

**WestminsterResearch**<http://www.westminster.ac.uk/westminsterresearch>**Estimation of Fourier Transform Using Alias-free Hybrid-Stratified Sampling****Tarczyński, A. and Ahmad, B.**

This is a copy of the author's accepted version of a paper subsequently to be published in the *IEEE Transactions on Signal Processing*, 1053-587X . The final definitive version is available from the publisher:

<http://dx.doi.org/10.1109/TSP.2016.2540602>

© 2016 IEEE . Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

---

The WestminsterResearch online digital archive at the University of Westminster aims to make the research output of the University available to a wider audience. Copyright and Moral Rights remain with the authors and/or copyright owners.

---

Whilst further distribution of specific materials from within this archive is forbidden, you may freely distribute the URL of WestminsterResearch: (<http://westminsterresearch.wmin.ac.uk/>).

In case of abuse or copyright appearing without permission e-mail [repository@westminster.ac.uk](mailto:repository@westminster.ac.uk)

# Estimation of Fourier Transform Using Alias-free Hybrid-Stratified Sampling

Andrzej Tarczynski, *Member, IEEE*, and Bashar I. Ahmad, *Member, IEEE*

**Abstract**— This paper proposes a novel method of estimating the Fourier Transform (FT) of deterministic, continuous-time signals, from a finite number  $N$  of their samples taken from a fixed-length observation window. It uses alias-free hybrid-stratified sampling to probe the processed signal at a mixture of deterministic and random time instants. The FT estimator, specifically designed to work with this sampling scheme, is unbiased, consistent and fast converging. It is shown that if the processed signal has continuous third derivative, then the estimator's rate of uniform convergence in mean square is  $N^{-5}$ . Therefore, in terms of frequency-independent upper bounds on the FT estimation error, the proposed approach significantly outperforms existing estimators that utilize alias-free sampling, such as total random, stratified sampling, and antithetical stratified whose rate of uniform convergence is  $N^{-1}$ . It is proven here that  $N^{-1}$  is a guaranteed minimum rate for all stratified-sampling-based estimators satisfying four weak conditions formulated in this paper. Owing to the alias-free nature of the sampling scheme, no constraints are imposed on the spectral support of the processed signal or the frequency ranges for which the Fourier Transform is estimated.

**Index Terms**— Fourier transform estimation, nonuniform sampling, alias-free sampling, stratified sampling, uniform convergence, digital alias-free signal processing

## I. INTRODUCTION

ESTIMATING the Fourier Transform (FT) from samples of the real-valued signal  $x(t)$  is an important task with applications in various areas of science and technology, including astronomy [1], seismology [2], biomedical sciences [3], NMR spectroscopy [4], and wireless communications where, for example, FT estimation is used for wideband spectrum sensing in cognitive radio networks [5] - [6]. When operating on sampled data, there is a possibility that the class of processed signals contains subsets within which all signals have identical discrete-time counterparts. This gives rise to the aliasing phenomenon and ambiguity in solving many DSP problems, including FT estimation. A standard way of avoiding aliasing is to considerably restrict the class of acquired signals and choose a sampling scheme that allows telling apart all the signals within that class.

Let  $\mathcal{F}$  be the spectral support of the signal  $x(t)$ , and  $\mathcal{F}_+$  its single-sided spectral support, where  $\mathcal{F} \subset R$ ,  $\mathcal{F}_+ \subset R_+$ ,  $R = (-\infty, \infty)$ , and  $R_+ = [0, \infty)$ . Then,  $[\mathcal{F}]$  and  $[\mathcal{F}_+]$  are the signal's spectral span and single-sided spectral span, i.e. the shortest intervals containing  $\mathcal{F}$  and  $\mathcal{F}_+$ , respectively. We denote by  $\mathcal{M}(\cdot)$  the Lebesgue measure of a set. To avoid aliasing when using uniform sampling, it suffices to select the sampling frequency  $f_S$  above the Nyquist rate  $f_N = \mathcal{M}([\mathcal{F}])$ :  $f_S > f_N$ . However, this popular solution could be inefficient unless  $x(t)$  is a baseband signal with known spectral support. An alternative way of selecting a uniform sampling rate is to use bandpass sampling [7], which exploits the fact that there exist uniform sampling rates  $f_S \in (f_{B,min}, f_{B,max}]$ , where  $f_{B,min} = 2\mathcal{M}([\mathcal{F}_+])$  and  $f_{B,max} = 4\mathcal{M}([\mathcal{F}_+])$ , which do not cause aliasing. If the processed signals are bandpass, these rates could be significantly lower than  $f_N$ . The theoretically lowest sampling rate that allows perfect signal reconstruction is the Landau rate  $f_L = \mathcal{M}(\mathcal{F})$  [8]. However, apart from some simple cases, such as processing lowpass signals, it is impossible to avoid aliasing while sampling signals uniformly at that rate. The solution is to deploy nonuniform sampling. For example, periodic nonuniform sampling was successfully used to sample multiband signals at rates arbitrarily close to  $f_L$ , without the adverse effects of the aliasing phenomenon [9].

When the spectral support  $\mathcal{F}$  is unknown, and instead its conservative approximation  $\hat{\mathcal{F}}$ :  $\hat{\mathcal{F}} \supset \mathcal{F}$  such that  $r = \mathcal{M}(\mathcal{F})/\mathcal{M}(\hat{\mathcal{F}}) \ll 1$ , has to be used to design the sampling scheme, the resultant sampling rates are likely to be excessive comparing to  $f_L$ ; typically by a factor of  $r^{-1}$ . Examples of such scenarios include instrumentation (e.g. when multiband signals with unknown central frequencies are acquired, as in spectrum analyzers), astronomy (e.g. detecting unknown periodic signals hidden in noise) and communication systems (e.g. wideband spectrum sensing). However, if an upper bound  $r_u$  of the ratio  $r$  is known and  $r \leq r_u < 1$  then the approaches, such as universal sampling [10] or compressed sensing [11], [12] and [13], offer solutions with sampling rates slightly exceeding  $f_{UL} = r_u \times \mathcal{M}(\hat{\mathcal{F}})$ . Another way of avoiding aliasing, while maintaining low sampling rates, emerges when signals are constrained in a domain other than frequency. Examples include signals with finite rate of innovation [14] where the sampling rate is linked

Copyright © 2015 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to [pubs-permissions@ieee.org](mailto:pubs-permissions@ieee.org).

A. Tarczynski is with the Faculty of Science and Technology, University of Westminster, 115 New Cavendish Street, London, W1W 6UW, UK (email: [tarczyna@wmin.ac.uk](mailto:tarczyna@wmin.ac.uk)).

B. I. Ahmad is with the Signal Processing and Communications Laboratory (SigProC), Department of Engineering, University of Cambridge, Trumpington Street, Cambridge, CB2 1PZ, UK (email: [bia23@cam.ac.uk](mailto:bia23@cam.ac.uk)).

to the rate of innovation, rather than the spectral support of the signal. Nevertheless, the methodologies described in [10]- [14] entail devising specialized processing algorithms that use advanced, and computationally or numerically demanding optimization techniques to determine the original continuous-time waveform or its specific properties, such as FT.

Interesting opportunities arise when the signals are sampled at random rather than deterministic time instants. In such cases, regardless of the nature of the original continuous-time signal (random or deterministic), the sampled signals are always random. Random sampling can be arranged in a way that different continuous-time signals always have their random, discrete-time counterparts distinct. Such sampling schemes are called *alias-free*, since they circumvent the reason of aliasing – a possibility that the same discrete-time signal can be obtained by sampling different continuous-time originals. Since alias-free sampling does not rely on knowledge of  $\mathcal{F}$ , it can be used when  $\mathcal{F}$  is unknown or when this set is fuzzy, i.e. there are no crisp borders between  $\mathcal{F}$  and its complement  $R - \mathcal{F}$ . For example, when a signal  $x(t)$  is observed over a finite-duration window, its spectrum is never perfectly confined to finite frequency bands.

