-120.

[5]   (2017). *PMA P970003/S207: FDA Summary of Safety and Effectiveness Data*.

[6]   J. P. O'Reardon, P. Cristancho, and A. D. Peshek, "Vagus Nerve Stimulation (VNS) and Treatment of Depression: To the Brainstem and Beyond," (in eng), *Psychiatry (Edgmont),* vol. 3, no. 5, pp. 54-63, 2006. [Online].

[7]   E. N. Nicolai *et al.*, "Sources of Off-Target Effects of Vagus Nerve Stimulation Using the Helical Clinical Lead in Domestic Pigs," *bioRxiv,* p. 2020.01.15.907246, 2020, doi: 10.1101/2020.01.15.907246.

[8]   G. M. De Ferrari *et al.*, "Long-term vagal stimulation for heart failure: Eighteen month results from the NEural Cardiac TherApy foR Heart Failure (NECTAR-HF) trial," *International Journal of Cardiology,* vol. 244, pp. 229-234, 2017/10/01/ 2017.

[9]   MayoClinic. "Vagus nerve stimulation."

[10]   M. a. P. M. D. A. Patricia O. Shafer RN, MSN, CNRN. "Vagus Nerve Stimulation (VNS)." LivaNova.

[11]   Cyberonics. "Cyberonics Announces 100,000th Patient Implant of VNS Therapy®."

[12]   gammaCore. "How gammaCore works."

[13]   A. Bhuyan *et al.*, "A 32×32 integrated CMUT array for volumetric ultrasound imaging," in *2013 IEEE International Ultrasonics Symposium (IUS)*, 21-25 July 2013 2013, pp. 545-548, doi: 10.1109/ULTSYM.2013.0141.

[14]   M. D. Menz, P. T. Oralkan O Fau - Khuri-Yakub, S. A. Khuri-Yakub Pt Fau - Baccus, and S. A. Baccus, "Precise neural stimulation in the retina using focused ultrasound," (in eng), no. 1529-2401 (Electronic).

[15]   W. J. Tyler, S. W. Lani, and G. M. Hwang, "Ultrasonic modulation of neural circuit activity," *Current Opinion in Neurobiology,* vol. 50, pp. 222-231, 2018/06/01/ 2018.

[16]   J. Blackmore, S. Shrivastava, J. Sallet, C. R. Butler, and R. O. Cleveland, "Ultrasound Neuromodulation: A Review of Results, Mechanisms and Safety," *Ultrasound in Medicine & Biology,* vol. 45, no. 7, pp. 1509-1536, 2019/07/01/ 2019.

[17]   in *Clinical Ultrasound (Third Edition)*, P. L. Allan, G. M. Baxter, and M. J. Weston Eds. Edinburgh: Churchill Livingstone, 2011, pp. I-1-I-41.

[18]   T. Costa, C. Shi, K. Tien, and K. L. Shepard, "A CMOS 2D Transmit Beamformer With Integrated PZT Ultrasound Transducers For Neuromodulation," in *2019 IEEE Custom Integrated Circuits Conference (CICC)*, 14-17 April 2019 2019, pp. 1-4, doi: 10.1109/CICC.2019.8780236.

[19] M. E. Poorman *et al.*, "Open-source, small-animal magnetic resonance-guided focused ultrasound system," (in eng), *J Ther Ultrasound,* vol. 4, no. 1, pp. 22-22, 2016, doi: 10.1186/s40349-016-0066-7.

[20] K. Lou, S. Granick, and F. Amblard, "How to better focus waves by considering symmetry and information loss," *Proceedings of the National Academy of Sciences,* vol. 115, no. 26, p. 6554, 2018, doi: 10.1073/pnas.1803652115.

[21]   C. Chen *et al.*, "A front-end ASIC with receive sub-array beamforming integrated with a 32 × 32 PZT matrix transducer for 3-D transesophageal echocardiography," in *2016 IEEE Symposium on VLSI Circuits (VLSI-Circuits)*, 15-17 June 2016 2016, pp. 1-2, doi: 10.1109/VLSIC.2016.7573470.

