© 2021 Hanfei Geng

AREA-CONSTRAINED FREQUENCY ESTIMATION FOR FOCAL PLANE ARRAYS

BY

HANFEI GENG

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Electrical and Computer Engineering
in the Graduate College of the
University of Illinois Urbana-Champaign, 2021

Urbana, Illinois

Adviser:

Professor Naresh R. Shanbhag

ABSTRACT

This thesis focuses on the problem of detecting the frequency and amplitude of signals generated by focal plane array (FPA) on a per-pixel basis, i.e., pixel-matched processing. Addressing this problem requires the design of signal processing algorithms and circuit architectures subject to stringent area constraints imposed by the pixel pitch, e.g., $144\,\mu\text{m}^2$. Imaging systems incorporating FPAs with such pixel-matched processors will be able to perform real-time, energy-efficient and low-latency feature extraction in a massively parallel manner.

Classical approaches to sinusoidal frequency estimation and detection such as those based on the Discrete Fourier Transform (DFT) and its complexity-reduced version, the Fast Fourier Transform (FFT), are unsuitable for this application their use of complex-valued coefficients/weights.

In this thesis, we study the detection performance of the Discrete Hartley Transform (DHT) in addition to its variants which employ real-valued weights, as possible alternatives to the DFT. Our analysis indicates that thresholding the block averaged squared magnitude of the DHT samples generates a detection accuracy close to that of the DFT. However, the DHT suffers from the *singularity problem* (detection failure) when the frequency of the input is perfectly aligned with one of the frequency bins.

To solve the singularity problem, we propose two variants of the DHT: the Dithered Discrete Hartley Transform (D-DHT) and the Jittered Discrete Hartley Transform (J-DHT). Both variants solve the singularity problem by spreading the input signal energy across the signal spectrum. However, due to its narrower spreading effect, the D-DHT requires $2\times$ fewer computations than the J-DHT to match the detection accuracy of the DFT.

Employing fixed-point analysis, we determine the minimum precision requirements for the D-DHT and the DFT when mapped to the direct dot-product (DP) architecture. The minimum precision values for the input,

weight, and the accumulator is combined with the area estimates of a 1-bit full adder (FA) and a 1-bit register in a 65 nm CMOS process to obtain the area costs of the D-DHT and the DFT. Results show that the D-DHT achieves $\sim 2\times$ area reduction over the DFT with no significant impact on detection accuracy.

However, D-DHT's area is still $\sim 3\times$ greater than the pixel pitch. Since the FPA sampling rates are at least an order-of-magnitude lower than the maximum throughput achievable by a multiply-accumulate (MAC) unit in 65 nm CMOS process, our study suggests the use of bit-serial time-multiplexed K-pixel D-DHT architecture as a reasonable solution.

To my parents, for their love and support.

ACKNOWLEDGMENTS

First and foremost, I would like to express my sincere gratitude to my advisor, Professor Naresh R. Shanbhag for his continuous guidance during the past two years as well as my four years of undergraduate study. This work would not have been possible without his support. I would like to acknowledge Mark Massie for providing the inspiration for this work. I am also very grateful for Charbel Sakr, Hassan Doubk, and Ameya Patel's mentoring. Their academic rigor always impresses me and motivates me to improve myself as a researcher. I would also like to thank all of my labmates: Saion Kumar Roy, Kuk-Hwan Kim, Howard Li, Hanmo Ou, and Michael Tuttle for creating a positive and friendly research environment.

I am also indebted to all of the dancing crews that I was lucky to be part of at UIUC for the amazing performance experience. Particularly, I would like to thank Michael Lian, Sungmin Jang, and Chulhee Cho for trusting me to plan and design performances for Mos Sick Crew that they founded. I am also grateful for the support from all of its crew members: Cong Li, Aaron Liu, Quanyu Zuo, Jiake Wen, Cloud Li, Krystal Zheng and Sherry Xu. I am glad that I was able to grow with this team and honored to perform with them for the Chinese Consulate General in Chicago in 2019.

Finally I would like to extend my deepest gratitude to my parents, David and Qing. This year marks my tenth year studying in America. Without their clairvoyance and sacrifices at the time, I would have neither been able to learn from the amazing people that I have met from all over the world nor become who I am today. This thesis is dedicated to them.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS viii

CHAPTER 1 INTRODUCTION 1
1.1 Motivation 1
1.2 Previous Work 3
1.3 Problem Definition 4
1.4 Thesis Contributions and Organization 6

CHAPTER 2 THE DETECTION PERFORMANCE OF DISCRETE HARTLEY-BASED TRANSFORMS 7
2.1 Discrete Hartley Transform 7
2.2 Squared Magnitude of Discrete Hartley Transform 8
2.3 The Singularity Problem in Discrete Hartley Transform 10
2.4 Dithered Discrete Hartley Transform 10
2.5 Jittered Discrete Hartley Transform 12
2.6 Simulation Results 14
2.7 Summary 19

CHAPTER 3 FIXED-POINT ANALYSIS ON DIRECT DOT-PRODUCT ARCHITECTURE 20
3.1 Direct Dot-Product Architecture 20
3.2 Fixed-Point Analysis 22
3.3 Simulation Results 25
3.4 Area Complexity Comparison in Fixed Point 29
3.5 Meeting the Area Constraint for Small Pixels 30
3.6 Summary 31

CHAPTER 4 CONCLUSION AND FUTURE WORK 32
4.1 Conclusion 32
4.2 Future Work 33

REFERENCES 34

APPENDIX A DERIVATIONS FOR ANALYTICAL EXPRESSIONS IN CHAPTER 2 38
A.1 Preliminaries 38
A.2 Derivation for Equation (2.3) 39
A.3 Derivation for Equation (2.7) 40
A.4 Derivation for Equations (2.8) and (2.9) 41
A.5 Derivation for Equations (2.10), (2.11), and (2.12) 43
A.6 Derivation for Equations (2.15) and (2.16) 44
A.7 Derivation for Equation (2.17) 45

LIST OF ABBREVIATIONS

BHT	Binary Hypothesis Test
D-DHT	Dithered Discrete Hartley Transform
DFT	Discrete Fourier Transform
DHT	Discrete Hartley Transform
DPA	Digital Processor Array
DP	Dot Product
DTFT	Discrete-Time Fourier Transform
FA	Full Adder
FL	Floating Point
FFT	Fast Fourier Transform
FPA	Focal Plane Array
FX	Fixed Point
IR	Infrared
J-DFT	Jittered Discrete Hartley Transform
MAC	Multiply and Accumulate
ML	Maximum Likelihood
PE	Processing Element
PSD	Power Spectral Density
ROIC	Read-Out Integrated Circuit
SNR	Signal to Noise Ratio
SM	Squared Magnitude

CHAPTER 1

INTRODUCTION

1.1 Motivation

A focal plane array (FPA) is a type of image sensor that plays an important role in many thermal infrared (IR) sensing applications. Such applications include 24/7 persistent surveillance, border patrol, and environmental monitoring [1–5]. Deployed in many imaging systems, FPAs pack large number of pixels in a tight 2-D array fashion, where each pixel is comprised of a photodiode in a *detector* layer and a subsequent read-out integrated circuit (ROIC) in a ROIC layer (see Fig. 1.1). Traditionally, when a FPA is exposed to illumination, the ROIC generates digital samples from the photodiode current I_{PD} and transports NL samples, where L is the number of consecutive blocks, each of N samples, to a digital processor. To facilitate the transfer for all pixels, a serializer circuit is employed [1]. The digital processor receives these samples and performs further processing such as image formation and spectral analysis using each pixel signal.

In mission-critical applications, there has been an increasing demand for large-pixel-count FPAs that can produce high-resolution images [1]. Along with this demand on image quality comes the need to interpret them in an energy-efficient and low-latency manner. One fundamental task required is to detect sinusoids on a per-pixel basis while estimating their parameters: amplitude, frequency and phase. Fast sinusoid detection and parameter estimation is key for real-time feature extraction, which has become a key component in wide-area IR imaging systems [1, 6].

For a K-pixel FPA, the conventional pixel processing scheme in Fig. 1.1 is expensive in terms latency and energy consumption for large K [6, 7]. Because more pixels get packed in an FPA to meet this application demand, time multiplexing is needed to schedule the transfer of samples from each

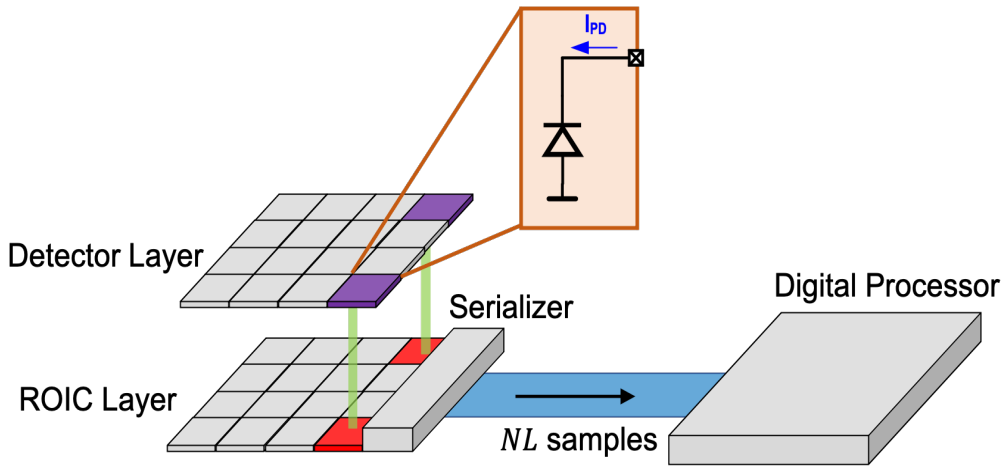


Figure 1.1: Traditional FPA pixel processing.

pixel. Moreover, the data communication between the ROIC and the processor induces high energy consumption. As a result, a new approach to processing these pixels is necessary.

In order to enable fast sinusoid detection and estimation, one solution is to leverage the massive parallelism in FPA by establishing a direct link between the pixel and the processor using 3-D integration (see Fig. 1.2) [7–10]. Instead of sending the samples to a central digital processor via peripheral circuitry as in Fig. 1.1, each ROIC in the ROIC layer sends its data to a pitch-matched processing element (PE) in a digital processor array (DPA). Between every pixel and every PE is a direct link so that the data transfer and processing occur independently among pixels.

The main challenge in the 3-D integrated approach in Fig. 1.2 is designing a pixel-matched PE. Typical area per pixel is $\sim 625\,\mu\text{m}^2$. Some technology nodes allow smaller area $\leq 144\,\mu\text{m}^2$ [1, 5, 7]. These stringent area budgets limit the number of full adders (FA) and registers allowed for the implemented algorithm. For example, in 65 nm technology, a 1-bit register occupies $9.9\,\mu\text{m}^2$ of area as does a 1-bit FA. As a result, an 8-bit multiply-and-accumulate (MAC) unit occupies an area of $633.6\,\mu\text{m}^2$, which far exceeds the area budget. Therefore, to avoid storing the input samples, the PE needs to process them in a streaming fashion with minimum storage. Consequently, it becomes crucial to search for a low-complexity fixed-point detection and estimation algorithm that provides the desired level of accuracy with minimum precision.

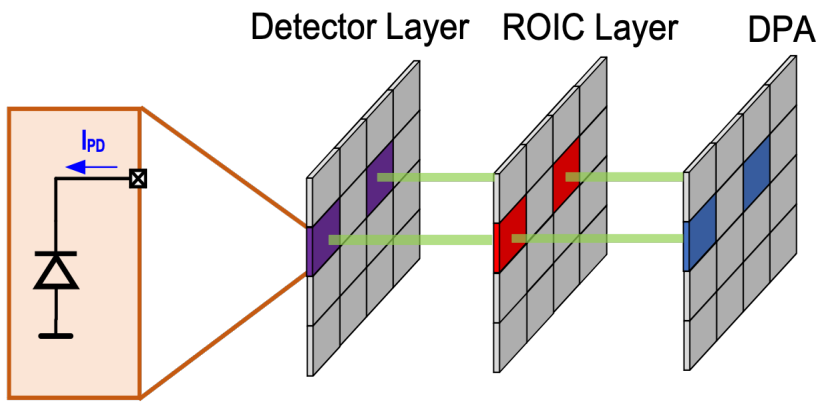


Figure 1.2: Pitch-matched FPA pixel processing using 3-D integration.

1.2 Previous Work

The problem of sinusoids detection and parameter estimation has been studied extensively [11–16]. Previous works formulate the sinusoid detection problem as a binary hypothesis test (BHT), where the null (H_0) and alternative (H_1) hypotheses correspond to the presence of noise only and the superposition of sinusoid and noise, respectively. Since the sinusoid's parameters are unknown, the set of all possible parameters leads to many possible BHTs. An ideal solution is to find a decision rule satisfying the Neyman-Pearson criterion for *all* parameter choices. However, such a decision rule often does not exist [13, 15].

An alternative strategy is to replace the unknown parameters with their maximum likelihood (ML) estimates [11, 13, 15]. In the case of $L = 1$, the ML estimator maximizes the periodogram of the observed signal in each pixel [11]. The associated frequency and amplitude at which such maximum occurs are the ML estimates. In this setup, the detection rule leads to an intuitive solution where the sufficient statistic used by the decision rule is the maximum periodogram value [11].