The use of alias-free sampling was first proposed in [15] and applied to estimating the power spectrum of random stationary signals. This notion was then revisited in various studies, e.g. [16], [17], and extended to the analysis of other classes of signals, and to target different signal processing objectives. The domain of signal processing that exploits alias-free sampling became known as Digital Alias-free Signal Processing (DASP). Few monographs devoted to the use of nonuniform sampling, e.g. [18] and [19], discuss selected issues of random sampling and alias-free signal processing, whilst [20] addresses DASP directly. A review of DASP estimators of FT can be found in [21], whereas [22] provides a general study of spectral analyses from irregularly sampled data. With aliasing being no longer a concern in DASP, the focus is shifted towards furnishing efficient estimators of the required features of the signal and establishing statistical accuracy as well as relevant properties of the results. The challenge is that these goals normally need to be achieved using a single realization of the random discrete-time signal, thus, when the accessible information about the signal is very limited.

This paper introduces a novel DASP method of Fourier transform estimation. The estimation is performed using  $N$  signal samples taken from a finite duration observation window. The approach proposed here, named Hybrid Stratified (HySt), is directly linked to the work in [23], [24], [25] and [26]. We demonstrate that the HySt estimator is unbiased, consistent and it uniformly converges in mean square much faster than its predecessors, namely the Total Random Sampling (ToRa) [23], [24], Stratified Sampling (StSa) [25] and Antithetical Stratified Sampling (AnSt) [26] estimators. It is reported in [25] that both pointwise and the uniform convergence rate of ToRa is exactly  $N^{-1}$ . In the case of StSa and AnSt,  $N^{-1}$  was given as a lower band of uniform convergence rate [25], [26], even though their *pointwise* convergence rates are as fast as  $N^{-3}$  for StSa and  $N^{-5}$  for AnSt. In this paper we produce new results that could be summarized as follows:

(a) The uniform convergence rate of  $N^{-1}$  is a guaranteed minimum for all FT estimators that use stratified sampling and satisfy mild assumptions of Theorem 1 presented in Section II. We demonstrate that these assumptions are satisfied by StSa, AnSt and HySt methods.

(b) The uniform convergence rate of StSa and AnSt estimators does not exceed  $N^{-1}$ . Hence, their rate of uniform convergence is exactly  $N^{-1}$ .

(c) The HySt estimator is unbiased, consistent and it uniformly converges in mean square at the rate  $N^{-5}$ . Thus, this rate is significantly faster than  $N^{-1}$  - the minimum rate guaranteed by Theorem 1 and the actual uniform convergence rate of the existing DASP FT estimators, namely ToRa, StSa and AnSt.

(d) The pointwise convergence rate of the HySt estimator is  $N^{-5}$ . By this measure, the HySt estimator outperforms ToRa and StSa, and matches the performance of AnSt.

We note that there is no difference between the pointwise and uniform convergence rates for ToRa ( $N^{-1}$ ) and HySt ( $N^{-5}$ ) methods, whereas StSa and AnSt are characterized by slow uniform and much faster pointwise convergence rates. Consequently, the FT estimation errors show different behavior for these two pairs of approaches. More specifically we demonstrate that

(e) The relation between the number of collected samples  $N$  and the FT estimation error for ToRa and HySt is hardly affected by the frequency for which the FT is estimated.

(f) In the case of the StSa and AnSt, the observed estimation errors at individual frequencies converge slowly (at the rate of  $N^{-1}$ ) when  $N$  is small. Once  $N$  passes a critical threshold, this rate accelerates to  $N^{-3}$  and  $N^{-5}$ , respectively. These thresholds, however, depend on the frequency for which the FT is estimated. The higher the frequency the more samples are needed to trigger the faster convergence.

The side-effect of the observations (d) - (f) above is that the performances of AnSt and the proposed HySt method are similar to each other when the FT is estimated at low frequencies. The advantage of HySt over AnSt becomes visible, and then grows when the estimation of the FT is shifted towards higher frequencies.

The work on the HySt approach has been motivated by the need of constructing low-cost wideband FT estimators that use a small number of signal samples and avoid computationally-expensive processing algorithms. The approaches with fast uniform convergence and simple algorithms, such as proposed HySt technique are good candidates for this role. Potential application areas include, but are not limited, to the domains mentioned in the first paragraph of this introduction. For example, taking measurements for FT analyses in NMR spectroscopy is a costly and relatively lengthy process. In areas such as biochemistry, the time available for collecting all NMR data is limited if the tested molecules, e.g. certain proteins, maintain their properties for a short period. In this case, reducing  $N$  can be a technological and/or economic necessity. Similarly, reducing the number of collected samples can result in significant cost savings in astronomy or seismology. In wideband FT-based spectrum sensing for cognitive radio networks, it is a challenge to maintain low sampling rates without knowing the extent of the monitored signals' spectral

support. DASP-type sensing methods do not require such prior information, that is often unavailable [6], and can provide a simple low-complexity, yet effective, low-sampling-rates solutions [27]. In applications, where data is stored before being processed, collecting less measurements reduces the memory requirements; and therefore can be used instead or in conjunction with compression techniques such as Huffman coding.

The remainder of the paper is organized as follows. In Section II, the problem of FT estimation is formulated. Section III overviews the existing DASP estimators of FT and provides new results on their properties. It is proven there that the rate of uniform convergence of StSa and AnSt estimators is exactly  $N^{-1}$ . Theorem 1 formulates sufficient conditions under which DASP estimators uniformly converge at least at that rate. The proposed hybrid-stratified estimator is introduced and its features are explored in Section IV; it is proven that its uniform and pointwise convergence rates in mean square are  $N^{-5}$ . Numerical simulations in Section V are used to compare the performance of the HySt approach and its DASP predecessors. Final remarks are stated and conclusions are drawn in Section VI.

## II. PROBLEM FORMULATION

Our objective is to estimate the FT  $X(f)$  of a deterministic, continuous-time, real-valued signal  $x(t)$  truncated to the interval  $\mathcal{T} = [0, H]$ , using a finite number  $N$  of its samples. The target FT is defined by

$$X(f) \triangleq \int_{\mathcal{T}} x(t)w(t)e^{-j2\pi ft} dt, \quad (1)$$

where the character  $\triangleq$  denotes that the quantity on its left-hand side is defined by the expression on the right side. The windowing function  $w(t)$ , bounded by  $0 \leq w(t) \leq 1$  for  $t \in \mathcal{T}$ , is used to taper  $x(t)$  and keep  $X(f)$  smoothed. More details on how different shapes of  $w(t)$  affect the spectrum (1) can be found in [28] and [29]. The range of frequencies for which  $X(f)$  is estimated is arbitrary. No assumptions are made about the signal's spectral support.

HySt approach tackles this FT estimation problem by using alias-free sampling and devising a suitable unbiased estimator. The quality of estimation is measured by the mean square error

$$Err_N(f) \triangleq E\left\{|\hat{X}_N(f) - X(f)|^2\right\}, \quad (2)$$

where  $\hat{X}_N(f)$  denotes the FT estimator constructed from  $N$  samples of the signal  $x(t)$ . For unbiased estimators  $Err_N(f)$  is identical with the variance of the estimator  $\sigma^2\{\hat{X}_N(f)\}$ . In relation to (1), we denote

$$\lambda(t, f) \triangleq w(t)e^{-j2\pi ft}, \quad (3)$$

$$x_w(t) \triangleq x(t)w(t). \quad (4)$$

We also define:

$$x_{k,max} \triangleq \sup_{t \in \mathcal{T}} |x^{(k)}(t)|, \quad (5)$$

$$\lambda_{k,max}(f) \triangleq \sup_{t \in \mathcal{T}} |\lambda^{(k)}(t, f)|, \quad (6)$$

where  $x^{(k)}(t)$  and  $\lambda^{(k)}(t, f)$  denote  $k^{th}$  derivatives of  $x(t)$  and  $\lambda(t, f)$  with respect to time  $t$ . Assuming that  $w(t) = 1$  for some  $t \in \mathcal{T}$ , which is true for all commonly used windowing functions, it is noted that

$$\lambda_{0,max}(f) = 1. \quad (7)$$

Common notation used here is summarized in Table I. Any departure from it is explained in the paper.