[22] J. D. Vandersand, "An Analog Multiphase Self-Calibrating DLL to Minimiz o Minimize the Effects of Process, Supply Voltage, and Temperature Variations ariations " PhD, University of Tennessee, Knoxville, 2008.

[23] S. Tripathi, "Design of Delay-locked loop in 0.18-um CMOS Technology," Master, Department of Electronics & Communication Engineering, Thapar University, 2010.

[24] J. Guo, "DLL Based Single Slope ADC For CMOS Image Sensor Column Readout," M.S. Master Thesis, Delft University of Technology, 2011.

[25] B. Razavi. "PLLs and Synthesizers." Electrical Engineering Department, University of California, Los Angeles.

[26]   L. Won-Hyo, C. Jun-Dong, and L. Sung-Dae, "A high speed and low power phase-frequency detector and charge-pump," in *Proceedings of the ASP-DAC '99 Asia and South Pacific Design Automation Conference 1999 (Cat. No.99EX198)*, 21-21 Jan. 1999 1999, pp. 269-272 vol.1, doi: 10.1109/ASPDAC.1999.760011.

[27] G. Souliotis, C. Laoudias, F. Plessas, and N. Terzopoulos, "Phase Interpolator with Improved Linearity," *Circuits, Systems, and Signal Processing,* vol. 35, no. 2, pp. 367-383, 2016/02/01 2016, doi: 10.1007/s00034-015-0082-9.

[28] G. Jung *et al.*, "A Reduced-Wire ICE Catheter ASIC With Tx Beamforming and Rx Time-Division Multiplexing," *IEEE Transactions on Biomedical Circuits and Systems,* vol. 12, no. 6, pp. 1246-1255, 2018, doi: 10.1109/TBCAS.2018.2881909.

[29] I. O. Wygant *et al.*, "An integrated circuit with transmit beamforming flip-chip bonded to a 2-D CMUT array for 3-D ultrasound imaging," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control,* vol. 56, no. 10, pp. 2145-2156, 2009, doi: 10.1109/TUFFC.2009.1297.

[30] PIEZO.COM. "Material Properties."

[31] Z. YU, "Low-Power Receive-Electronics for a Miniature 3D Ultrasound Probe," PhD, Master of Science in Electrical Engineering Technische Universiteit Delft, 2012.

[32] G. Li *et al.*, "Investigation of High-Power Properties of PIN-PMN-PT Relaxor-Based Ferroelectric Single Crystals and PZT-4 Piezoelectric Ceramics," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control,* vol. 67, no. 8, pp. 1641-1646, 2020, doi: 10.1109/TUFFC.2020.2979217.

[33] PIEZO.COM. "T110-P1NO-0505 Piezoelectric Plate."

[34] M. Kumar, "Level Shifter Design For Low Power Applications," *International journal of computer science & information Technology (IJCSIT)* vol. 2, 2010.

[35]   B. Zhang, L. Liang, and X. Wang, "A New Level Shifter with Low Power in Multi-Voltage

System," in *2006 8th International Conference on Solid-State and Integrated Circuit Technology Proceedings*, 23-26 Oct. 2006 2006, pp. 1857-1859, doi: 10.1109/ICSICT.2006.306488.

[36]   C. Q. Tran, H. Kawaguchi, and T. Sakurai, "Low-power high-speed level shifter design for block-level dynamic voltage scaling environment," in *2005 International Conference on Integrated Circuit Design and Technology, 2005. ICICDT 2005.*, 9-11 May 2005 2005, pp. 229-232, doi: 10.1109/ICICDT.2005.1502637.

[37]   T. Hao-Yen, "High Voltage Level-Shifter Circuit Design for Efficiently High Voltage Transducer Driving," Msc, EECS Department, University of California, Berkeley, 2014.

[38]   M. Tan, "A Front-end ASIC with High-Voltage Transmit Switching and Receive Digitization for Forward-Looking Intra-Vascular Ultrasound," Msc, Delft University ofTechnology, 2016.