Obtaining the periodogram requires computing the discrete-time Fourier transform (DTFT) which is practically infeasible due to the need for an infinite number of frequency samples. Therefore, the joint-detection-and-estimation problem is usually divided into two steps. The first step involves using the discrete Fourier transform (DFT) to coarsely estimate the parameters of the sinusoid, followed by a second step to fine-tune the coarse estimates [12]. The DFT is a prevalent choice for coarse estimation as it is

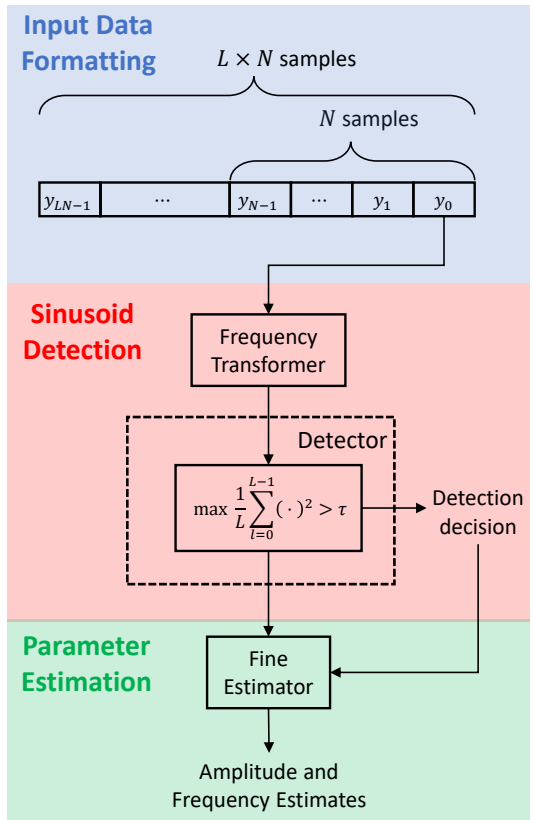


Figure 1.3: The event-driven approach to joint sinusoid detection and estimation.

a sampled alternative to the DTFT. Given N times samples, the DFT correlates the input using N complex sinusoids of different frequencies, i.e., N frequency bins. For real-valued signals, correlation using $\frac{N}{2}$ frequency bins is sufficient by virtue of DFT's conjugate symmetry property.

1.3 Problem Definition

Following the approach in [12], we adopt an event-driven approach to the joint-detection-and-estimation problem (see Fig. 1.3). After a pixel's ROIC reads out L blocks each of N observation samples y_n, the frequency transform

of each block is computed. The detector then averages the squared magnitude (SM) of the resultant frequency samples. It outputs a detection decision by thresholding the maximum of the resultant average. The fine estimator is activated only if a sinusoid is detected. It produces the final estimation results by treating the thresholded squared amplitude and its associated frequency as coarse estimates.

In the BHT setup, we write the observed signal y_n as:

$$\begin{cases} H_1 : y_n = x_n + v_n \\ H_0 : y_n = v_n \end{cases} \tag{1.1}$$

where $x_n = A\cos(\omega_o n + \phi)$ is a sinusoid of unknown amplitude A, digital frequency ω_o, and phase ϕ. The real-valued background noise v_n are samples drawn from a zero-mean Gaussian distribution. In a typical application, the input signal-to-noise ratio (SNR) ranges from $-10\,\text{dB}$ to $30\,\text{dB}$.

The key to the success of this event-driven approach is a low-complexity frequency transformer that meets the area constraint while retaining high detection accuracy. An accurate and robust frequency transformer enables accurate detection. It also reduces energy consumption since it avoids activating the fine estimator in the absence of a signal. As it is a complex-valued transform, the DFT needs two real-valued MAC branches to generate one frequency sample of a real-valued pixel signal. Moreover, the fixed-point (FX) fast Fourier Transform (FFT) [17–21] has a large storage requirement because of the need to store intermediate results, which renders it unsuitable for area-constrained platforms. Therefore, we consider real-valued transforms such as the Discrete Hartley Transform (DHT) [22] in the context of sinusoid detection. To minimize the area complexity, we only consider half-band frequency transforms, i.e., for an even-length block of N observation samples, only the first $\frac{N}{2}$ frequency bins are computed. Hence, we assume N is even. According to the conjugate symmetry property of the DFT, it does not suffer from any losses due to the use of half of the frequency bins, whereas the impact on the DHT is unknown in the context of detection.

1.4 Thesis Contributions and Organization

In this thesis, we focus on the design of the frequency transformer in Fig. 1.3 to explore the possibility of replacing the DFT with DHT, which is a real-valued frequency transform. The contributions of this thesis are summarized as follows: (1) we find that the DHT suffers from the singularity problem, where it is unable to detect sinusoids of frequency $\omega_o = \frac{2\pi k_o}{N}, k_o \in \mathbb{Z}$, i.e., the sinusoids that lie at the center of a frequency bin; (2) by analyzing why this singularity problem occurs, we propose the Dithered Discrete Hartley Transform (D-DHT) and the Jittered Discrete Hartley Transform (J-DHT) as solutions and show their robustness; (3) we show through simulation that the D-DHT is a better choice as a frequency detector as it needs $2\times$ less block averaging compared to the J-DHT to match the detection accuracy of the DFT; and (4) using FX analysis, we find the minimal precision requirement for the D-DHT and the DFT, and estimate the area required by them implemented on a direct dot-product architecture. We show that the D-DHT achieves $2\times$ area reduction relative to the DFT, but there remains a gap between its area cost and the constraint posted by high-resolution FPAs. To address this issue, we suggest time multiplexing of the processing of multiple pixel signals using a single D-DHT processor with bit-serial multiplier and adder.

The rest of the thesis is organized as follows: Chapter 2 analyzes the DHT, D-DHT, and J-DHT's floating-point (FL) detection performance relative to the DFT. The derivations for analytic expressions in Chapter 2 is found in Appendix A. Chapter 3 finds the minimal precision requirement for the D-DHT and the DFT, in addition to using these values and estimating the required area of both algorithms. Finally Chapter 4 concludes this thesis and lists possible future directions of this work.

CHAPTER 2

THE DETECTION PERFORMANCE OF DISCRETE HARTLEY-BASED TRANSFORMS

In this chapter, we compare DHT's FL detection performance to that of the DFT for input frequency $\omega_o = \frac{2\pi k_o}{N}$. Analysis of the DHT reveals detection failure when $k_o \in \mathbb{Z}$. By analyzing the root cause of this singularity problem, we present the D-DHT and the J-DHT as potential solutions.

Through simulations, we compare the detection performance of the D-DHT and the J-DHT to that of the DFT, and conclude that among the discrete Hartley-based transforms (DHT, D-DHT, and J-DHT), the D-DHT is the most promising candidate to implement an area-constrained frequency transformer.

2.1 Discrete Hartley Transform

Unlike the DFT, the DHT is a real-valued frequency transform. For a length N sinusoidal block with amplitude A, frequency ω_o, and phase ϕ, $\{x_n \in \mathbb{R} \mid x_n = A\cos(\omega_o n + \phi),\ n \in [N]\}$, each DHT frequency sample H_k is given by:

$$H_k = \frac{\sqrt{2}}{N} \sum_{n=0}^{N-1} x_n \cos\Big(\frac{2\pi kn}{N} - \frac{\pi}{4}\Big) \tag{2.1}$$

On the other hand, the DFT X_k frequency sample is given by:

$$X_k = \frac{1}{N}\Big[\sum_{n=0}^{N-1} x_n \cos\Big(\frac{2\pi kn}{N}\Big) - j\sum_{n=0}^{N-1} x_n \sin\Big(\frac{2\pi kn}{N}\Big)\Big] \tag{2.2}$$

Comparing (2.1) and (2.2) reveals that the DHT requires only one real-valued dot-product (DP) computation, whereas the DFT requires two of them. Thus, the DHT is better suited for area-constrained implementations compared to the DFT.

As derived in Appendix A.2, the DHT can be used to obtain the DFT according to:

$$X_k = \frac{H_{N-k} + H_k}{2} + j\frac{H_{N-k} - H_k}{2} \tag{2.3}$$

Thus, it is possible to use the DFT frequency samples for detection by first computing the DHT frequency samples. However, this approach requires a full-band DHT computation which is area expensive. Therefore, we discuss the approach of matching DFT's SM by averaging DHT's SM over L consecutive blocks in Section 2.2.

2.2 Squared Magnitude of Discrete Hartley Transform

The detector in Fig. 1.3 uses the maximum averaged squared magnitude (SM) of the frequency sample from L blocks, each of which consists of N samples, as the sufficient statistic. It is thus necessary to quantify the deviation of the DHT's SM from that of the DFT.

For the lth sinusoidal block, where $l \in [L]$, each of length N with amplitude A, frequency ω_o, and phase ϕ, $\{x_{n,l} \in \mathbb{R} \mid x_{n,l} = A\cos(\omega_o(n + Nl) + \phi), n \in [N]\}$, we can write the the SM of the lth block DHT at the kth bin $|H_{k,l}|^2$ as follows:

$$|H_{k,l}|^2 = |X_{k,l}|^2 + \delta_{k,l} \tag{2.4}$$

where $\delta_{k,l}$ is the per-block deviation between the SMs of two transforms, and $H_{k,l}$ and $X_{k,l}$ are the associated DHT and DFT frequency samples, respectively, given by:

$$H_{k,l} = \frac{\sqrt{2}}{N}\sum_{n=0}^{N-1} x_{n,l}\cos\left(\frac{2\pi kn}{N} - \frac{\pi}{4}\right) \tag{2.5}$$

$$X_{k,l} = \frac{1}{N}\Big[\sum_{n=0}^{N-1} x_{n,l}\cos\left(\frac{2\pi kn}{N}\right) - j\sum_{n=0}^{N-1} x_{n,l}\sin\left(\frac{2\pi kn}{N}\right)\Big] \tag{2.6}$$

If $k_o \notin \mathbb{Z}$, $\delta_{k,l}$ is given by:

$$\begin{aligned}\delta_{k,l} = &\frac{A^2 \sin^2\left(\frac{NE}{2}\right)}{4N^2 \sin^2\left(\frac{E}{2}\right)} \sin\Big((N-1)E + 2\phi + 2\omega_o N l\Big) \\ &- \frac{A^2 \sin^2\left(\frac{NF}{2}\right)}{4N^2 \sin^2\left(\frac{F}{2}\right)} \sin\Big((N-1)F + 2\phi + 2\omega_o N l\Big) \\ &- \frac{A^2\Big[1 - \cos(N\omega_o)\Big]}{2N^2\Big[\cos(\frac{2\pi k}{N}) - \cos(\omega_o)\Big]} \sin\left(\frac{2\pi k}{N}\right)\end{aligned} \tag{2.7}$$

where $E = \omega_o + \frac{2\pi k}{N}$ and $F = \omega_o - \frac{2\pi k}{N}$. The detailed derivation can be found in Appendix A.3. The block-averaged deviation $\Delta_k(L) = \frac{1}{L}\sum_{l=0}^{L-1} \delta_{k,l}$ has the following asymptotic behavior as L goes to infinity:

$$\Delta_k(\infty) = \lim_{L\to\infty} \Delta_k(L) = -\frac{A^2\Big[1 - \cos(N\omega_o)\Big]}{2N^2\Big[\cos(\frac{2\pi k}{N}) - \cos(\omega_o)\Big]} \sin\left(\frac{2\pi k}{N}\right) \tag{2.8}$$

where $\Delta_k(\infty)$ is the asymptotic deviation of DHT's block-averaged SM for large L.

To further understand (2.8), we find the following upper bound for $\Delta_k(\infty)$ when N is large:

$$|\Delta_k(\infty)| \leq \frac{A^2 \pi k}{N} \tag{2.9}$$

Appendix A.4 contains the detailed derivations for (2.8) and (2.9). Equation (2.9) reveals the following insight: in the absence of noise, given sufficient block averaging, DHT's deviation approaches zero as N increases. Therefore, this property suggests that the DHT has the potential to replace the DFT when $k_o \notin \mathbb{Z}$. Yet the trade-off is the increased latency and computational complexity due to the need for block averaging.

2.3 The Singularity Problem in Discrete Hartley Transform

When $k_o = k \in \mathbb{Z}$, the deviation $\delta_{k,l}$ is given by:

$$\delta_{k,l} = -\frac{A^2}{4}\sin(2\phi) \tag{2.10}$$

and the resultant DFT and DHT's SMs at the lth block are given by:

$$|X_{k,l}|^2 = \frac{A^2}{4} \tag{2.11}$$

$$|H_{k,l}|^2 = \frac{A^2}{4}\Big[1 - \sin(2\phi)\Big] \tag{2.12}$$

where $X_{k,l}$ and $H_{k,l}$ are the DFT and DHT frequency samples, respectively. Appendix A.5 contains the detailed derivation. Equation (2.12) mathematically defines the singularity problem. For any block l, when ϕ belongs to the set $\{\phi \in \mathbb{R} \mid \phi = 2m\pi + \frac{\pi}{4}, m \in \mathbb{Z}\}$, the resultant DHT's SM $|H_{k,l}|^2$ will be *zero* despite the SM of the DFT $|X_{k,l}|^2$ being non-zero. The DHT's signal energy completely shifts to the $(N-k)^{\text{th}}$ bin. In this case, the input signal is orthogonal to the cosine basis of the DHT. Since the SM of the DHT at l^{th} block is independent of block index l, block averaging cannot function as a remedy as it does for the case when $k_o \notin \mathbb{Z}$ described in Section 2.2.

2.4 Dithered Discrete Hartley Transform

The fundamental difference between the DHT and the DFT is that the DHT correlates the input with a cosine as opposed to the DFT which correlates the input with *both* a sine and a cosine basis. As a result, only the DHT is susceptible to the singularity problem whereas the DFT is not.

Therefore, one solution is to correlate half of the time samples with a cosine basis and the other half with a sine basis, and combine the correlation results at last. Following this philosophy, we propose the D-DHT as an alternative to the DHT.

The D-DHT computes a frequency sample as follows: for the lth sinusoidal block of length N with amplitude A, frequency ω_o, and phase ϕ, $\{x_{n,l} \in$

$\mathbb{R} \mid x_{n,l} = A\cos(\omega_o(n+Nl)+\phi), n \in [N]\}$, the D-DHT frequency sample at the kth bin $D_{k,l}$ is given as:

$$D_{k,l} = \frac{\sqrt{2}}{N}\sum_{n=0}^{N-1} x_{n,l}\cos\Big(\frac{2\pi kn}{N} - \frac{\pi}{4} + \phi_n\Big) \tag{2.13}$$

where $\phi_n = \frac{\pi}{4}\Big(1-(-1)^n\Big)$, which is a binary alternating sequence between 0 and $\frac{\pi}{2}$. Based on trigonometric identities, (2.13) can be re-formulated as a summation of a cosine term $C_{k,l}$ and a sine term $S_{k,l}$ as follows:

$$\begin{aligned} D_{k,l} &= \frac{\sqrt{2}}{N}\Big[C_{k,l} + S_{k,l}\Big] \\ &= \frac{\sqrt{2}}{N}\Big[\sum_{n=0}^{\frac{N}{2}-1} x_{2n,l}\cos\Big(\frac{2\pi k(2n)}{N} - \frac{\pi}{4}\Big) \\ &- \sum_{n=0}^{\frac{N}{2}-1} x_{2n+1,l}\sin\Big(\frac{2\pi k(2n+1)}{N} - \frac{\pi}{4}\Big)\Big] \end{aligned} \tag{2.14}$$

Equation (2.14) shows that the D-DHT correlates the even input samples with a cosine basis and the odd samples with a sine basis.