TABLE I  
NOTATION

$x(t)$	Analyzed continuous-time signal
$X(f)$	Fourier transform of $x(t)$
$w(t)$	Windowing function
$\mathcal{T}$	A finite-duration observation window
$H$	Length of $\mathcal{T}$
$N$	Number of processed samples
$\hat{X}_N(f), \hat{X}_{ToRa,N}(f), \hat{X}_{StSa,N}(f), \hat{X}_{AnSt,N}(f), \hat{X}_{HySt,N}(f)$	FT estimates from $N$ samples
$E\{X\}$	Expected value of random variable $X$
$\sigma^2\{X\}$	Variance of random variable $X$
$Err_N(f)$	Mean square estimation error
$\tau_n$ or $\tau_{N,l}$	Random sampling instants
$L_N$	Number of strata
$\Delta_{N,l}$	Length of the $l^{th}$ stratum
$t_{N,l}$	Edges of strata
$\mathcal{T}_{N,l}$	The $l^{th}$ stratum given by $[t_{N,l}, t_{N,l+1}]$
$c_{N,l}$	Centre of the $l^{th}$ stratum
$g(t)$	Stratifying function
$I_{N,l}(f)$	FT of the $x(t)$ truncated to $\mathcal{T}_{N,l}$
$\hat{I}_{N,l}(f)$	Estimate of $I_{N,l}$

## III. OVERVIEW OF ALIAS-FREE FT ESTIMATORS

One of the early DASP estimators of FT is ToRa [23] - [24]. ToRa uses samples of  $x(t)$  collected at time instants  $\tau_n \in \mathcal{T}$ ,  $n = 0, \dots, N-1$  that are IID random variables with the Probability Density Function (PDF)  $f_{ToRa}(\tau)$ . The estimator defined by:  $\hat{X}_{ToRa,N}(f) = N^{-1} \sum_{n=0}^{N-1} x(\tau_n) \lambda(\tau_n, f) / f_{ToRa}(\tau_n)$  is unbiased for any  $N$  or  $f$ , i.e.  $E\{\hat{X}_{ToRa,N}(f)\} = X(f)$ . Its variance is  $\sigma^2\{\hat{X}_{ToRa,N}(f)\} = N^{-1} \sigma_{ToRa,lim}^2(f)$  where  $\sigma_{ToRa,lim}^2(f) = \int_{\mathcal{T}} x_w^2(\tau) / f_{ToRa}(\tau) d\tau - |X(f)|^2$ . Since  $\sigma_{ToRa,lim}^2(f) \leq \int_{\mathcal{T}} x_w^2(\tau) / f_{ToRa}(\tau) d\tau$ , where the right hand side does not depend on frequency, we note that  $\hat{X}_{ToRa,N}(f)$  uniformly converges in mean square to  $X(f)$  at rate  $N^{-1}$ . ToRa estimation method does not impose any significant constraints on the signal  $x(t)$  or windowing function  $w(t)$ . The above results hold as long as the integral (1) exists.

Two improvements to ToRa, namely StSa and AnSt, have been proposed in [25] and [26]. They rely on stratification of the interval  $\mathcal{T}$ . In the following subsections, we explore some of the properties of FT estimators that use stratification and demonstrate how these relate to StSa and AnSt.

### A. Stratification in FT Estimation

Stratification entails selecting  $L_N + 1$  time instants:  $0 = t_{N,0} < t_{N,1} < \dots < t_{N,L_N} = H$  and defining  $L_N$  strata by

$$\mathcal{T}_{N,l} \triangleq [t_{N,l}, t_{N,l+1}], \quad l = 0, \dots, L_N - 1. \quad (8)$$

The  $l^{th}$  stratum has the length

$$\Delta_{N,l} \triangleq t_{N,l+1} - t_{N,l} \quad (9)$$

and its center point is

$$c_{N,l} \hat{=} (t_{N,l+1} + t_{N,l})/2. \quad (10)$$

The FT (1) can be expressed by

$$X(f) = \sum_{l=0}^{L_N-1} I_{N,l}(f), \quad (11)$$

where

$$I_{N,l}(f) = \int_{\mathcal{T}_{N,l}} x(t)w(t)e^{-j2\pi ft} dt. \quad (12)$$

The estimator  $\hat{X}_N(f)$  of  $X(f)$  is constructed as a sum of the  $L_N$  estimators  $\hat{I}_{N,l}(f)$  of  $I_{N,l}(f)$

$$\hat{X}_N(f) = \sum_{l=0}^{L_N-1} \hat{I}_{N,l}(f). \quad (13)$$

In this paper, the strata are created with use of a stratifying function  $g(t)$ . This function is continuous on  $\mathcal{T}$ , and separated from 0 by  $g_{min}$

$$g(t) \geq g_{min} > 0. \quad (14)$$

Its average value on  $\mathcal{T}$  is one, i.e.

$$H^{-1} \int_{\mathcal{T}} g(t) dt = 1. \quad (15)$$

The boundaries  $t_{N,l}$ ,  $l = 0, \dots, L_N$  are solutions to the following equation

$$\int_0^{t_{N,l}} g(t) dt = H \frac{l}{L_N}. \quad (16)$$

This stratification is equivalent to that proposed in [25] and [26]. We note that the PDF  $h(t)$  used in [25] and [26] is related to the stratifying function  $g(t)$  by  $g(t) = H \times h(t)$ .

It follows from (16) that  $\int_{\mathcal{T}_{N,l}} g(t) dt = L_N^{-1} H$ . Since  $g(t)$  is continuous, the mean value theorem implies the existence of  $t_{h,l} \in \mathcal{T}_{N,l}$  such that

$$\Delta_{N,l} = \frac{1}{g(t_{h,l})} \frac{H}{L_N}. \quad (17)$$

This in turn implies that the lengths of strata are upper-bounded by

$$\Delta_{N,l} \leq \frac{1}{g_{min}} \frac{H}{L_N}. \quad (18)$$

Theorem 1 below provides sufficient conditions under which the estimator (13) is guaranteed to uniformly converge in mean square to  $X(f)$  at the rate  $N^{-1}$  or faster. Let  $\mathcal{D}_N \subset \mathcal{T}$  be the set of time instants at which the signal  $x(t)$  is sampled and  $\mathcal{D}_{N,l} \subset \mathcal{D}_N$  be a subset containing those sampling instants that are used to calculate  $\hat{I}_{N,l}(f)$ . The subsets  $\mathcal{D}_{N,l}$  may or may not overlap each other.

*Theorem 1:* Suppose that each  $\hat{I}_{N,l}(f)$  used in (13) is a linear combination of  $S$  samples of the signal  $x(t)$ , i.e. each  $\mathcal{D}_{N,l}$  contains exactly  $S$  time instants, denoted by  $\tau_{N,l,1}, \tau_{N,l,2}, \dots, \tau_{N,l,S}$ . Hence,

$$\hat{I}_{N,l}(f) = \sum_{r=1}^S a_{N,l,r}(f) x(\tau_{N,l,r}), \quad (19)$$

where  $a_{N,l,r}(f)$ ,  $l = 1, \dots, L_N$ ,  $r = 1, \dots, S$  are the estimator's multipliers whose values are selected appropriately to the method used for FT estimation.

If

$$(A.1) \quad \text{Estimator } \hat{X}_N(f) \text{ is unbiased, i.e. } E\{\hat{X}_N(f)\} = X(f);$$

$$(A.2) \quad \text{Estimators } \hat{I}_{N,l}(f), l = 1, \dots, L_N \text{ are independent from each other;}$$

$$(A.3) \quad \text{There exist } A_r, r = 1, \dots, S, \text{ such that for any } l, r \text{ and } f, \text{ we have } |a_{N,l,r}(f)| \leq \Delta_{N,l} A_r;$$

$$(A.4) \quad \text{There exists } D > 0 \text{ independent of } N \text{ such that } \sum_{l=0}^{L_N-1} \Delta_{N,l}^2 \leq N^{-1} D.$$

Then there exists  $B > 0$  independent of  $N$  such that for any  $f$ :  $Err_N(f) \leq N^{-1} B$ . (20)

*Proof of Theorem 1:* According to assumption (A.1), the estimator  $\hat{X}_N(f)$  is unbiased. Hence,  $Err_N(f) = \sigma^2\{\hat{X}_N(f)\}$ . It follows from (13) and assumption (A.2) that  $Err_N(f) = \sum_{l=0}^{L_N-1} \sigma^2\{\hat{I}_{N,l}(f)\}$ . Given (19) and that  $\sigma^2\{\hat{I}_{N,l}(f)\} \leq E\{|\hat{I}_{N,l}(f)|^2\}$ , we conclude:  $Err_N(f) \leq \sum_{l=0}^{L_N-1} E\{|\hat{I}_{N,l}(f)|^2\} = \sum_{l=0}^{L_N-1} E\{|\sum_{r=1}^S a_{N,l,r}(f) x(\tau_{N,l,r})|^2\} \leq \sum_{l=0}^{L_N-1} E\{(\sum_{r=1}^S |a_{N,l,r}(f)| |x(\tau_{N,l,r})|)^2\}$ . By (5) and assumptions (A.3) and (A.4) we obtain:  $Err_N(f) \leq x_{0,max}^2 \sum_{l=0}^{L_N-1} \Delta_{N,l}^2 \sum_{r=1}^S A_r^2$ , which leads to  $Err_N(f) \leq N^{-1} x_{0,max}^2 D \sum_{r=1}^S A_r^2$ . Therefore,  $B = x_{0,max}^2 D \sum_{r=1}^S A_r^2$ , which completes the proof of Theorem 1.

Theorem 1 reveals potential flexibilities when using stratification. The assumptions of this theorem could be satisfied even when not all sampling instants are chosen in a random manner, or some signal samples  $x(\tau_n)$  are used to calculate more than one  $\hat{I}_{N,l}(f)$ . These possibilities are exploited in the proposed HySt approach. It is noted that the convergence rate of an estimator satisfying Theorem 1 could be faster than  $N^{-1}$ . Such accelerated convergences may or may not be uniform. Lemma 1 below formulates a sufficient condition of assumption (A.4) of Theorem 1.

*Lemma 1:* If the strata borders  $t_l$ ,  $l = 0, \dots, L_N$  satisfy (16), and the number of samples  $N$  and strata  $L_N$  are related by

$$N = vL_N + \varpi, \quad (21)$$

where  $v \geq 1$  and  $\varpi \geq 0$ ; then  $\sum_{l=0}^{L_N-1} \Delta_{N,l}^2 \leq N^{-1} D$ , where  $D = (v + \varpi) H^2 / g_{min}^2$ .

*Proof of Lemma 1:* Since  $L_N \geq 1$ , it follows from (21) that  $N$  is lower-bounded by:  $N \geq v + \varpi$ . Additionally, we have  $L_N^{-1} = v/(N - \varpi) = N^{-1} v / (1 - \varpi N^{-1})$ . Thus,  $L_N^{-1} \leq N^{-1} \frac{v}{1 - \varpi/(v + \varpi)} = N^{-1} (v + \varpi)$ . The latter identity and (18) lead to:  $\sum_{l=0}^{L_N-1} \Delta_{N,l}^2 \leq L_N^{-1} H^2 / g_{min}^2 \leq N^{-1} (v + \varpi) H^2 / g_{min}^2 = N^{-1} D$ , where  $D = (v + \varpi) H^2 / g_{min}^2$ , which completes the proof of Lemma 1.

The three stratified approaches StSa, AnSt and HySt considered here use stratification strategies defined by (16) and, as shown in this paper, they satisfy (21). Therefore, Lemma 1 is used to prove that assumption (A.4) of Theorem 1 holds for StSa, AnSt and HySt schemes.

### B. StSa Estimation of Fourier Transform

In StSa FT estimation [25], the sampling instants are independent random variables distributed one per stratum, i.e.  $N = L_N$ . This satisfies (21) and thereby assumption (A.4) of Theorem 1. The PDF of the  $l^{\text{th}}$  sampling instant  $\tau_{N,l}$  is

$$f_{\text{StSa},N,l}(\tau) = \begin{cases} \Delta_{N,l}^{-1} & \text{if } \tau \in \mathcal{T}_{N,l} \\ 0 & \text{if } \tau \notin \mathcal{T}_{N,l} \end{cases} \quad (22)$$

The StSa estimator  $\hat{I}_{N,l}(f)$  of  $I_{N,l}(f)$  is given by

$$\hat{I}_{N,l}(f) = \Delta_{N,l} x(\tau_{N,l}) \lambda(\tau_{N,l}, f), \quad (23)$$

which means that  $S$  defined in assumption (A.3) of Theorem 1 is  $S = 1$ . The FT estimator is given by  $\hat{X}_{\text{StSa},N}(f) = \sum_{l=1}^N \Delta_{N,l} \lambda(\tau_{N,l}, f) x(\tau_{N,l})$ . According to [25],  $\hat{X}_{\text{StSa},N}(f)$  is unbiased. Since  $\tau_{N,l}$  are independent from each other and consequently  $\hat{I}_{N,l}(f)$  are also independent, StSa satisfies assumptions (A.1) and (A.2) of Theorem 1. To confirm (A.3), we deploy (7), (19) and (23) and note that  $a_{N,l,1}(f) = \Delta_{N,l} \lambda(\tau_{N,l}, f)$ . This leads to  $|a_{N,l,1}(f)| = \Delta_{N,l} |\lambda(\tau_{N,l}, f)| \leq \Delta_{N,l}$ . Since the StSa estimator satisfies all four assumptions of Theorem 1, its uniform convergence rate is at least  $N^{-1}$ .

Corollary 1 below asserts that  $N^{-1}$  is the fastest rate at which the StSa estimator can be guaranteed to uniformly converge in mean square. Let  $\mathcal{P}_{\text{StSa}}$  be the set of all rates at which the StSa estimator can be guaranteed to uniformly converge to  $X(f)$ . In other words, it contains all real numbers  $p > 0$  that for any signal  $x(t)$ , weighting function  $w(t)$  and stratifying function  $g(t)$  there exists  $\mathcal{B} > 0$  such that for any  $N$  and frequency  $f$ :

$$\sigma^2\{\hat{X}_{\text{StSa},N}(f)\} < N^{-p}\mathcal{B}. \quad (24)$$

*Corollary 1:*  $\mathcal{P}_{\text{StSa}} = (0, 1]$

*Proof of Corollary 1:* if  $p \in \mathcal{P}_{\text{StSa}}$ , then for any  $p_1 \in (0, p)$  and for any frequency  $f$  we have:  $\sigma^2\{\hat{X}_{\text{StSa},N}(f)\} < N^{-p}\mathcal{B} < N^{-p_1}\mathcal{B}$ , hence,  $p_1 \in \mathcal{P}_{\text{StSa}}$ . Following this observation and the fact that Theorem 1 stipulates that  $1 \in \mathcal{P}_{\text{StSa}}$ , we note that  $(0, 1] \subset \mathcal{P}_{\text{StSa}}$ . To complete the proof, it suffices to present an example of a signal  $x(t)$ , functions  $w(t)$  and  $g(t)$  where for any  $\mathcal{B}$  and  $p > 1$ , there exist  $N$  and  $f$  such that (24) does not hold. Let  $H = 1$  and  $x(t) = w(t) = g(t) = 1$ . Hence the length of each stratum is  $\Delta_{N,l} = N^{-1}$ . We note that  $I_{N,l}(f) = N^{-1} \text{sinc}(f/N) \exp(-j2\pi f c_{N,l})$  and its StSa estimate is  $\hat{I}_{N,l}(f) = N^{-1} \exp(-j2\pi f \tau_{N,l})$ . Thus,  $\sigma^2\{\hat{I}_{N,l}(f)\} = \text{E}\{|\hat{I}_{N,l}(f)|^2\} - |I_{N,l}(f)|^2 = N^{-2}[1 - \text{sinc}^2(f/N)]$  and  $\sigma^2\{\hat{X}_{\text{StSa},N}(f)\} = N\sigma^2\{\hat{I}_{N,l}(f)\} = N^{-1}[1 - \text{sinc}^2(f/N)]$ . For given  $\mathcal{B}$  and  $p > 1$ , we select any  $N$  that satisfies  $N > \sqrt[p-1]{\mathcal{B}}$  and frequency  $f = N$ . In this case,  $\sigma^2\{\hat{X}_{\text{StSa},N}(f)\} = N^{-1}$  and  $N^{-p}\mathcal{B} < N^{-p}N^{p-1} = N^{-1} = \sigma^2\{\hat{X}_{\text{StSa},N}(f)\}$ . Therefore,  $\sigma^2\{\hat{X}_{\text{StSa},N}(f)\} > N^{-p}\mathcal{B}$ , implying that if  $p > 1$ , then  $p \notin \mathcal{P}_{\text{StSa}}$ . This completes the proof of Corollary 1.