The D-DHT solves the singularity problem thanks to the following property of the SMs at bin k and bin $\frac{N}{2}-k$, when $k_o = k \in \mathbb{Z}$ but $k \neq \frac{N}{4}$:

$$|D_{k,l}|^2 = \frac{A^2}{8}\Big[1+\cos(2\phi)\Big] \tag{2.15}$$

$$\Big|D_{\frac{N}{2}-k,l}\Big|^2 = \frac{A^2}{8}\Big[1-\cos(2\phi)\Big] \tag{2.16}$$

where $|D_{k,l}|^2$ and $\Big|D_{\frac{N}{2}-k,l}\Big|^2$ are SMs at bin k and bin $\frac{N}{2}-k$, respectively. The opposite signs in front of the cosine basis in (2.15) and (2.16) guarantee the singularity problem will not occur, since $|D_{k,l}|^2$ and $\Big|D_{\frac{N}{2}-k,l}\Big|^2$ cannot both be zero for any value of the input phase ϕ. Comparing (2.11) and (2.15)-(2.16) reveals that the D-DHT spreads the signal energy across bins k and $\frac{N}{2}-k$. Therefore, the D-DHT necessitates block averaging to match DFT's detection performance.

When $k_o = k = \frac{N}{4}$, $|D_{k,l}|^2$ is given by:

$$|D_{k,l}|^2 = \left|D_{\frac{N}{2}-k,l}\right|^2 = \frac{A^2}{4}\Big[1 + \sin(2\phi)\Big] \tag{2.17}$$

The detailed derivation is found in Appendix A.6. Similar to (2.12), the singularity problem manifests itself again. Bins k and $\frac{N}{2} - k$ are the same bin. As a result, block averaging will not help. However, the difference is that the singularity problem occurs *only* when N is a multiple of four, whereas in the case of the DHT, it occurs for any bin k. Hence, choosing N not to be a multiple of four ensures that the D-DHT will not suffer from the singularity problem.

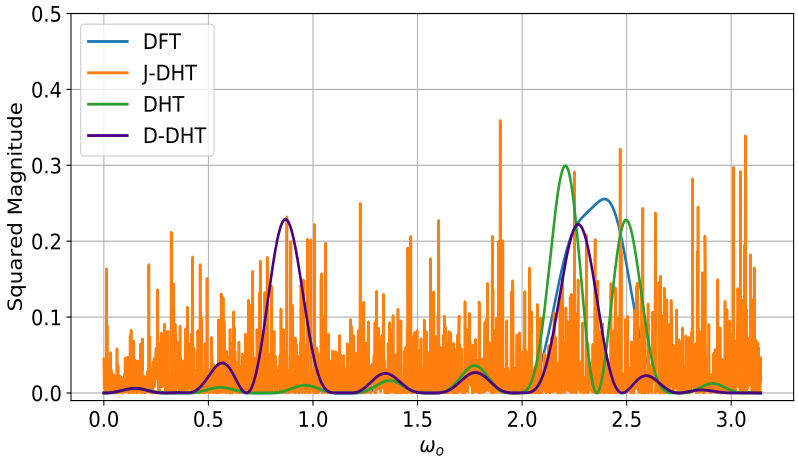
2.5 Jittered Discrete Hartley Transform

In this section, we propose the J-DHT as another approach to address the singularity problem. The J-DHT assigns a random phase to the DHT coefficients while computing a frequency sample. For the lth input block of length N with amplitude A, frequency ω_o, and phase ϕ, $\{x_{n,l} \in \mathbb{R} \mid x_{n,l} = A\cos(\omega_o(n + Nl) + \phi), n \in [N]\}$, the J-DHT computes the kth frequency sample $J_{k,l}$ as follows:

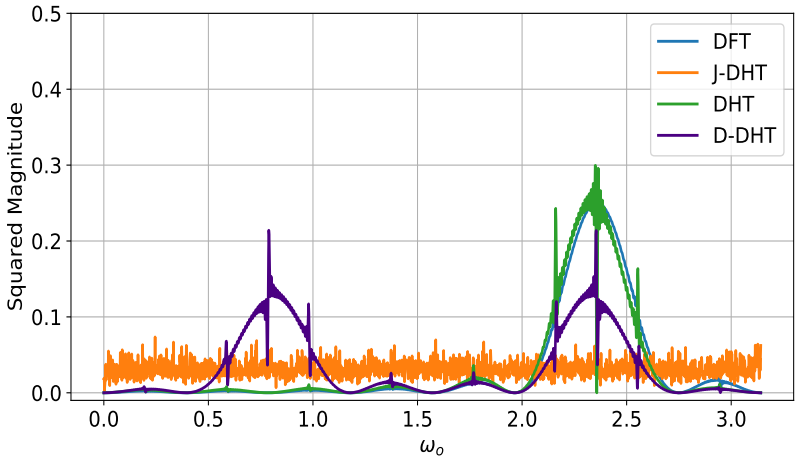
$$J_{k,l} = \frac{\sqrt{2}}{N}\sum_{n=0}^{N-1} x_{n,l}\cos\Big(\frac{2\pi kn}{N} - \frac{\pi}{4} + \Phi_n\Big) \tag{2.18}$$

where $\Phi_n \sim Unif(0, 2\pi)$, where $Unif(a, b)$ is the uniform distribution with domain $[a, b]$. To understand J-DHT's SM response, we set $A = 1, k = 6, N = 16$, and sweep ω_o while computing the L-block averaged SM $\frac{1}{L}\sum_{l=0}^{L-1}|J_{k,l}|^2$ for an arbitrary bin ($k = 6$).

Figure 2.1 compares J-DHT's SM to those of the DFT, DHT, and D-DHT. In the case of the DFT, the signal induces a high SM in the kth bin if the input frequency ω_o is in the proximity of the frequency $\frac{2\pi k}{N}$. The DHT shows similar behavior and approaches the DFT with block averaging except at $\omega_o = \frac{2\pi k}{N}$. Due to the spreading of the signal energy characterized by (2.15) and (2.16), the D-DHT induces high SMs but lower than that of the DFT in the proximity of frequencies $\frac{2\pi k}{N}$ and $\frac{2\pi}{N}\Big(\frac{N}{2} - k\Big)$. Yet, the J-DHT, unlike all prior frequency transforms, induces a nonzero SM for any value of ω_o.



(a)



(b)

Figure 2.1: Block averaged SM at bin $k = 6$ and $N = 16$ vs. ω_o for DFT, DHT, J-DHT, and D-DHT at: (a) L=1, and (b) $L = 20$ when $\phi = \frac{\pi}{4}$.

Similar to the D-DHT, the J-DHT solves the singularity problem by spreading the signal energy. The difference is that the D-DHT only spreads the signal energy across two bins as opposed to the J-DHT which spreads the signal energy across the entire spectrum. This wider spreading leads to further reduced SM, which indicates that the J-DHT would need more block averaging to match DFT's detection performance than the D-DHT. Although it also motivates an even simpler detector using single-frequency-bin computation, such an approach is at the expense of detection accuracy and leads to a much more complex estimator.

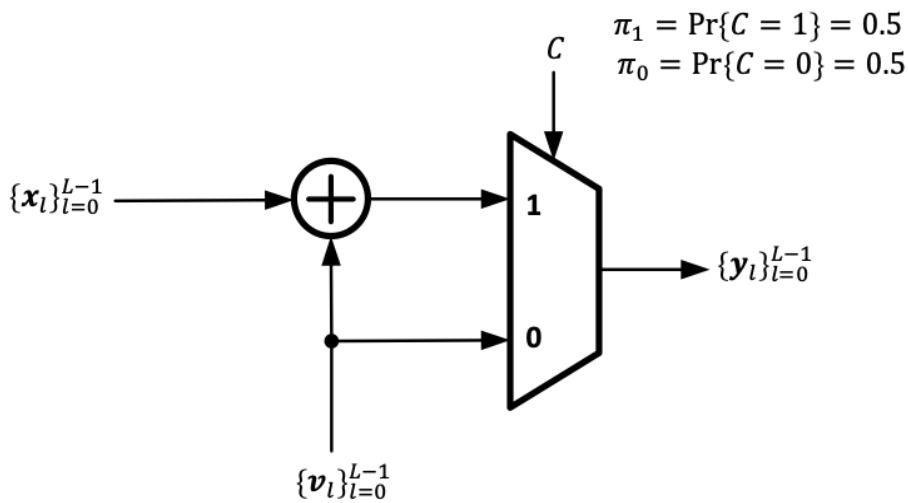


Figure 2.2: Data generation setup.

2.6 Simulation Results

In this section, we compare the detection performance of the DHT, D-DHT, and J-DHT to that of the DFT as a function of input SNR. We show that the singularity problem occurs in the DHT but neither D-DHT nor J-DHT. Furthermore, we quantify the minimum number of blocks L needed to obtain detection accuracy approaching that of the DFT.

2.6.1 Data Generation

We generate the input data for $N_d = 20000$ decisions using the BHT setup shown in Fig. 2.2. Each decision is obtained using L blocks of observations $\{\boldsymbol{y}_l\}_{l=0}^{L-1}$, each of size $N = 16$. Each block of observation $\boldsymbol{y}_l, l \in [L]$ is given by:

$$\boldsymbol{y}_l = \begin{cases} \boldsymbol{x}_l + \boldsymbol{v}_l, & \text{if } C = 1 \\ \boldsymbol{v}_l, & \text{if } C = 0 \end{cases} \tag{2.19}$$

where $C \sim Ber(0.5)$ is a Bernoulli random variable with parameter 0.5 that selects $\boldsymbol{y}_l$ to either be a pure noise block $\boldsymbol{v}_l \sim \mathcal{N}(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_N)$, where $\boldsymbol{I}_N$ is the $N \times N$ identity matrix and $\mathcal{N}(\boldsymbol{m}, \sigma^2 \boldsymbol{I}_N)$ is the independent multivariate Gaussian distribution with mean $\boldsymbol{m}$ and variance σ^2, or a superposition of $\boldsymbol{v}_l$ and a sinusoidal block $\boldsymbol{x}_l$, where $\boldsymbol{x}_l = [x_{0,l}, x_{1,l}, \cdots, x_{N-1,l}]^T$ has amplitude A, frequency ω_o, and phase ϕ. The variance σ^2 is swept to obtain detection

performance as a function of input SNR.

2.6.2 Evaluation Metric

The detection decision $\hat{H}$ is given by:

$$\hat{H} = \begin{cases} H_1 & \text{if } \max_{k\in[\frac{N}{2}]} \frac{1}{L}\sum_{l=0}^{L-1} |F_{k,l}|^2 \geq \tau \\ H_0 & \text{if } \max_{k\in[\frac{N}{2}]} \frac{1}{L}\sum_{l=0}^{L-1} |F_{k,l}|^2 < \tau \end{cases} \tag{2.20}$$

where $F_{k,l}$ is the frequency sample at bin k obtained using input $\boldsymbol{y}_l$ via one of the four frequency transforms discussed so far: DFT, DHT, D-DHT, or J-DHT, and τ is the detection threshold. Table 2.1 summarizes four probability metrics used to evaluate $\hat{H}$. We are interested in two of them: the true positive rate p_{tp} and the false positive rate p_{fp}. The value of p_{fp} determines the value of the threshold. Knowing p_{tp} and p_{fp} is sufficient as we can deduce the other two given p_{tp} and p_{fp}. The false positive rate p_{fp} is extremely important in the context of this thesis since it is not desired to activate the fine estimator when there is no signal. Hence, as the evaluation

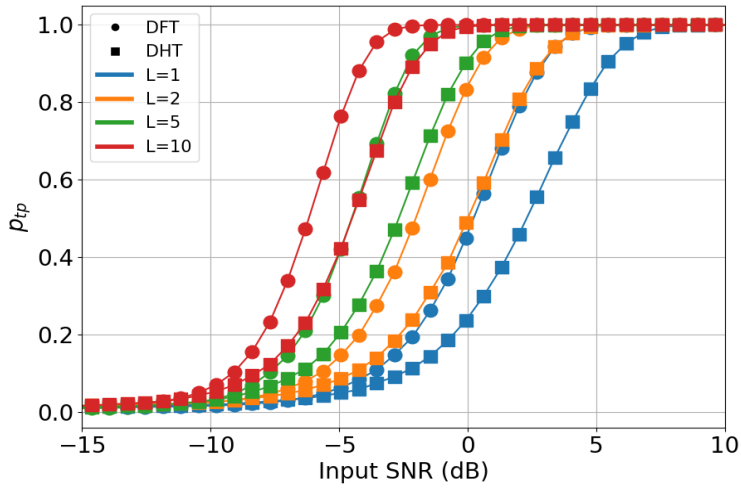
Table 2.1: Four types of probability metrics for BHT.