It is shown in [25] that if the derivative of  $x_w(t)$  is continuous, then at individual frequencies the convergence rate of the StSa estimator can be as fast as  $N^{-3}$ . Specifically, for each frequency  $f$ , the following holds:  $\lim_{N \rightarrow \infty} N^3 \sigma^2\{\hat{X}_{\text{StSa},N}(f)\} = \sigma_{\text{StSa},\text{lim}}^2(f)$ , where  $\sigma_{\text{StSa},\text{lim}}^2(f) = A_{\text{StSa}} f^2 + B_{\text{StSa}}$ ,  $A_{\text{StSa}} = H^3 \pi^2 \int_{\mathcal{T}} x_w^2(t) / 3g^3(t) dt$ , and

$B_{\text{StSa}} = H^3 \int_{\mathcal{T}} [x_w^{(1)}(t)]^2 / 12g^3(t) dt$ . We show here that in order to observe this accelerated convergence, the number of samples  $N$  that have to be collected increases with frequency  $f$ . The proof is by contradiction. Let's assume that opposite is true, i.e. for any  $\varepsilon > 0$ , there exists  $N_0$  such that for any  $N \geq N_0$  and frequency  $f$ :  $|\sigma_{\text{StSa},\text{lim}}^2(f) - N^3 \sigma^2\{\hat{X}_{\text{StSa},N}(f)\}| < \varepsilon$ . If this was true, we would have:  $\sigma_{\text{StSa},\text{lim}}^2(f) < N_0^3 \sigma^2\{\hat{X}_{\text{StSa},N_0}(f)\} + \varepsilon$  for any  $N > N_0$  and  $f$ . However, it follows from Theorem 1 that  $\sigma^2\{\hat{X}_{\text{StSa},N_0}(f)\} < N_0^{-1}B$ . Therefore,  $\sigma_{\text{StSa},\text{lim}}^2(f) < N_0^2 B + \varepsilon$ , and consequently for any  $f$ :  $A_{\text{StSa}} f^2 + B_{\text{StSa}} < N_0^2 B + \varepsilon$ . Since the left hand side of the last expression goes to infinity when  $f \rightarrow \infty$ , this relation cannot hold for all  $f$  regardless of how  $N_0$  was selected. This confirms that, in general, in order to observe the accelerated convergence, the number of collected samples  $N_0$  has to increase with  $f$ .

### C. AnSt Fourier Transform Estimates

In the AnSt approach [26], two samples of  $x(t)$  are collected in each stratum, i.e.  $N = 2L_N$ , which conforms with (21) and satisfies assumption (A.4) of Theorem 1. The first sample is selected randomly in the same way as for StSa. The second one is taken in an antithetical manner, i.e. if the first sampling time in the  $l^{\text{th}}$  stratum is  $\tau_{N,2l}$ , the second one is  $\tau_{N,2l+1} = 2c_{N,l} - \tau_{N,2l}$ . The estimators of  $I_{N,l}(f)$  are

$$\hat{I}_{N,l}(f) = 0.5 \Delta_{N,l} [\lambda(\tau_{N,2l}, f) x(\tau_{N,2l}) + \lambda(\tau_{N,2l+1}, f) x(\tau_{N,2l+1})]. \quad (25)$$

Consequently, the antithetical stratified FT estimator of  $X(f)$  is given by  $\hat{X}_{\text{AnSt},N}(f) = 0.5 \sum_{l=1}^N \Delta_{N,l} [\lambda(\tau_{N,2l}, f) x(\tau_{N,2l}) + \lambda(\tau_{N,2l+1}, f) x(\tau_{N,2l+1})]$ . It is shown in [26] that  $\hat{X}_{\text{AnSt},N}(f)$  is unbiased. Hence, assumption (A.1) is satisfied. Since for  $l \neq k$ , the pair of random time instants  $(\tau_{N,2l}, \tau_{N,2l+1})$  is statistically independent from  $(\tau_{N,2k}, \tau_{N,2k+1})$ , assumption (A.2) also holds. Additionally, it follows from (19), (25) and (7) that  $S = 2$ ,  $|a_{N,l,1}(f)| = 0.5 \Delta_{N,l} |\lambda(\tau_{N,2l}, f)| \leq 0.5 \Delta_{N,l}$ , and  $|a_{N,l,2}(f)| = 0.5 \Delta_{N,l} |\lambda(\tau_{N,2l+1}, f)| \leq 0.5 \Delta_{N,l}$ , confirming that (A.3) of Theorem 1 is fulfilled with  $A_1 = A_2 = 0.5$ . Thus, according to Theorem 1 the AnSt estimator uniformly converges in mean square to  $X(f)$  at least at rate  $N^{-1}$ .

Similarly to the StSa case, we show that the uniform convergence rate of AnSt estimation is exactly  $N^{-1}$ . This is formally stated by Corollary 2. Let  $\mathcal{P}_{\text{AnSt}}$  be the set of all rates at which the AnSt estimator can be guaranteed to uniformly converge to  $X(f)$ , i.e. a collection of all real numbers  $p > 0$  that for any signal  $x(t)$ , weighting function  $w(t)$  and stratifying function  $g(t)$  there exists  $\mathcal{B} > 0$  such that for any  $N$  and frequency  $f$ :

$$\sigma^2\{\hat{X}_{\text{AnSt},N}(f)\} < N^{-p}\mathcal{B}, \quad p \in \mathcal{P}_{\text{AnSt}}. \quad (26)$$

*Corollary 2:*  $\mathcal{P}_{\text{AnSt}} = (0, 1]$

*Proof of Corollary 2:* By repeating the initial steps of the proof of Corollary 1, with the relevant changes, we can show that  $(0, 1] \subset \mathcal{P}_{\text{AnSt}}$ . To complete the proof it suffices to present an example of a signal  $x(t)$  and functions  $w(t)$  and  $g(t)$  where for any  $\mathcal{B}$  and  $p > 1$ , there exist  $N$  and  $f$  such that (26) does not hold. Again, we use:  $H = 1$  and  $x(t) = w(t) = g(t) = 1$ , hence  $\Delta_{N,l} = N^{-1}$  and  $I_{N,l}(f) =$

$N^{-1} \text{sinc}(f/N) \exp(-j2\pi f c_{N,l})$ . However, the AnSt estimate of  $I_{N,l}(f)$  is  $\hat{I}_{N,l}(f) = N^{-1} \exp(-j2\pi f c_{N,l}) \cos(2\pi f(c_n - \tau_{N,2l}))$ . Hence,  $E\{|\hat{I}_{N,l}(f)|^2\} = N^{-2}[0.5 + 0.5 \text{sinc}(2f/N)]$ . Using  $|I_{N,l}(f)|^2 = N^{-2} \text{sinc}^2(f/N)$ , we attain:  $\sigma^2\{\hat{I}_{N,l}(f)\} = E\{|\hat{I}_{N,l}(f)|^2\} - |I_{N,l}(f)|^2 = N^{-2}[0.5 + 0.5 \text{sinc}(2f/N) - \text{sinc}^2(f/N)]$ . Taking into account that  $\sigma^2\{\hat{X}_{AnSt,N}(f)\}$  is the sum of  $L_N = N/2$  components,  $\sigma^2\{\hat{I}_{N,l}(f)\}$ , we get  $\sigma^2\{\hat{X}_{AnSt,N}(f)\} = N^{-1}[0.25 + 0.25 \text{sinc}(2f/N) - 0.5 \text{sinc}^2(f/N)]$ . For a given  $\mathcal{B}$  and  $p > 1$ , we select any  $N$  satisfying  $N > \sqrt[p-1]{4\mathcal{B}}$  and choose frequency  $f = N$ . This leads to  $\sigma^2\{\hat{X}_{AnSt,N}(f)\} = 0.25N^{-1}$ . We also note that  $N^{-p}\mathcal{B} < 0.25N^{p-1}N^{-p} = 0.25N^{-1} = \sigma^2\{\hat{X}_{AnSt,N}(f)\}$ . Subsequently,  $\sigma^2\{\hat{X}_{AnSt,N}(f)\} > N^{-p}\mathcal{B}$ , which means that if  $p > 1$  then  $p \notin \mathcal{P}_{AnSt}$ . This finalizes the proof of Corollary 2.

It has been shown in [26] that if  $x_w(t)$  has continuous second derivative, the AnSt estimator can converge as fast as  $N^{-5}$  at individual frequencies. Specifically, for a given frequency  $f$ , the following holds:  $\lim_{N \rightarrow \infty} N^5 \sigma^2\{\hat{X}_{AnSt,N}(f)\} = \sigma_{AnSt,lim}^2(f)$  where  $\sigma_{AnSt,lim}^2(f) = A_{AnSt}f^4 + B_{AnSt}f^2 + C_{AnSt}$ ,  $A_{AnSt} = \frac{32}{45}H^5\pi^4 \int_{\mathcal{T}} \frac{x_w^{(2)}(t)}{g^5(t)} dt$ ,  $B_{AnSt} = \frac{8}{45}H^5\pi^2 \int_{\mathcal{T}} \frac{2[x_w^{(1)}(t)]^2 - x_w^{(2)}(t)x_w(t)}{g^5(t)} dt$ , and  $C_{AnSt} = \frac{2}{45}H^5 \int_{\mathcal{T}} \frac{[x_w^{(2)}(t)]^2}{g^5(t)} dt$ . Similarly to the StSa case, for this accelerated convergence be observed the number of processed samples  $N$  must generally increase with frequency  $f$ . We prove this fact by contradiction. Let's assume that for any  $\varepsilon > 0$  there exists  $N_0$  such that for any  $N \geq N_0$  and frequency  $f$ :  $|N^5 \hat{X}_{AnSt,N}(f) - \sigma_{AnSt,lim}^2(f)| < \varepsilon$ . This assumption implies  $\sigma_{AnSt,lim}^2(f) < N_0^5 \sigma^2\{\hat{X}_{AnSt,N_0}(f)\} + \varepsilon$ . By Theorem 1:  $\sigma^2\{\hat{X}_{AnSt,N_0}(f)\} < N_0^{-1}B$ . Thus,  $\sigma_{AnSt,lim}^2(f) < N_0^4B + N_0^5\varepsilon$ , and  $A_{AnSt}f^4 + B_{AnSt}f^2 + C_{AnSt} < N_0^2B + N_0^3\varepsilon$ . Since this relation cannot hold for any  $f$  regardless of how  $N_0$  is selected, we conclude that  $N_0$  must in general increase with the frequency  $f$ .

#### IV. HYBRID-STRATIFIED SAMPLING AND FOURIER TRANSFORM ESTIMATION

In this section, we introduce the HySt method and explore its features. In particular, we show that the uniform convergence of the HySt estimator significantly outperforms its predecessors described in the previous section.

##### A. HySt Estimator of Fourier Transform

In the HySt approach, the sampling instants are a mixture of deterministic and random variables. The random instants  $\tau_{N,l}$ ,  $l = 0, \dots, L_N - 1$  are selected in the same manner as those in StSa. The deterministic ones are the strata borders:  $t_{N,l}$ ,  $l = 0, \dots, L_N$ . The total number of processed samples is

$$N = 2L_N + 1. \quad (27)$$

It follows from (18) that  $\Delta_{N,l} \leq \frac{2}{g_{min}} \frac{H}{N-1}$ . Since  $L_N \geq 1$  and consequently  $N \geq 3$  we get

$$\Delta_{N,l} \leq \frac{3}{N} \frac{H}{g_{min}}. \quad (28)$$

The HySt estimator of  $\hat{I}_{N,l}(f)$  is a linear combination of  $x(t_{N,l})$ ,  $x(\tau_{N,l})$  and  $x(t_{N,l+1})$

$$\hat{I}_{N,l}(f) \hat{=} \alpha_{N,l}(f)x(t_{N,l}) + \beta_{N,l}(f)x(\tau_{N,l}) + \gamma_{N,l}(f)x(t_{N,l+1}), \quad (29)$$

where

$$\alpha_{N,l}(f) \hat{=} \Delta_{N,l}^{-1} \int_{\mathcal{T}_{N,l}} (t_{N,l+1} - t)\lambda(t, f) dt \quad (30)$$

$$- (t_{N,l+1} - \tau_{N,l})\lambda(\tau_{N,l}, f), \quad (31)$$

$$\beta_{N,l}(f) \hat{=} \Delta_{N,l}\lambda(\tau_{N,l}, f),$$

$$\gamma_{N,l}(f) \hat{=} \Delta_{N,l}^{-1} \int_{\mathcal{T}_{N,l}} (t - t_{N,l})\lambda(t, f) dt \quad (32)$$

$$- (\tau_{N,l} - t_{N,l})\lambda(\tau_{N,l}, f).$$

Hence, the estimator of  $X(f)$  is

$$\hat{X}_{HySt,N}(f) = \sum_{l=0}^{L_N-1} \alpha_{N,l}(f)x(t_{N,l}) + \beta_{N,l}(f)x(\tau_{N,l}) + \gamma_{N,l}(f)x(t_{N,l+1}). \quad (33)$$

We demonstrate that  $\hat{X}_{HySt,N}(f)$  satisfies the assumptions (A.1) - (A.4) of Theorem 1 and therefore uniformly converges in mean square to  $X(f)$  at the rate  $N^{-1}$  or faster. According to (11) and (13), assumption (A.1) holds if all  $\hat{I}_{N,l}(f)$  are unbiased. In fact:  $E\{\hat{I}_{N,l}(f)\} = E\{\alpha_{N,l}(f)\}x(t_{N,l}) + E\{\beta_{N,l}(f)\}x(\tau_{N,l}) + E\{\gamma_{N,l}(f)\}x(t_{N,l+1})$ . Now, we note that  $E\{\alpha_{N,l}(f)\} = \Delta_{N,l}^{-1} \int_{\mathcal{T}_{N,l}} (t_{N,l+1} - t)\lambda(t, f) dt - E\{(t_{N,l+1} - \tau_{N,l})\lambda(\tau_{N,l}, f)\} = \Delta_{N,l}^{-1} \int_{\mathcal{T}_{N,l}} (t_{N,l+1} - t)\lambda(t, f) dt - \Delta_{N,l}^{-1} \int_{\mathcal{T}_{N,l}} (t_{N,l+1} - \tau)\lambda(\tau, f) dt = 0$ ,  $E\{\beta_{N,l}(f)\}x(\tau_{N,l}) = \Delta_{N,l}^{-1} \int_{\mathcal{T}_{N,l}} \Delta_{N,l}\lambda(\tau, f)x(\tau) d\tau = \int_{\mathcal{T}_{N,l}} \lambda(\tau, f)x(\tau) d\tau = I_{N,l}(f)$ , and  $E\{\gamma_{N,l}(f)\} = \Delta_{N,l}^{-1} \int_{\mathcal{T}_{N,l}} (t - t_{N,l})\lambda(t, f) dt - E\{(\tau_{N,l} - t_{N,l})\lambda(\tau_{N,l}, f)\} = \Delta_{N,l}^{-1} \int_{\mathcal{T}_{N,l}} (t - t_{N,l})\lambda(t, f) dt - \Delta_{N,l}^{-1} \int_{\mathcal{T}_{N,l}} (\tau - t_{N,l})\lambda(\tau, f) dt = 0$ . By combining these observations, we confirm that  $E\{\hat{I}_{N,l}(f)\} = I_{N,l}(f)$  and thus  $E\{\hat{X}_{HySt,N}(f)\} = X(f)$ .