	H_0	H_1
$\hat{H} = H_0$	true positive (p_{tp})	false negative (p_{fn})
$\hat{H} = H_1$	false positive (p_{fp})	true negative (p_{tn})

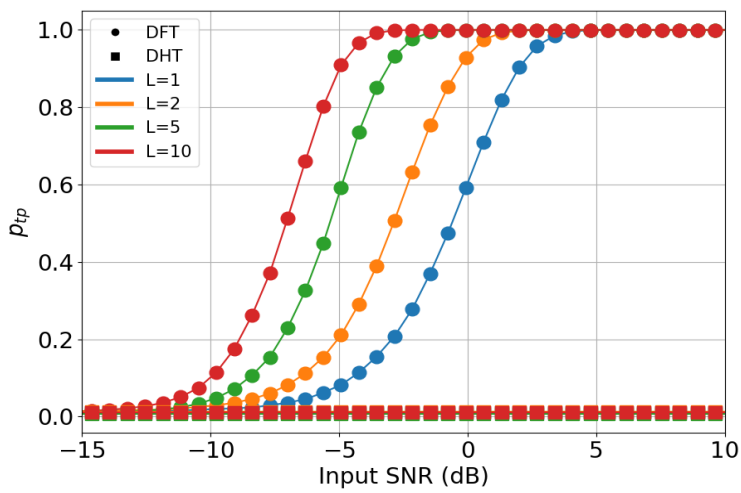
metric, we plot the true positive rate p_{tp} vs. input SNR subject to a low false positive rate ($p_{fp} = 0.01$) for different values of L. We assume the subsequent estimator is no more than $100\times$ computationally complex than the frequency transformer in terms of the number of required MACs. There are two types of ω_o that we are interested in: $\omega_o = \frac{2\pi k_o}{N}$ with $k_o \in \mathbb{Z}$ and $k_o \notin \mathbb{Z}$.

2.6.3 Detection Performance of DHT vs. DFT

Figure 2.3(a) shows that both the DHT and the DFT have similar accuracies (p_{tp}) for input SNR ≥ 10 dB. For low SNR (< 10 dB), the DHT needs block averaging to match the DFT. Figure 2.3(a) reveals that DHT($L = 10$) (DHT with 10-block averaging) achieves the same p_{tp} as DFT($L = 5$). As a result,



(a)

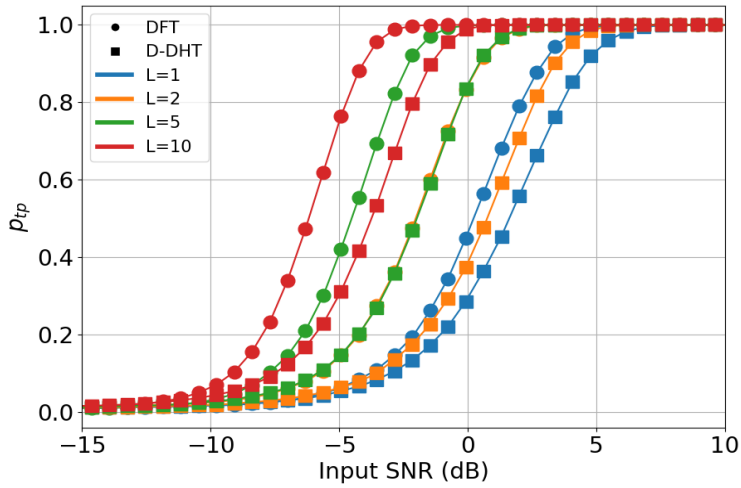


(b)

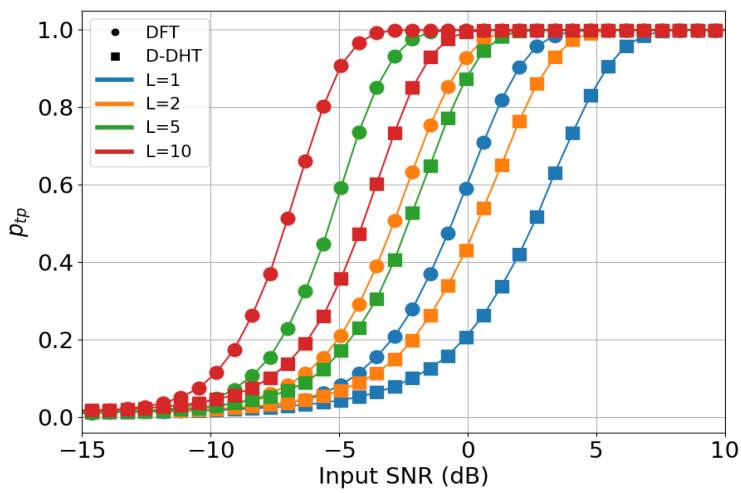
Figure 2.3: True positive rate of DHT and DFT vs. input SNR for different values of L at $\phi = \frac{\pi}{4}$ when: (a) $k_o = 3.25$, and (b) $k_o = 3$.

the DHT($L = 10$) needs to process 2× more time samples and requires the same number of MACs compared to DFT($L = 5$). Yet, this trade-off is acceptable since the MAC unit in the DHT is half as complex as the one used in computing the DFT.

Figure 2.3(b) demonstrates the singularity problem as the DHT fails to detect the input sinusoid that matches the frequency bins. When this failure occurs, the detector cannot distinguish the signal from the noise. As a result, p_{tp} is the same as p_{fp} threshold regardless of the input SNR as shown in



(a)



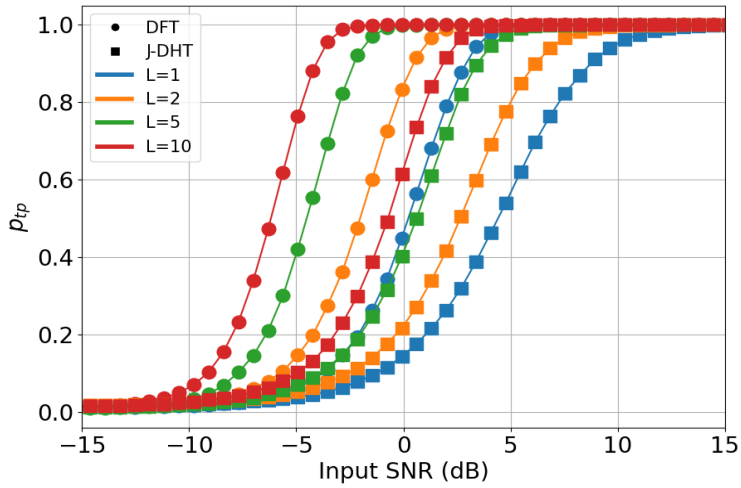
(b)

Figure 2.4: True positive rate of D-DHT and DFT vs. input SNR for different values of L at $\phi = \frac{\pi}{4}$ when: (a) $k_o = 3.25$, and (b) $k_o = 3$.

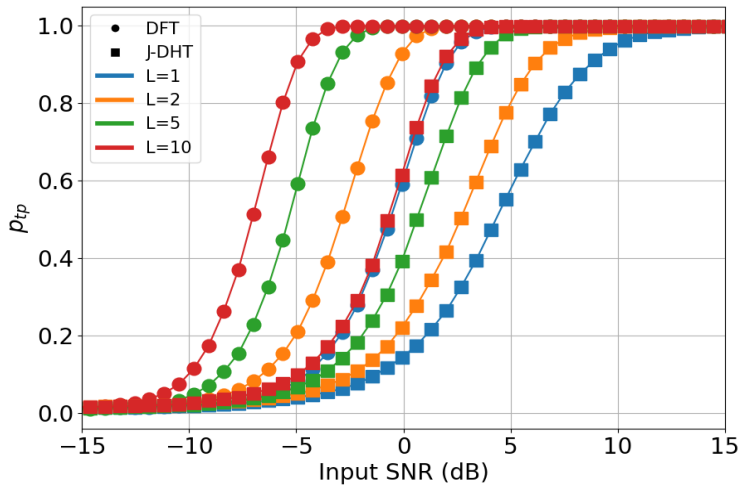
Fig. 2.3(b), which motivates the D-DHT and the J-DHT as solutions to this problem.

2.6.4 Detection Performance of D-DHT vs. DFT

Figure 2.4(a) compares p_{tp} between the D-DHT and the DFT for $k_o = 3.25$. It shows that D-DHT($L = 5$) matches the p_{tp} for DFT($L = 2$). Both trans-



(a)



(b)

Figure 2.5: True positive rate of J-DHT and DFT vs. input SNR of different values for L at $\phi = \frac{\pi}{4}$ when: (a) $k_o = 3.25$, and (b) $k_o = 3$.

forms result in true positive rates close to unity for input SNR $\geq$ 10 dB. Similar to the DHT, the D-DHT needs to process 2.5× more time samples and requires 1.25× more MACs compared to the DFT. In addition, Figure 2.4(b) demonstrates that it does not suffer from the singularity problem. Thus, the D-DHT is a robust option compared to the DHT.

2.6.5 Detection Performance of J-DHT vs. DFT

Finally, we evaluate J-DHT's detection performance relative to the DFT. Although Fig. 2.5(b) shows it solves the singularity problem, Fig. 2.5(a) reveals that the J-DHT needs to average at least five blocks in order to match DFT($L = 1$). Therefore, the J-DHT needs to process at least 5× more times samples and requires at minimum 2.5× more MACs to match DFT's detection performance. This observation corroborates our analysis in Section 2.5.

2.7 Summary

In this chapter, we analyzed the detection performance of the DHT relative to the DFT. By analyzing its SM, we demonstrated DHT's inability to detect input frequencies at the center of frequency bins, referred to as the singularity problem. Thus, we proposed two DHT's variants: the D-DHT and the J-DHT to address this problem. Both variants solve the singularity problem by spreading the input signal energy across the signal spectrum. Although they both are able to match DFT's detection accuracy via block averaging, the J-DHT requires 2× more blocks to achieve DFT's detection accuracy compared to the D-DHT. Thus, among the DFT, DHT, D-DHT, and J-DHT, the D-DHT is the most promising candidate for realizing an area-constrained frequency transformer.

CHAPTER 3

FIXED-POINT ANALYSIS ON DIRECT DOT-PRODUCT ARCHITECTURE

In Chapter 2, we analyzed the detection performance of DHT, D-DHT and J-DHT relative to the DFT under a FL implementation. The D-DHT was found more suitable than the J-DHT as a replacement for the DFT, as the D-DHT needs to observe 2× less data and requires 2× fewer MACs compared to the J-DHT in order to match DFT's detection performance.

In this chapter, we evaluate the impact of quantization on the D-DHT using FX analysis. We compare its precision requirement to that of the DFT. Using the area parameters of a 1-bit register and 1-bit FA in 65 nm technology, we estimate the area cost of the D-DHT and the DFT on a direct dot-product architecture and show that the D-DHT achieves $\sim 2\times$ reduction in terms of area complexity. Yet, the area estimation also reveals a gap to the area constraint of small pixels. Therefore, we propose techniques to close this gap at the end of this chapter.

3.1 Direct Dot-Product Architecture

A DFT frequency sample can be computed with the direct dot-product (direct-DP) architecture shown in Fig. 3.1. In the following, we use the notation x_n instead of y_n to refer to the ROIC output. First, the ROIC output $x(t)$ is sampled every period T to obtain x_n with precision B_x. Next, x_n is multiplied with both the real and the imaginary parts of the weight $w_{k,n}^{(F)} = \frac{1}{N} e^{\frac{-j2\pi kn}{N}}$, respectively, each with precision B_w. The frequency sample X_k is given by:

$$X_k = \sum_{n=0}^{N-1} w_{k,n}^{(F)} x_n \tag{3.1}$$

where N is the DP dimension. Both accumulators in the two branches of the direct-DP architecture in Fig. 3.1 have precision of B_a bits.

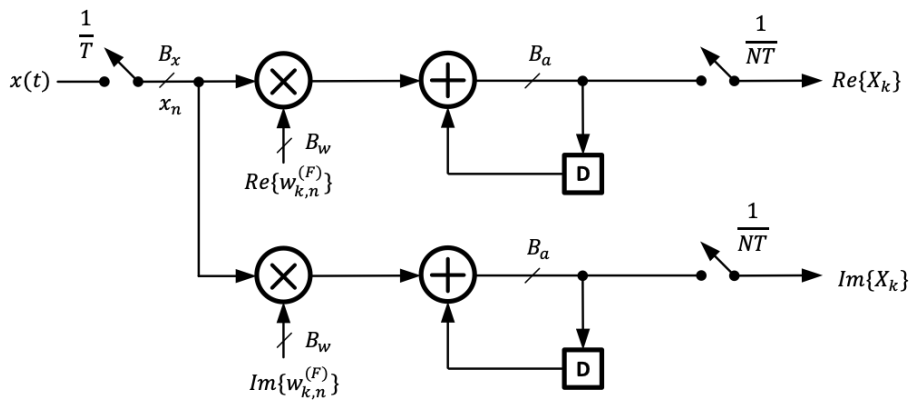


Figure 3.1: Direct-DP architecture for the DFT.

The computation of the D-DHT frequency samples, on the other hand, requires only one DP branch (see Fig. 3.2). A B_x bit input x_n is obtained by sampling $x(t)$ with period T. The signal x_n is multiplied with weight $w_{k,n}^{(D)} = \frac{\sqrt{2}}{N}\cos\left(\frac{2\pi kn}{N} - \frac{\pi}{4} + \phi_n\right)$ that has precision B_w, where $\phi_n = \frac{\pi}{4}\left(1 - (-1)^n\right)$. The frequency sample D_k is computed as follows:

$$D_k = \sum_{n=0}^{N-1} w_{k,n}^{(D)} x_n \tag{3.2}$$

Both multiplication and addition required for (3.2) are real-valued, and the output D_k is quantized to B_a bits.

The direct-DP architecture is suitable for area-constrained platforms due to the following three advantages. First, the computation begins as soon as the first time sample x_0 is ready. The second advantage is its flexibility. By adjusting the mechanism that externally generates the weights, the direct-DP architecture can be used for any frequency transform as long as it can be formulated as a DP. Finally, this architecture minimizes the storage requirement for intermediate values. Figure 3.2 shows that only the partial sum needs to be stored during computation.

To reduce the area complexity of the frequency transformer, the values of B_x, B_w, and B_a need to be minimized while maintaining sufficient detection SNR. Therefore, in this chapter, we perform a SNR vs. precision analysis for FX implementations of the DFT and the D-DHT.

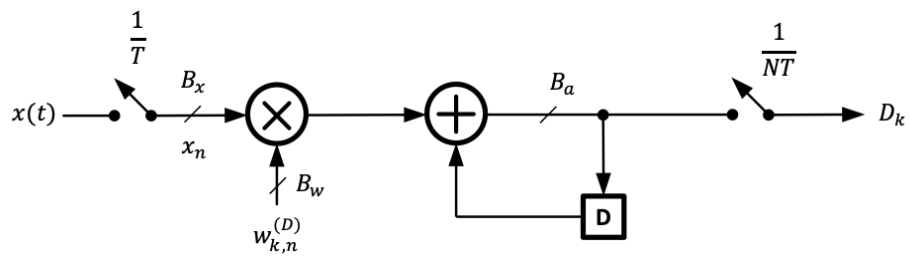


Figure 3.2: Direct-DP architecture for the D-DHT.

3.2 Fixed-Point Analysis

3.2.1 Quantization Noise Model

We assume an additive quantization noise model. For a real-valued signed random signal $x \in [-x_m, x_m]$, where x_m is the maximum absolute value of x, quantized to B bits, the quantized signal x_q is expressed as $x_q = x + q_x$, where q_x is the quantization noise term assumed to be independent of x and uniformly distributed over the range $[-\frac{\Delta_x}{2}, \frac{\Delta_x}{2}]$, where $\Delta_x = x_m 2^{1-B}$. The variance of the quantization noise term q_x is given by $Var(x_q) = \frac{\Delta_x^2}{12}$.

3.2.2 Fixed-Point Dot Product

Consider an ideal FL frequency transform with real-valued weights defined as:

$$F_k = \boldsymbol{w}_k^T \boldsymbol{x} = \sum_{n=0}^{N-1} w_{k,n} x_n \tag{3.3}$$

where F_k is the frequency sample at frequency bin k obtained by computing the DP between the weight vector $\boldsymbol{w}_k = [w_{k,0}, w_{k,1}, \ldots, w_{k,N-1}]^T$ and the input vector $\boldsymbol{x} = [x_0, x_1, \ldots, x_{N-1}]^T$.

The FX realization of the DP in (3.3), with input, weight, and partial sum precisions B_x, B_w, and B_a, respectively, is given by:

$$\begin{aligned} F_k^{(q)} &= q_k^{(F)} + (\boldsymbol{w}_k + \boldsymbol{\Delta w}_k)^T (\boldsymbol{x} + \boldsymbol{q}_x) \\ &\approx \boldsymbol{w}_k^T \boldsymbol{x} + \boldsymbol{w}_k^T \boldsymbol{q}_x + \boldsymbol{\Delta w}_k^T \boldsymbol{x} + q_k^{(F)} \\ &= F_k + q_k^{(i)} + q_k^{(w)} + q_k^{(F)} \end{aligned} \tag{3.4}$$

where $\boldsymbol{\Delta w}_k$ is the deterministic quantization noise vector of the weight vector

$\boldsymbol{w}_k$, $\boldsymbol{q}_x$ is the quantization noise vector of the input vector $\boldsymbol{x}$, $\boldsymbol{w}_k^{(q)} = \boldsymbol{w}_k + \boldsymbol{\Delta w}_k$ and $\boldsymbol{x}_q = \boldsymbol{x} + \boldsymbol{q}_x$ are the quantized weight and input quantized to B_w and B_x bits, respectively, $F_k^{(q)}$ is the FX DP with a precision of B_a bits, $q_k^{(i)}$ and $q_k^{(w)}$ are the input and weight quantization noise seen at the output F_k, respectively, and $q_k^{(F)}$ is the total round-off noise introduced by quantizing the partial sum. Note that $\boldsymbol{\Delta w}_k$ is a deterministic vector since the weight vectors of both the D-DHT and the DFT are deterministic.

3.2.3 Fixed-Point Analysis for D-DHT

The next step in FX analysis involves finding the variances of signals defined in (3.4). For D-DHT, those variances are given by:

$$\begin{aligned} \sigma_{F_k}^2 &= \left(\boldsymbol{w}_k^{(D)}\right)^T \boldsymbol{R}_{xx} \left(\boldsymbol{w}_k^{(D)}\right);\ \sigma_a^2 = \frac{N\Delta_{F_k}^2}{12}; \\ \sigma_x^2 &= \frac{\Delta_i^2}{12}\left\|\boldsymbol{w}_k^{(D)}\right\|_2^2;\ \sigma_w^2 = \left(\boldsymbol{\Delta w}_k^{(D)}\right)^T \boldsymbol{R}_{xx} \left(\boldsymbol{\Delta w}_k^{(D)}\right) \end{aligned} \tag{3.5}$$

where $\sigma_{F_k}^2$, σ_x^2, σ_w^2, and σ_a^2 are the noise variances for F_k, $q_k^{(i)}$, and $q_k^{(w)}$, $q_k^{(F)}$ defined in (3.4), respectively, $\boldsymbol{w}_k^{(D)}$ is the D-DHT weight vector, $\boldsymbol{R}_{xx} = \mathbb{E}[\boldsymbol{x}\boldsymbol{x}^T]$ is the input covariance matrix, $\Delta_i = x_{\max} 2^{1-B_x}$ and $\Delta_{F_k} = F_{k,\max} 2^{1-B_a}$ are the quantization step-sizes for input and partial sum, respectively, with $x_{\max}$ and $F_{k,\max}$ being the maximum absolute value of $\boldsymbol{x}$ and F_k, respectively.

3.2.4 Fixed-Point Analysis for DFT

Equation (3.4) can also be used to analyze FX frequency transforms with complex-valued weights. In such a setup, vector transpose operators are replaced with the Hermitian operators. For DFT, we obtain:

$$\begin{aligned} \sigma_{F_k}^2 &= \left(\boldsymbol{w}_k^{(F)}\right)^H \boldsymbol{R}_{xx} \left(\boldsymbol{w}_k^{(F)}\right);\ \sigma_a^2 = \frac{N\Delta_{F_k}^2}{6}; \\ \sigma_x^2 &= \frac{\Delta_i^2}{12}\|\boldsymbol{w}_k\|_2^2;\ \sigma_w^2 = \left(\boldsymbol{\Delta w}_k^{(F)}\right)^H \boldsymbol{R}_{xx} \left(\boldsymbol{\Delta w}_k^{(F)}\right) \end{aligned} \tag{3.6}$$

where $\boldsymbol{w}_k^{(F)}$ is the DFT weight vector. Note that the quantization noise term for the partial sum is twice as large as that of the D-DHT due to the complex-valued nature of the DFT.

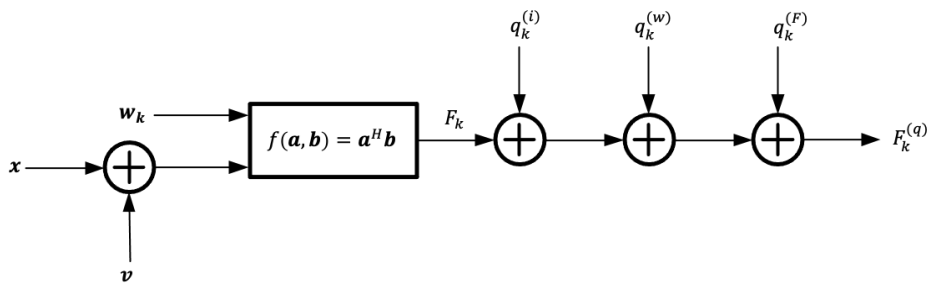


Figure 3.3: Unified quantization noise model for DFT and D-DHT. The conjugate transpose operator is used to accommodate complex-valued weights.

3.2.5 SNR Metrics for D-DHT and DFT

Given an input $\boldsymbol{x} + \boldsymbol{v}$, where $\boldsymbol{x}$ is a sinusoidal input vector with amplitude A, frequency $\omega_o = \frac{2\pi k_o}{N}$, and phase ϕ, and $\boldsymbol{v}$ is the input noise vector, the following describes a unified quantization noise model that applies to both half-band D-DHT and DFT, as shown in Fig. 3.3:

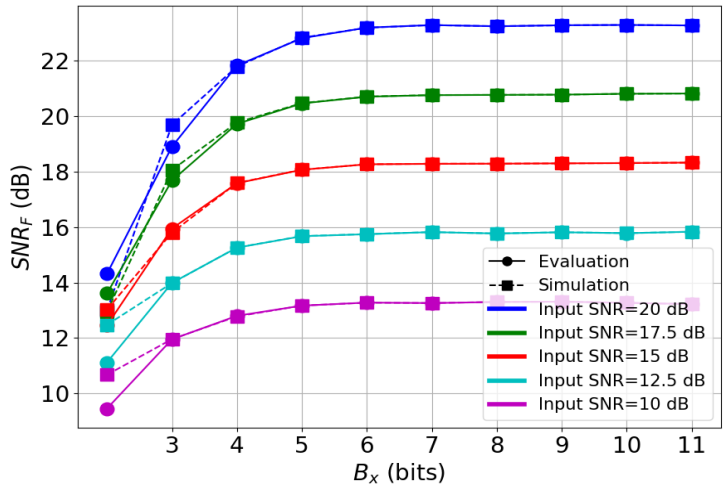
$$F_k^{(q)} = F_k + \eta_k + q_k^{(i)} + q_k^{(w)} + q_k^{(F)}, k \in \left[\frac{N}{2}\right] \tag{3.7}$$

where F_k is the ideal frequency sample, η_k is the output-referred input noise contributed by $\boldsymbol{v}$, $\boldsymbol{w}_k$ is the weight vector for bin k, and $q_k^{(i)}$, $q_k^{(w)}$, and $q_k^{(F)}$ are the output-referred quantization noise terms corresponding to input, weight and partial-sum quantization, respectively.

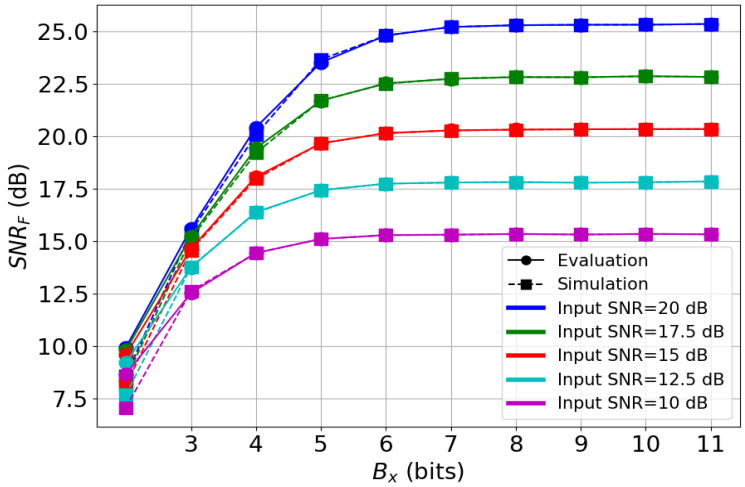
We use the SNR in the frequency domain $\mathrm{SNR_F}$ as the SNR metric to compare DFT and D-DHT's precision requirements. Specifically, $\mathrm{SNR_F}$ is defined as:

$$\mathrm{SNR_F} = \frac{|F_{k_r}|^2}{\frac{2}{N}\sum_{k=0}^{\frac{N}{2}-1}\left(\sigma_a^2 + \sigma_w^2 + \sigma_x^2 + \sigma_{\eta_k}^2\right)} \tag{3.8}$$

where $|F_{k_r}|^2$ is the SM of either the DFT or the D-DHT at bin $k_r = \mathrm{round}(k_o)$, $\sigma_{\eta_k}^2 = \boldsymbol{w}_k{}^H \boldsymbol{R}_{vv} \boldsymbol{w}_k$ with $\boldsymbol{R}_{vv}$ being the covariance matrix of $\boldsymbol{v}$, and σ_x^2, σ_w^2, and σ_a^2 are the output-referred quantization noise variances corresponding to input, weight and partial-sum quantization, respectively. These quantization noise variances are calculated according to (3.5) for the D-DHT and (3.6) for the DFT. The SNR in (3.8) evaluates the ratio of the signal power to the average noise variance across all frequency bins. $\mathrm{SNR_F}$ achieves maximum $\mathrm{SNR_{F,\,max}} = \frac{|F_{k_r}|^2}{\frac{2}{N}\sum_{k=0}^{\frac{N}{2}-1}\sigma_{\eta_k}^2}$ when no quantization noise source is present, i.e., a



(a)



(b)

Figure 3.4: $\mathrm{SNR_F}$ vs. B_x for: (a) D-DHT, and (b) DFT at different input SNR using both sample-accurate simulation and evaluation of FX model in Section 3.2.

FL implementation.

3.3 Simulation Results

In this section, we validate the FX models above and find minimal precision requirements of the DFT and the D-DHT in order to compare their complex-

ities. To do so, we plot $\mathrm{SNR_F}$ vs. B_x, B_w, and B_a at different input SNR separately in order to isolate the impact of quantization noise sources.

For each quantization noise source, we compare the numerically calculated $\mathrm{SNR_F}$ obtained via sample-accurate Python simulations to the one obtained by evaluating the analytical expressions in Section 3.2. The goal is to find the minimum values for B_x, B_w, and B_a such that $\mathrm{SNR_F}$ is close to $\mathrm{SNR_{F,\,max}}$. In the remainder of this chapter, we fix N to 16. We also choose the input SNR to be at least 10 dB, which was shown, in Chapter 2, to enable a detection rate close to unity for both the D-DHT and the DFT without block averaging.

3.3.1 Input Quantization

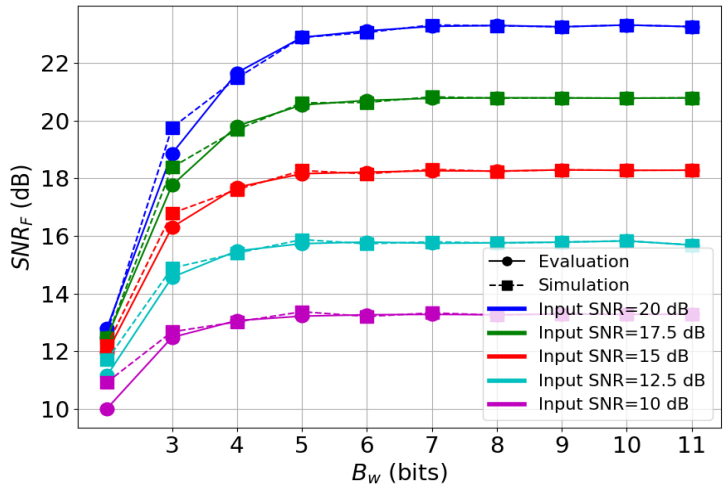
We first compare the impact of input quantization on $\mathrm{SNR_F}$ of the D-DHT and the DFT. Figure 3.4 shows that $\mathrm{SNR_F}$ increases with increasing B_x and saturates at the $\mathrm{SNR_{F,\,max}}$, consistent with the prediction of the FX DP model. The $\mathrm{SNR_{F,\,max}}$ of the DFT is higher than that of the D-DHT as expected according to the analysis in Chapter 2. To achieve an $\mathrm{SNR_F}$ within 1 dB of $\mathrm{SNR_{F,\,max}}$, the D-DHT requires a minimum B_x of 5 bits while the DFT requires a minimum B_x of 6 bits.