Assumption (A.2) is also satisfied since for  $l \neq k$ , the random time instants  $\tau_{N,l}$  and  $\tau_{N,k}$  are independent from each other. Consequently,  $\hat{I}_{N,l}(f)$  and  $\hat{I}_{N,k}(f)$  are also independent. To confirm assumption (A.3), we note that  $S = 3$ . By using (7), we get:  $|\alpha_{N,l}| \leq \Delta_{N,l}^{-1} \int_{\mathcal{T}_{N,l}} (t_{N,l+1} - t)|\lambda(t, f)| dt + (t_{N,l+1} - \tau_{N,l})|\lambda(\tau_{N,l}, f)| \leq \Delta_{N,l}^{-1} \int_{\mathcal{T}_{N,l}} (t_{N,l+1} - t) dt + \Delta_{N,l} = 1.5\Delta_{N,l}$ .

Hence

$$|\alpha_{N,l}| \leq 1.5\Delta_{N,l}, \quad (34)$$

and then,

$$|\beta_{N,l}| \leq \Delta_{N,l} |\lambda(\tau_{N,l}, f)| \leq \Delta_{N,l} \quad (35)$$

and  $|\gamma_{N,l}| \leq \Delta_{N,l}^{-1} \int_{\mathcal{T}_{N,l}} (t - t_{N,l}) |\lambda(t, f)| dt + (\tau_{N,l} - t_{N,l}) |\lambda(\tau_{N,l}, f)| \leq \Delta_{N,l}^{-1} \int_{\mathcal{T}_{N,l}} (t - t_{N,l}) dt + \Delta_{N,l} = 1.5\Delta_{N,l}$ . This yields

$$|\gamma_{N,l}| \leq 1.5\Delta_{N,l}. \quad (36)$$

The relationships (34) - (36) confirm that (A.3) is satisfied with  $A_1 = A_3 = 1.5$  and  $A_2 = 1$ . Finally, we use (27) and Lemma 1 to affirm that assumption (A.4) holds. This concludes the proof that HySt estimator uniformly converges to  $X(f)$  at least at rate  $N^{-1}$ .

Before we investigate further the properties of the HySt estimator, we use elementary calculations to show that

$$\mathcal{L}_{0,l} \hat{=} \alpha_{N,l}(f) + \beta_{N,l}(f) + \gamma_{N,l}(f) = \int_{\mathcal{T}_{N,l}} \lambda(t, f) dt, \quad (37)$$

$$\begin{aligned} \mathcal{L}_{1,l} \hat{=} & -\alpha_{N,l}(f) \frac{\Delta_{N,l}}{2} + \beta_{N,l}(f) (\tau_{N,l} - c_{N,l}) \\ & + \gamma_{N,l}(f) \frac{\Delta_{N,l}}{2} \end{aligned} \quad (38)$$

$$= \int_{\mathcal{T}_{N,l}} (t - c_{N,l}) \lambda(t, f) dt,$$

$$\begin{aligned} \mathcal{L}_{2,l} \hat{=} & \alpha_{N,l}(f) \frac{\Delta_{N,l}^2}{8} + \beta_{N,l}(f) \frac{(\tau_{N,l} - c_{N,l})^2}{2} \\ & + \gamma_{N,l}(f) \frac{\Delta_{N,l}^2}{8} \\ & = \frac{\Delta_{N,l}^2}{8} \int_{\mathcal{T}_{N,l}} \lambda(t, f) dt \\ & - \lambda(\tau_{N,l}, f) \left[ \frac{\Delta_{N,l}^3}{8} \right. \\ & \left. - \Delta_{N,l} \frac{(\tau_{N,l} - c_{N,l})^2}{2} \right]. \end{aligned} \quad (39)$$

and according to (34) - (36), we have

$$\mathcal{M}_l \hat{=} |\alpha_{N,l}| + |\beta_{N,l}| + |\gamma_{N,l}| \leq 4\Delta_{N,l}. \quad (40)$$

### B. Fast Uniform Convergence of HySt Estimation

In this subsection, we prove that if the signal  $x(t)$  has a continuous third derivative in some open interval  $\mathcal{T}_B$  comprising  $\mathcal{T}$ :  $\mathcal{T} \subset \mathcal{T}_B$  then the rate of uniform convergence of the HySt estimator is at least  $N^{-5}$ . The analyses in the next subsection combined with this result prove a stronger statement, namely that this rate is exactly  $N^{-5}$ . Let

$$\mathcal{E}_{N,l}(f) \hat{=} \hat{I}_{N,l}(f) - I_{N,l}(f), \quad (41)$$

denote the error of estimating  $I_{N,l}(f)$ . Subsequently, the variance of  $\hat{I}_{N,l}(f)$  is

$$\sigma^2\{\hat{I}_{N,l}(f)\} = \mathbb{E}\{|\mathcal{E}_{N,l}(f)|^2\}. \quad (42)$$

Since for  $l \neq k$   $\hat{I}_{N,l}(f)$  and  $\hat{I}_{N,k}(f)$  are independent from each other, the variance of HySt estimator is given by

$$\sigma^2\{\hat{X}_{HySt,N}(f)\} = \sum_{l=0}^{L_N-1} \mathbb{E}\{|\mathcal{E}_{N,l}(f)|^2\}. \quad (43)$$

Consider the following second order Taylor expansions of  $x(t)$  about each stratum center  $c_{N,l}$

$$\begin{aligned} x(t) = & x(c_{N,l}) + (t - c_{N,l})x^{(1)}(c_{N,l}) \\ & + \frac{(t - c_{N,l})^2}{2}x^{(2)}(c_{N,l}) + r_{N,l}(t). \end{aligned} \quad (44)$$

Since  $x^{(3)}(t)$  is continuous in  $\mathcal{T}_B$  and therefore bounded in  $\mathcal{T}$ , the remainder  $r_{N,l}(t)$  can be put in the Lagrange form  $r_{N,l}(t) = \frac{(t - c_{N,l})^3}{6}x^{(3)}(\tilde{t})$ , where  $\tilde{t} \in \mathcal{T}_{N,l}$  is chosen to satisfy (44) and  $|x^{(3)}(\tilde{t})| \leq x_{3,max} < \infty$ . For any  $t \in \mathcal{T}_{N,l}$ :  $|t - c_{N,l}| \leq 0.5\Delta_{N,l}$ , we get  $|r_{N,l}(t)| \leq \frac{\Delta_{N,l}^3}{48}|x^{(3)}(\tilde{t})|$ , and

$$|r_{N,l}(t)| \leq \frac{\Delta_{N,l}^3}{48}x_{3,max}. \quad (45)$$

By substituting (44) in (12), and deploying (7) and (37)-(38), we obtain

$$\begin{aligned} I_{N,l}(f) = & \mathcal{L}_{0,l}x(c_{N,l}) + \mathcal{L}_{1,l}x^{(1)}(c_{N,l}) \\ & + \int_{\mathcal{T}_{N,l}} \frac{(t - c_{N,l})^2}{2}\lambda(t, f) dt x^{(2)}(c_{N,l}) \\ & + \chi_l \end{aligned} \quad (46)$$

where  $\chi_l = \int_{\mathcal{T}_{N,l}} \lambda(t, f)r_{N,l}(t) dt$ . It follows from (7) and (45) that  $|\chi_l| \leq \Delta_{N,l}^4 x_{3,max}/48$ . By utilizing (28), we get

$$|\chi_l| \leq N^{-4} \frac{27}{16} \frac{H^4}{g_{min}^4} x_{3,max}. \quad (47)$$