3.3.2 Weight Quantization

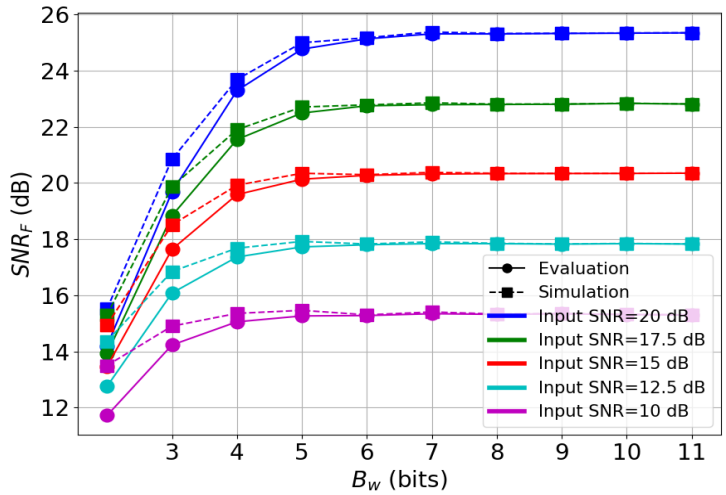
Next, we evaluate the impact of weight quantization on the D-DHT and the DFT. Figure 3.5 shows that both transforms require a minimum of 5 bits for B_w in order to maintain an $\mathrm{SNR_F}$ within 1 dB of $\mathrm{SNR_{F,\,max}}$.

3.3.3 Partial Sum Quantization

Finally, we evaluate the impact of quantizing the partial sum on the D-DHT and the DFT in Fig. 3.6. To prevent an 1 dB drop with respect to $\mathrm{SNR_{F,\,max}}$, Figure 3.6 reveals both transforms require 8 bits.



(a)



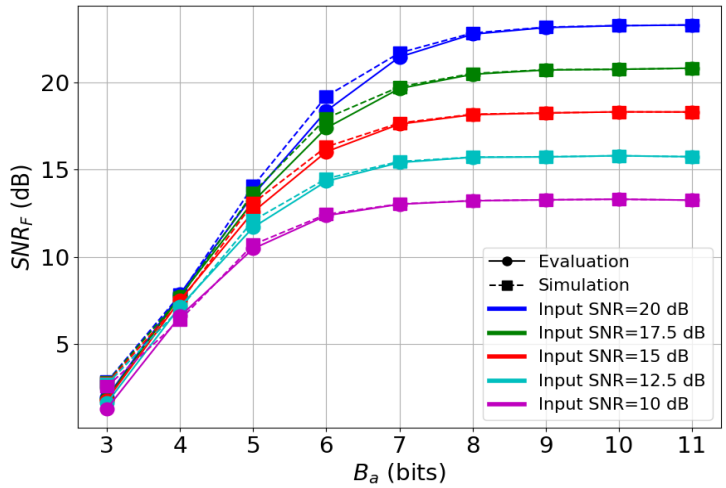
(b)

Figure 3.5: SNR_F vs. B_w for: (a) D-DHT, and (b) DFT at different input SNR using both sample-accurate simulation and evaluation of FX model in Section 3.2.

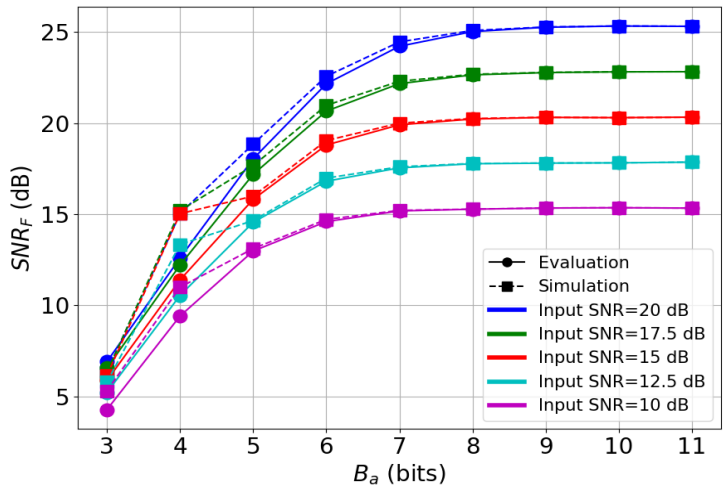
3.3.4 Full FX Model

Given the above simulation results, we quantize all signals: input, weight, and partial sum, and find the (B_x, B_w, B_a) triplets which lead to an SNR_F within 1 dB of its FL baseline for the D-DHT and the DFT, respectively.

Table 3.1 shows the reported SNR_F and the associated precisions B_x, B_w, B_a values. Reported results are for different input SNR obtained from evaluation



(a)



(b)

Figure 3.6: $\mathrm{SNR_F}$ vs. B_a for: (a) D-DHT, and (b) DFT at different input SNR using both sample-accurate simulation and evaluation of FX model in Section 3.2.

of the FX model in Section 3.2 and sample-accurate Python simulations. The table also shows the corresponding $\mathrm{SNR_F}$ under FL implementation. Consistent with prior simulation results, the DFT and the D-DHT require the same number of bits for weight and partial sum, but the DFT requires one more bit for the input. Next, we study the area complexity of the frequency transformer based on the results of our FX analysis.

Table 3.1: Minimum required B_x, B_w, B_a for D-DHT and DFT so that $\mathrm{SNR_F}$ is within 1 dB of its FL $\mathrm{SNR_F}$.

Frequency Transform	Input SNR (dB)	B_x	B_w	B_a	Evaluated $\mathrm{SNR_F}$ (dB)	Simulated $\mathrm{SNR_F}$ (dB)	FL $\mathrm{SNR_F}$ (dB)
D-DHT	20	5	5	8	22.34	22.36	23.28
DFT	20	6	5	8	24.43	24.46	25.33
D-DHT	17.5	5	5	8	20.13	20.17	20.76
DFT	17.5	6	5	8	22.08	22.29	22.83
D-DHT	15	5	5	8	17.86	17.84	18.27
DFT	15	6	5	8	19.86	20.23	20.33
D-DHT	12.5	5	5	8	15.50	15.51	15.77
DFT	12.5	6	5	8	17.55	17.73	17.83
D-DHT	10	5	5	8	13.10	13.07	13.29
DFT	10	6	5	8	15.13	15.37	15.34

3.4 Area Complexity Comparison in Fixed Point

Given the minimum required B_x, B_w and B_a, we estimate the area costs for the D-DHT and the DFT using the direct-DP architecture. As both transforms require computation of $\frac{N}{2}$ frequency samples, we only need to compare the area required for a single frequency sample. The required area to compute one DFT or D-DHT frequency sample $\mathcal{A}$ is given as:

$$\mathcal{A} = \mathcal{A}_{\mathcal{R}}\mathcal{C}_{\mathcal{R}} + \mathcal{A}_{\mathcal{C}}\mathcal{C}_{\mathcal{C}} \tag{3.9}$$

where $\mathcal{A}_{\mathcal{R}}$ and $\mathcal{A}_{\mathcal{C}}$ are the area occupied by a 1-bit register and a 1-bit FA, respectively, $\mathcal{C}_{\mathcal{R}}$ is the total number of bits required to store the partial sum, and $\mathcal{C}_{\mathcal{C}}$ is the number of 1-bit FAs required to implement the real-valued operations, i.e., multiplications and additions. The values of $\mathcal{A}_{\mathcal{R}}$ and $\mathcal{A}_{\mathcal{C}}$ are specific to the technology node. In 65 nm technology, $\mathcal{A}_{\mathcal{R}} = \mathcal{A}_{\mathcal{C}} = 9.9\,\mu\mathrm{m}^2$.

We assume that the coefficients are externally generated and the inputs are streamed in. Thus, only the partial sum is needed to be stored. In addition, we also assume the multiplier and the adder are implemented using the bit-parallel architecture. Thus, for D-DHT, $\mathcal{C}_{\mathcal{R}}$ and $\mathcal{C}_{\mathcal{C}}$ are given by:

$$\mathcal{C}_{\mathcal{R}} = B_a;\ \mathcal{C}_{\mathcal{C}} = B_x B_w + (B_x + B_w + \log_2\lceil N\rceil - 1) \tag{3.10}$$

Table 3.2: Associated $\mathcal{C}_{\mathcal{R}}$, $\mathcal{C}_{\mathcal{C}}$, and $\mathcal{A}$ evaluated using precision values from Table 3.1.

Transform	B_x	B_w	B_a	$\mathcal{C}_{\mathcal{R}}$ (bits)	$\mathcal{C}_{\mathcal{C}}$ (#FA)	$\mathcal{A}$ (μm^2)
D-DHT	5	5	8	8	38	455.4
DFT	6	5	8	16	88	1029.6

for DFT, $\mathcal{C}_{\mathcal{R}}$ and $\mathcal{C}_{\mathcal{C}}$ are given as:

$$\mathcal{C}_{\mathcal{R}} = 2B_a;\ \mathcal{C}_{\mathcal{C}} = 2\Big[B_x B_w + (B_x + B_w + \log_2\lceil N\rceil - 1)\Big] \tag{3.11}$$

The additional factor of two accounts for complex-valued arithmetic required by the DFT.

Table 3.2 summarizes the evaluation results of (3.9) using B_x, B_w, B_a values obtained in Section 3.3. Results indicate that the D-DHT achieves $\sim 2\times$ reduction in area complexity compared to the DFT. Therefore, the D-DHT is a more suitable frequency transform algorithm compared to the DFT for pitch-matched frequency transformer in the FPA.

3.5 Meeting the Area Constraint for Small Pixels

According to the estimated area costs in Table 3.2, although the D-DHT requires less area than the DFT, it remains challenging for the direct-DP D-DHT architecture to meet the area constraint of $144\,\mu\text{m}^2$ imposed by the pixel pitch in high-resolution FPAs.

However, since the typical sampling rate of high resolution FPAs ($< 10\,\text{kHz}$ [1]) is at least an order-of-magnitude smaller than the maximum achievable throughput of the direct-DP architecture (a few hundreds of MHz [23–25]), it is possible to employ bit-serial and time-multiplexed architectures to meet the area constraints.

Bit-serial architectures process data one bit at a time requiring 1-bit arithmetic units and some additional storage but operating at a higher clock frequency and smaller area than the bit-parallel architectures assumed in Table 3.2. In spite of its higher clock frequency, bit-serial architectures are slower than their bit-parallel counterparts. Therefore, this is one way to trade-off latency with area complexity.

Another method is to time-multiplex pixel data onto the same hardware architecture. For example, assuming a 8×8-bit MAC frequency of 100 MHz, and $N = 16$, we find that the direct-DP architecture can compute frequency samples at a rate of 6.2 MHz. Such an architecture can cycle through 620 pixels computing one frequency sample for each. The area occupied by 620 pixels may be sufficient to implement the direct-DP architecture for the D-DHT along with a follow-on detector. Note that time-multiplexing can be done on a per-input sample basis as well resulting in a different trade-off between the overall latency and area.

3.6 Summary

In this chapter, we have used FX analysis to compare the precision requirements for the D-DHT and the DFT implemented using the direct-DP architecture. Based on the area parameters for a 1-bit register and 1-bit FA in the 65 nm process, we compare the area requirements and find that the D-DHT needs half the area as the DFT, thus showing D-DHT's advantage in area-constrained frequency detection and estimation. To meet the area constraint posted by high-resolution FPAs, we show that it is desired to explore bit-serial implementation of the multiplier and the adder in the direct-DP architecture. Moreover, it is also desired to time-multiplex the processing of multiple pixel signals using a single D-DHT transformer, since the throughput of the D-DHT engine is likely to far exceed the sampling rate of the input signal.

CHAPTER 4

CONCLUSION AND FUTURE WORK

4.1 Conclusion

Although DFT is the ideal solution to frequency detection and coarse estimation in many applications, it introduces high area cost due to its complex-valued nature, which is unsuitable for area-constrained frequency detection and estimation. As a result, we have analyzed the performance of real-valued Hartley-based transforms (DHT, D-DHT, and J-DHT) in this thesis to explore their possibilities as DFT's replacements. Through our analysis, we show D-DHT's following advantages:

- The D-DHT does *not* suffer from the singularity problem seen in the DHT.
- The D-DHT requires $2\times$ less block averaging than the J-DHT to match DFT's performance.
- Given the minimal precision requirements for the D-DHT and the DFT obtained through FX analysis, the area cost for the D-DHT is estimated to be $\sim 2\times$ less than that of the DFT.

Therefore, compared to DFT, J-DHT, and DHT, the D-DHT is the preferred algorithm for pitched-matched frequency detector used in FPA.

Meeting the area constraint for small pixels remains challenging. There is still a gap given the estimated area cost. As a result, it is suggested to have multiple pixels share a single D-DHT transformer. The number of pixels to be grouped should be based on the difference between the sampling rate of the pixel signal and the throughput of the D-DHT transformer.

4.2 Future Work

4.2.1 Fine Estimator for Dithered Discrete Hartley Transform

As shown in Fig. 1.3, the frequency detector needs a fine estimator to achieve the ML estimates for the amplitude and the frequency. Since the D-DHT spreads the signal energy across two frequency bins, there is an additional frequency ambiguity problem that the fine estimator has to solve. One potential solution is to compute the frequency bin $k_{\max}$ corresponding to the maximum block-averaged SM and send the SMs in bin $k_{\max}$ and bin $\frac{N}{2} - k_{\max}$ to the fine estimator. A drawback of this approach is that it doubles the amount of data that needs to be transferred from the detector to the fine estimator.

4.2.2 Spatial Dithered Discrete Hartley Transform

In high-resolution FPAs, it is highly likely that neighboring pixels will observe sinusoids with the same frequency due to decreasing distances among pixels. Based on this principle, one can have each pixel compute one frequency sample and combine the computation results at the end. In this way, it is possible to implement the MAC operation in the ROIC to reduce the area complexity. The challenge is that the amplitude of the sinusoid may be different in neighboring pixels. Therefore, it needs to be addressed by the subsequent detector and estimator.

4.2.3 Area-constrained Multi-Sinusoid Detection and Estimation

In this thesis, we assume a single-sinusoid joint detection and parameter estimation problem. A more practical scenario is to relax this assumption and simultaneously estimate multiple frequencies. In this case, the DTFT becomes a sub-optimal method. This kind of estimation problem has also been studied extensively [26–29]. Yet these proposed solutions involve huge matrix multiplication and inversion, which leads to significant area complexity. Exploring area-constrained algorithms and architectures for multiple-sinusoid detection is an open problem.

REFERENCES

[1] M. G. Brown, J. Baker, C. Colonero, J. Costa, T. Gardner, M. Kelly, K. Schultz, B. Tyrrell, and J. Wey, "Digital-pixel focal plane array development," in *Quantum Sensing and Nanophotonic Devices VII*, M. Razeghi, R. Sudharsanan, and G. J. Brown, Eds., vol. 7608, International Society for Optics and Photonics. SPIE, 2010, pp. 765 – 774.

[2] T. Martin, R. Brubaker, P. Dixon, M.-A. Gagliardi, and T. Sudol, "640×512 InGaAs focal plane array camera for visible and SWIR imaging," in *Infrared Technology and Applications XXXI*, B. F. Andresen and G. F. Fulop, Eds., vol. 5783, International Society for Optics and Photonics. SPIE, 2005, pp. 12 – 20.

[3] S. Mintenig, I. Int-Veen, M. Löder, S. Primpke, and G. Gerdts, "Identification of microplastic in effluents of waste water treatment plants using focal-plane-array-based micro-Fourier-transform infrared imaging," in *Water Research*, vol. 108, 2017, pp. 365–372.

[4] D. A. Scribner, M. R. Kruer, and J. M. Killiany, "Infrared focal plane array technology," in *Proceedings of the IEEE*, vol. 79, no. 1, 1991, pp. 66–85.

[5] A. Rogalski, "Progress in focal plane array technologies," in *Progress in Quantum Electronics*, vol. 36, no. 2, 2012, pp. 342–473.

[6] R. Domínguez-Castro, S. Espejo, Á. Rodríguez-Vázquez, R. Carmona, P. Földesy, Á. Zarándy, P. Szolgay, T. Szirányi, and T. Roska, "A 0.8-μm CMOS two-dimensional programmable mixed-signal focal-plane array processor with on-chip binary imaging and instructions storage," in *IEEE Journal of Solid-State Circuits*, vol. 32, 1997, pp. 1013–1026.

[7] D. Temple, C. A. Bower, D. Malta, J. E. Robinson, P. R. Coffman, M. R. Skokan, and T. B. Welch, "High density 3-D integration technology for massively parallel signal processing in advanced infrared focal plane array sensors," in *2006 International Electron Devices Meeting*, 2006, pp. 1–4.

[8] C. Keast, B. Aull, J. Burns, C. Chen, J. Knecht, B. Tyrrell, K. Warner, B. Wheeler, V. Suntharaligam, P. Wyatt, and D. Yost, "Three-dimensional integration technology for advanced focal planes," in *MRS Proceedings*, 2007.

[9] C. A. Bower, D. Malta, D. Temple, J. E. Robinson, P. R. Coffinan, M. R. Skokan, and T. B. Welch, "High density vertical interconnects for 3-D integration of silicon integrated circuits," in *56th Electronic Components and Technology Conference*, 2006, pp. 309–403.

[10] F. Forsberg, A. Lapadatu, G. Kittilsland, S. Martinsen, N. Roxhed, A. C. Fischer, G. Stemme, B. Samel, P. Ericsson, N. Høivik, T. Bakke, M. Bring, T. Kvisterøy, A. Rør, and F. Niklaus, "CMOS-integrated Si/SiGe quantum-well infrared microbolometer focal plane arrays manufactured with very large-scale heterogeneous 3-D integration," in *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 21, no. 4, 2015, pp. 30–40.

[11] D. Rife and R. Boorstyn, "Single tone parameter estimation from discrete-time observations," in *IEEE Transactions on Information Theory*, vol. 20, no. 5, 1974, pp. 591–598.

[12] M. D. Macleod, "Fast nearly ML estimation of the parameters of real or complex single tones or resolved multiple tones," in *IEEE Transactions on Signal Processing*, vol. 46, no. 1, 1998, pp. 141–148.

[13] S. M. Kay and J. R. Gabriel, "Optimal invariant detection of a sinusoid with unknown parameters," in *IEEE Transactions on Signal Processing*, vol. 50, no. 1, 2002, pp. 27–40.

[14] A. Polydoros and C. Nikias, "Detection of unknown-frequency sinusoids in noise: spectral versus correlation domain," in *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 35, no. 6, 1987, pp. 897–900.

[15] P. Moulin and V. V. Veeravalli, *Statistical Inference for Engineers and Data Scientists*. Cambridge University Press, 2018.

[16] S. Djukanović and V. Popović-Bugarin, "Efficient and accurate detection and frequency estimation of multiple sinusoids," in *IEEE Access*, vol. 7, 2019, pp. 1118–1125.

[17] B. M. Baas, "A low-power, high-performance, 1024-point FFT processor," in *IEEE Journal of Solid-State Circuits*, vol. 34, no. 3, 1999, pp. 380–387.

[18] W. Chang and T. Q. Nguyen, "On the fixed-point accuracy analysis of FFT algorithms," in *IEEE Transactions on Signal Processing*, vol. 56, no. 10, 2008, pp. 4673–4682.

[19] Tran-Thong and Bede Liu, "Fixed-point fast Fourier transform error analysis," in *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 24, no. 6, 1976, pp. 563–573.

[20] P. Welch, "A fixed-point fast Fourier transform error analysis," in *IEEE Transactions on Audio and Electroacoustics*, vol. 17, no. 2, 1969, pp. 151–157.

[21] O. Sarbishei and K. Radecka, "Analysis of mean-square-error (MSE) for fixed-point FFT units," in *2011 IEEE International Symposium of Circuits and Systems (ISCAS)*, 2011, pp. 1732–1735.

[22] R. V. L. Hartley, "A more symmetrical Fourier analysis applied to transmission problems," in *Proceedings of the IRE*, vol. 30, no. 3, 1942, pp. 144–150.

[23] R. Pawar and S. S. Shriramwar, "Design implementation of area efficient low power high speed MAC unit using FPGA," in *2017 IEEE International Conference on Power, Control, Signals and Instrumentation Engineering (ICPCSI)*, 2017, pp. 2683–2687.

[24] K. Lilly, S. Nagaraj, B. Manvitha, and K. Lekhya, "Analysis of 32-bit multiply and accumulate unit (MAC) using Vedic multiplier," in *2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE)*, 2020, pp. 1–4.

[25] P. Jagadeesh, S. Ravi, and K. H. Mallikarjun, "Design of high performance 64-bit MAC unit," in *2013 International Conference on Circuits, Power and Computing Technologies (ICCPCT)*, 2013, pp. 782–786.

[26] O. Y. Bushuev and O. L. Ibryaeva, "Choosing an optimal sampling rate to improve the performance of signal analysis by Prony's method," in *2012 35th International Conference on Telecommunications and Signal Processing (TSP)*, 2012, pp. 634–638.

[27] M. Ruan, Y. Cheng, T. Zhang, A. Wang, and H. Xue, "Improved Prony's method for high-frequency-resolution harmonic and interharmonic analysis," in *2019 IEEE 2nd International Conference on Electronics Technology (ICET)*, 2019, pp. 585–589.

[28] E. Djermoune and M. Tomczak, "Statistical analysis of the Kumaresan-Tufts and matrix pencil methods in estimating a damped sinusoid," in *2004 12th European Signal Processing Conference*, 2004, pp. 1261–1264.

[29] A. Okhovat and J. R. Cruz, "Statistical analysis of the Tufts-Kumaresan and principal Hankel components methods for estimating damping factors of single complex exponentials," in *International Conference on Acoustics, Speech, and Signal Processing*, 1989, pp. 2286–2289 vol.4.

APPENDIX A

DERIVATIONS FOR ANALYTICAL EXPRESSIONS IN CHAPTER 2

A.1 Preliminaries

Many analytical expressions in Chapter 2 require finding a closed expression for finite trigonometric series. We will give the derivation for the general case here and apply it in latter sections of this appendix.

Consider the following summation of a complex exponential:

$$Z = \sum_{n=0}^{N-1} e^{j(\Omega n+\phi)} \tag{A.1}$$

Evaluating Z is equivalent to evaluating the following two sums:

$$C = \sum_{n=0}^{N-1} \cos(\Omega n + \phi) = Re\{Z\} \tag{A.2}$$

$$S = \sum_{n=0}^{N-1} \sin(\Omega n + \phi) = Im\{Z\} \tag{A.3}$$

Since Z is a geometric series with common ratio $e^{j\Omega}$, we have the following:

$$\begin{aligned} Z &= \sum_{n=0}^{N-1} e^{j(\Omega n+\phi)} \\ &= e^{j\phi}\frac{e^{j\Omega N}-1}{e^{j\Omega}-1} \\ &= e^{j\phi}\frac{e^{j\frac{\Omega N}{2}}}{e^{j\frac{\Omega}{2}}}\frac{e^{j\frac{\Omega N}{2}}-e^{-j\frac{\Omega N}{2}}}{e^{j\frac{\Omega}{2}}-e^{-j\frac{\Omega}{2}}} \\ &= \frac{\sin\left(\frac{N\Omega}{2}\right)}{\sin\left(\frac{\Omega}{2}\right)}e^{j\left(\frac{N-1}{2}\Omega+\phi\right)} \end{aligned} \tag{A.4}$$

Therefore, C and S are given by:

$$C = Re\{Z\} = \frac{\sin\left(\frac{N\Omega}{2}\right)}{\sin\left(\frac{\Omega}{2}\right)} \cos\left(\frac{N-1}{2}\Omega + \phi\right) \tag{A.5}$$

$$S = Im\{Z\} = \frac{\sin\left(\frac{N\Omega}{2}\right)}{\sin\left(\frac{\Omega}{2}\right)} \sin\left(\frac{N-1}{2}\Omega + \phi\right) \tag{A.6}$$

A.2 Derivation for Equation (2.3)

We show the derivation of (2.3). First, we write (2.1) as follows:

$$H_k = \frac{1}{N}\Big[\sum_{n=0}^{N-1} x_n \cos\left(\frac{2\pi kn}{N}\right) + \sum_{n=0}^{N-1} x_n \sin\left(\frac{2\pi kn}{N}\right)\Big] \tag{A.7}$$

Next, we substitute k in (A.7) with $N-k$ to obtain:

$$\begin{aligned} H_{N-k} &= \frac{1}{N}\Big[\sum_{n=0}^{N-1} x_n \cos\left(\frac{2\pi(N-k)n}{N}\right) + \sum_{n=0}^{N-1} x_n \sin\left(\frac{2\pi(N-k)n}{N}\right)\Big] \\ &= \frac{1}{N}\Big[\sum_{n=0}^{N-1} x_n \cos\left(\frac{2\pi kn}{N}\right) - \sum_{n=0}^{N-1} x_n \sin\left(\frac{2\pi kn}{N}\right)\Big] \end{aligned} \tag{A.8}$$

Using (A.7) and (A.8), we compute:

$$\frac{H_{N-k} + H_k}{2} = \frac{1}{N}\sum_{n=0}^{N-1} x_n \cos\left(\frac{2\pi kn}{N}\right) \tag{A.9}$$

$$\frac{H_{N-k} - H_k}{2} = -\frac{1}{N}\sum_{n=0}^{N-1} x_n \sin\left(\frac{2\pi kn}{N}\right) \tag{A.10}$$

Comparing (2.2), (A.9), and (A.10), we can see that (A.9) and (A.10) constitute the real and imaginary parts of X_k in (2.2), respectively. Therefore, we obtain (2.3).

A.3 Derivation for Equation (2.7)

Here we shall give the derivation for (2.7). First, we rewrite $|X_{k,l}|^2$ in terms of the real and imaginary parts of $X_{k,l}$ as follows:

$$|X_{k,l}|^2 = |Re\{X_{k,l}\}|^2 + |Im\{X_{k,l}\}|^2 \tag{A.11}$$

Using (A.9) and (A.10), we can show that $|H_{k,l}|^2$ is given as:

$$|H_{k,l}|^2 = |Re\{X_{k,l}\}|^2 + |Im\{X_{k,l}\}|^2 - 2Re\{X_{k,l}\}Im\{X_{k,l}\} \tag{A.12}$$

Comparing (2.4), (A.11), and (A.12), we can see that $\delta_{k,l}$ in (2.4) is given by:

$$\delta_{k,l} = -2Re\{X_{k,l}\}Im\{X_{k,l}\} \tag{A.13}$$

Next, we derive the closed expression for $Re\{X_{k,l}\}$:

$$\begin{aligned} Re\{X_{k,l}\} &= \frac{1}{N}\sum_{n=0}^{N-1} x_{n,l}\cos\left(\frac{2\pi kn}{N}\right) \\ &= \frac{A}{N}\sum_{n=0}^{N-1}\cos(\omega_o(n+Nl)+\phi)\cos\left(\frac{2\pi kn}{N}\right) \end{aligned} \tag{A.14}$$

Based on the product-to-sum formula $\cos(a)\cos(b) = \frac{1}{2}[\cos(a+b)-\cos(a-b)]$, we have:

$$\begin{aligned} Re\{X_{k,l}\} = \frac{A}{2N}\Big[&\sum_{n=0}^{N-1}\cos(En+\omega_o Nl+\phi) \\ &+\sum_{n=0}^{N-1}\cos(Fn+\omega_o Nl+\phi)\Big] \end{aligned} \tag{A.15}$$

where $E = \omega_o + \frac{2\pi k}{N}$ and $F = \omega_o - \frac{2\pi k}{N}$. Using (A.5), we have:

$$\begin{aligned} Re\{X_{k,l}\} = \frac{A}{2N}\Big[&\frac{\sin(\frac{NE}{2})}{\sin(\frac{E}{2})}\cos\left(\frac{N-1}{2}E+\omega_o Nl+\phi\right) \\ &+\frac{\sin(\frac{NF}{2})}{\sin(\frac{F}{2})}\cos\left(\frac{N-1}{2}F+\omega_o Nl+\phi\right)\Big] \end{aligned} \tag{A.16}$$

Similarly, we repeat this process for the imaginary part $Im\{X_{k,l}\}$:

$$\begin{aligned} Im\{X_{k,l}\} &= -\frac{1}{N}\sum_{n=0}^{N-1} x_{n,l}\sin\Big(\frac{2\pi kn}{N}\Big) \\ &= \frac{A}{N}\sum_{n=0}^{N-1}\cos(\omega_o(n+Nl)+\phi)\sin\Big(\frac{2\pi kn}{N}\Big) \\ &= -\frac{A}{2N}\Big[\sum_{n=0}^{N-1}\sin(En+\omega_o Nl+\phi) \\ &- \sum_{n=0}^{N-1}\sin(Fn+\omega_o Nl+\phi)\Big] \end{aligned} \tag{A.17}$$

Using (A.6), we have:

$$\begin{aligned} Im\{X_{k,l}\} = -\frac{A}{2N}\Big[&\frac{\sin(\frac{NE}{2})}{\sin(\frac{E}{2})}\sin\Big(\frac{N-1}{2}E+\omega_o Nl+\phi\Big) \\ &- \frac{\sin(\frac{NF}{2})}{\sin(\frac{F}{2})}\sin\Big(\frac{N-1}{2}F+\omega_o Nl+\phi\Big)\Big] \end{aligned} \tag{A.18}$$

Finally, we use (A.13), (A.16), and (A.18) to obtain (2.7).

A.4 Derivation for Equations (2.8) and (2.9)

Here we derive (2.8) and (2.9). First, we compute the blocked average deviation after L blocks as follows:

$$\begin{aligned} \Delta_k(L) &= \frac{1}{L}\sum_{l=0}^{L-1}\delta_{k,l} \\ &= \frac{1}{L}\sum_{l=0}^{L-1}\Big[\frac{A^2\sin^2\left(\frac{NE}{2}\right)}{4N^2\sin^2\left(\frac{E}{2}\right)}\sin\Big((N-1)E+2\phi+2\omega_o Nl\Big) \\ &- \frac{A^2\sin^2\left(\frac{NF}{2}\right)}{4N^2\sin^2\left(\frac{F}{2}\right)}\sin\Big((N-1)F+2\phi+2\omega_o Nl\Big) \\ &- \frac{A^2\Big[1-\cos(N\omega_o)\Big]}{2N^2\Big[\cos(\frac{2\pi k}{N})-\cos(\omega_o)\Big]}\sin\Big(\frac{2\pi k}{N}\Big)\Big] \end{aligned} \tag{A.19}$$

where $E = \omega_o + \frac{2\pi k}{N}$ and $F = \omega_o - \frac{2\pi k}{N}$. Using (A.6), we have:

$$\begin{aligned}
\Delta_k(L) = &-\frac{A^2\Big[1-\cos(N\omega_o)\Big]}{2N^2\Big[\cos(\frac{2\pi k}{N})-\cos(\omega_o)\Big]}\sin\Big(\frac{2\pi k}{N}\Big) \\
&+\frac{A^2\sin^2\Big(\frac{NE}{2}\Big)\sin\Big(\omega_o NL\Big)}{4LN^2\sin^2\Big(\frac{E}{2}\Big)\sin(\omega_o N)}\sin\Big((N-1)E+2\phi+\omega_o N(L-1)\Big) \\
&-\frac{A^2\sin^2\Big(\frac{NF}{2}\Big)\sin\Big(\omega_o NL\Big)}{4LN^2\sin^2\Big(\frac{F}{2}\Big)\sin(\omega_o N)}\sin\Big((N-1)F+2\phi+\omega_o N(L-1)\Big)
\end{aligned} \tag{A.20}$$

Next we take the limit of (A.20) as L goes to ∞ to obtain (2.8):

$$\begin{aligned}
\Delta_k(\infty) = \lim_{L\leftarrow\infty}\Delta_k(L) &= \lim_{L\to\infty}\frac{1}{L}\sum_{l=0}^{L-1}\delta_{k,l} \\
&= -\frac{A^2\Big[1-\cos(N\omega_o)\Big]}{2N^2\Big[\cos(\frac{2\pi k}{N})-\cos(\omega_o)\Big]}\sin\Big(\frac{2\pi k}{N}\Big)
\end{aligned} \tag{A.21}$$

The limit is obtained since the limit of the last two terms in (A.20) go to 0 due to $\lim_{L\to\infty}\frac{\sin(\omega_o NL)}{L} = 0$.

When N is large, we use the fact that $\sin\Big(\frac{2\pi k}{N}\Big) \approx \frac{2\pi k}{N}$ and $\cos\Big(\frac{2\pi k}{N}\Big) \approx 1$ to arrive at (2.9):

$$\begin{aligned}
|\Delta_k(\infty)| &\approx \frac{A^2\Big[1-\cos(N\omega_o)\Big]}{2N^2\Big[1-\cos(\omega_o)\Big]}\frac{2\pi k}{N} \\
&\leq \frac{A^2N^2}{2N^2}\frac{2\pi k}{N} \\
&= \frac{A^2\pi k}{N}
\end{aligned} \tag{A.22}$$

The maximum of $\frac{1-\cos(N\omega_o)}{1-\cos(\omega_o)}$ is obtained using L'Hospital's rule.

A.5 Derivation for Equations (2.10), (2.11), and (2.12)

Here we derive (2.10), (2.11), and (2.12) when the singularity problem occurs. First, we find a closed expression for $Re\{X_{k,l}\}$ and $Im\{X_{k,l}\}$ in the case of $\omega_o = \frac{2\pi k_o}{N}$, where $k_o = k \in \mathbb{Z}$. Based on (A.15), we have:

$$\begin{aligned} Re\{X_{k,l}\} &= \frac{A}{2N}\Big[\sum_{n=0}^{N-1}\cos\Big(\frac{4\pi k}{N}n + \omega_o N l + \phi\Big) + \sum_{n=0}^{N-1}\cos(\phi)\Big] \\ &= \frac{A}{2N}\sum_{n=0}^{N-1}\cos(\phi) \\ &= \frac{A}{2}\cos(\phi) \end{aligned} \tag{A.23}$$

Based on (A.17), we have:

$$\begin{aligned} Im\{X_{k,l}\} &= -\frac{A}{2N}\Big[\sum_{n=0}^{N-1}\sin\Big(\frac{4\pi k}{N}n + \omega_o N l + \phi\Big) - \sum_{n=0}^{N-1}\sin(\phi)\Big] \\ &= \frac{A}{2N}\sum_{n=0}^{N-1}\sin(\phi) \\ &= \frac{A}{2}\sin(\phi) \end{aligned} \tag{A.24}$$

Next, we compute $\delta_{k,l}$ using (A.13) to obtain (2.10):

$$\delta_{k,l} = -2 \cdot \frac{A}{2}\sin(\phi) \cdot \frac{A}{2}\cos(\phi) = -\frac{A^2}{4}\sin(2\phi) \tag{A.25}$$

Using (A.12), we can also compute $|H_{k,l}|^2$ to obtain (2.12):

$$\begin{aligned} |H_{k,l}|^2 &= |Re\{X_{k,l}\}|^2 + |Im\{X_{k,l}\}|^2 - 2Re\{X_{k,l}\}Im\{X_{k,l}\} \\ &= \frac{A^2}{4}\sin^2(\phi) + \frac{A^2}{4}\cos^2(\phi) - \frac{A^2}{4}\sin(2\phi) \\ &= \frac{A^2}{4}\Big[1 - \sin(2\phi)\Big] \end{aligned} \tag{A.26}$$

Equation (2.11) can be obtained using (A.25) and (A.26).

A.6 Derivation for Equations (2.15) and (2.16)

Here we shall derive (2.15) and (2.16) from (2.14). First, we assume $\omega_o = \frac{2\pi k}{N}$ and $k \neq \frac{N}{4}$ while evaluating $C_{k,l}$ and $S_{k,l}$:

$$\begin{aligned}
C_{k,l} &= \sum_{n=0}^{\frac{N}{2}-1} A\cos(\omega_o(2n) + \phi + \omega_o N l)\cos\Big(\frac{2\pi k(2n)}{N} - \frac{\pi}{4}\Big) \\
&= \frac{A}{2}\Big[\sum_{n=0}^{\frac{N}{2}-1}\cos\Big(\frac{8\pi k}{N}n + \phi + \omega_o N l - \frac{\pi}{4}\Big) + \sum_{n=0}^{\frac{N}{2}-1}\cos\Big(\phi + \omega_o N l + \frac{\pi}{4}\Big)\Big] \\
&= \frac{AN}{4}\cos\Big(\phi + \frac{\pi}{4}\Big)
\end{aligned} \tag{A.27}$$

$$\begin{aligned}
S_{k,l} &= -\sum_{n=0}^{\frac{N}{2}-1} x_{2n+1,l}\sin\Big(\frac{2\pi k(2n+1)}{N} - \frac{\pi}{4}\Big) \\
&= -\sum_{n=0}^{\frac{N}{2}-1} A\cos(\omega_o(2n+1) + \phi + \omega_o N l)\sin\Big(\frac{2\pi k(2n+1)}{N} - \frac{\pi}{4}\Big) \\
&= -\frac{A}{2}\Big[\sum_{n=0}^{\frac{N}{2}-1}\sin\Big(\frac{4\pi k}{N}(2n+1) + \phi + \omega_o N l - \frac{\pi}{4}\Big) \\
&- \sum_{n=0}^{\frac{N}{2}-1}\sin\Big(\phi + \omega_o N l + \frac{\pi}{4}\Big)\Big] \\
&= \frac{AN}{4}\sin\Big(\phi + \frac{\pi}{4}\Big)
\end{aligned} \tag{A.28}$$

Next, we evaluate $|D_{k,l}|^2$ to arrive at (2.15):

$$\begin{aligned}
|D_{k,l}|^2 &= \frac{2}{N^2}|C_{-k,l} + S_{k,l}|^2 \\
&= \frac{2}{N^2}\frac{N^2A^2}{16}\Big|\cos\Big(\phi + \frac{\pi}{4}\Big) + \sin\Big(\phi + \frac{\pi}{4}\Big)\Big|^2 \\
&= \frac{2}{N^2}\frac{N^2A^2}{16}\cdot 2\cos^2(\phi) \\
&= \frac{A^2}{4}\frac{1+\cos(2\phi)}{2} \\
&= \frac{A^2}{8}\Big[1 + \cos(2\phi)\Big]
\end{aligned} \tag{A.29}$$

Deriving (2.16) uses a similar strategy but we replace k with $\frac{N}{2} - k$ while evaluating (A.27) and (A.28).

A.7 Derivation for Equation (2.17)

Here we derive (2.17). When $\omega_o = \frac{2\pi}{N}\frac{N}{4} = \frac{\pi}{2}$, i.e, ω_o, matches the bin $k = \frac{N}{4}$, we evaluate $C_{\frac{N}{4},l}$ and $S_{\frac{N}{4},l}$:

$$\begin{aligned}
C_{\frac{N}{4},l} &= \sum_{n=0}^{\frac{N}{2}-1} A\cos(\omega_o(2n) + \phi + \omega_o N l)\cos\left(\frac{2\pi k(2n)}{N} - \frac{\pi}{4}\right) \\
&= \sum_{n=0}^{\frac{N}{2}-1} A\cos(\omega_o(2n) + \phi + \omega_o N l)\cos\left(\pi n - \frac{\pi}{4}\right) \\
&= \sum_{n=0}^{\frac{N}{2}-1} A\cos(\pi n + \phi)\cos\left(\pi n - \frac{\pi}{4}\right) \\
&= \frac{AN}{2}\left[\cos\left(\phi - \frac{\pi}{4}\right) + \cos\left(\phi + \frac{\pi}{4}\right)\right] \\
&= \frac{AN}{2}\cos\left(\frac{\pi}{4}\right)\cos\left(\phi\right)
\end{aligned} \tag{A.30}$$

$$\begin{aligned}
S_{\frac{N}{4},l} &= -\sum_{n=0}^{\frac{N}{2}-1} A\cos(\omega_o(2n+1) + \phi + \omega_o N l)\sin\left(\frac{\pi}{2}(2n+1) - \frac{\pi}{4}\right) \\
&= -\sum_{n=0}^{\frac{N}{2}-1} A\cos(\frac{\pi}{2}(2n+1) + \phi)\sin\left(\frac{\pi}{2}(2n+1) - \frac{\pi}{4}\right) \\
&= \frac{AN}{2}\left[\sin\left(\phi - \frac{\pi}{4}\right) + \sin\left(\phi + \frac{\pi}{4}\right)\right] \\
&= \frac{AN}{2}\sin\left(\frac{\pi}{4}\right)\sin\left(\phi\right)
\end{aligned} \tag{A.31}$$

Using (A.30) and (A.31), we finish the derivation for (2.17):

$$\begin{aligned}
\left|D_{\frac{N}{4},l}\right|^2 &= \frac{2}{N^2}\left|C_{\frac{N}{4},l} + S_{\frac{N}{4},l}\right|^2 \\
&= \frac{2}{N^2}\frac{A^2N^2}{4}\cos^2\left(\phi - \frac{\pi}{4}\right) \\
&= \frac{A^2}{2}\frac{1 + \cos\left(2\phi - \frac{\pi}{2}\right)}{2} \\
&= \frac{A^2}{4}\left[1 + \cos(2\phi - \frac{\pi}{2})\right] \\
&= \frac{A^2}{4}\left[1 + \sin(2\phi)\right]
\end{aligned} \tag{A.32}$$