From (44), the signal samples  $x(t_{N,l})$ ,  $x(\tau_{N,l})$  and  $x(t_{N,l+1})$  can be expressed as:  $x(t_{N,l}) = x(c_{N,l}) - \frac{\Delta_{N,l}}{2}x^{(1)}(c_{N,l}) + \frac{\Delta_{N,l}^2}{8}x^{(2)}(c_{N,l}) + r_{N,l}(t_{N,l})$ ,  $x(\tau_{N,l}) = x(c_{N,l}) + (\tau_l - c_{N,l})x^{(1)}(c_{N,l}) + \frac{(\tau_l - c_{N,l})^2}{2}x^{(2)}(c_{N,l}) + r_{N,l}(\tau_{N,l})$ , and  $x(t_{N,l+1}) = x(c_{N,l}) + \frac{\Delta_{N,l}}{2}x^{(1)}(c_{N,l}) + \frac{\Delta_{N,l}^2}{8}x^{(2)}(c_{N,l}) + r_{N,l}(t_{N,l+1})$ . By substituting these in (29), and using (30) - (32) and (37) - (39), we obtain

$$\begin{aligned} \hat{I}_{N,l}(f) = & \mathcal{L}_{0,l}x(c_{N,l}) + \mathcal{L}_{1,l}x^{(1)}(c_{N,l}) + \mathcal{L}_{2,l}x^{(2)}(c_{N,l}) \\ & + \hat{\chi}_l \end{aligned} \quad (48)$$

where  $\hat{\chi}_l = \alpha_{N,l}(f)r_{N,l}(t_l) + \beta_{N,l}(f)r_{N,l}(\tau_l) + \gamma_{N,l}(f)r_{N,l}(t_{l+1})$ . An upper bound for  $|\hat{\chi}_l|$  can be calculated using (40) and (45):  $|\hat{\chi}_l| \leq |\alpha_{N,l}(f)||r_{N,l}(t_l)| + |\beta_{N,l}(f)||r_{N,l}(\tau_l)| + |\gamma_{N,l}(f)||r_{N,l}(t_{l+1})| \leq \mathcal{M}_l \frac{\Delta_{N,l}^3}{48}x_{3,max} \leq \frac{\Delta_{N,l}^4}{12}x_{3,max}$ . Subsequently, by (28), we get

$$|\hat{\chi}_l| \leq N^{-4} \frac{27}{4} \frac{H^4}{g_{min}^4} x_{3,max}. \quad (49)$$

Now we derive the estimation error (41). By subtracting (46) from (48), we get

$$\begin{aligned} \mathcal{E}_{N,l}(f) = & \left\{ \frac{\Delta_{N,l}^2}{8} \int_{\mathcal{T}_{N,l}} \lambda(t, f) dt - \lambda(\tau_l, f) \left[ \frac{\Delta_{N,l}^3}{8} - \right. \right. \\ & \left. \left. \Delta_{N,l} \frac{(\tau_{N,l} - c_{N,l})^2}{2} \right] - \int_{\mathcal{T}_{N,l}} \frac{(t - c_{N,l})^2}{2} \lambda(t, f) dt \right\} x^{(2)}(c_{N,l}) + \hat{\chi}_l - \chi_l. \end{aligned}$$

Therefore,



$$\mathcal{E}_{N,l}(f) = 0.5 \left[ \int_{\mathcal{T}_{N,l}} z_{N,l}(t) \lambda(t, f) dt - \Delta_{N,l} z_{N,l}(\tau_{N,l}) \lambda(\tau_{N,l}, f) \right] x^{(2)}(c_{N,l}) + \bar{\chi}_l \quad (50)$$

where  $z_{N,l}(t) \triangleq \frac{\Delta_{N,l}^2}{4} - (t - c_{N,l})^2 = \left(\frac{\Delta_{N,l}}{2} - t + c_{N,l}\right) \left(\frac{\Delta_{N,l}}{2} + t - c_{N,l}\right) = (t_{N,l+1} - t)(t - t_{N,l})$  and  $\bar{\chi}_l = \hat{\chi}_l - \chi_l$ . It can be checked that the function  $z_{N,l}(t)$  has the following properties

$$\int_{\mathcal{T}_{N,l}} z_{N,l}(t) dt = \frac{\Delta_{N,l}^3}{6} \quad (51)$$

$$\int_{\mathcal{T}_{N,l}} z_{N,l}^2(t) dt = \frac{\Delta_{N,l}^5}{30} \quad (52)$$

$$\int_{\mathcal{T}_{N,l}} z_{N,l}(t) |t - c_{N,l}| dt = \frac{\Delta_{N,l}^4}{32} \quad (53)$$

and, if  $t \in \mathcal{T}_{N,l}$ , then

$$0 \leq z_{N,l}(t) \leq \Delta_{N,l}^2/4 \quad (54)$$

Since  $E\{z_{N,l}(\tau_{N,l})\} = \Delta_{N,l}^{-1} \int_{\mathcal{T}_{N,l}} z_{N,l}(\tau) d\tau$ , then (51) implies

$$E\{z_{N,l}(\tau_{N,l})\} = \Delta_{N,l}^2/6. \quad (55)$$

From (47) and (49), we derive

$$|\bar{\chi}_l| \leq N^{-4} \frac{135}{16} \frac{H^4}{g_{min}^4} x_{3,max}. \quad (56)$$

It follows from (50), (5) and (7) that  $|\mathcal{E}_{N,l}(f)| \leq 0.5 \left[ \int_{\mathcal{T}_{N,l}} z_{N,l}(t) dt + \Delta_{N,l} z_{N,l}(\tau_{N,l}) \right] x_{2,max} + |\bar{\chi}_l|$ . Consequently, we have:

$$|\mathcal{E}_{N,l}(f)| \leq 0.5 \left( \frac{\Delta_{N,l}^3}{6} + \frac{\Delta_{N,l}^3}{4} \right) x_{2,max} + |\bar{\chi}_l| \leq \frac{5}{24} \Delta_{N,l}^3 x_{2,max} + |\bar{\chi}_l| \leq N^{-3} \left( \frac{45}{8} \frac{H^3}{g_{min}^3} x_{2,max} + N^{-1} \frac{135}{16} \frac{H^4}{g_{min}^4} x_{3,max} \right).$$

Since  $N \geq 3$ , we reach

$$|\mathcal{E}_{N,l}(f)| \leq N^{-3} \left[ \frac{45}{8} \frac{H^3}{g_{min}^3} x_{2,max} + \frac{45}{16} \frac{H^4}{g_{min}^4} x_{3,max} \right] \quad (57)$$

Substituting (57) in (43) yields  $\sigma^2\{\hat{X}_{HySt,N}(f)\} \leq L_N N^{-6} \left[ \frac{45}{8} \frac{H^3}{g_{min}^3} x_{2,max} + \frac{45}{16} \frac{H^4}{g_{min}^4} x_{3,max} \right]^2$ . Since  $L_N \leq 0.5N$ , we conclude that

$$\sigma^2\{\hat{X}_{HySt,N}(f)\} \leq 0.5N^{-5} \left[ \frac{45}{8} \frac{H^3}{g_{min}^3} x_{2,max} + \frac{45}{16} \frac{H^4}{g_{min}^4} x_{3,max} \right]^2 \quad (58)$$

which proves the following theorem:

**Theorem 2:** If the signal  $x(t)$  has continuous third derivative in  $\mathcal{T}_B$ , the HySt estimator converges uniformly to  $X(f)$  at least at the rate  $N^{-5}$ .

### C. Asymptotic variance of the HySt estimator

In this section, we derive the asymptotic variance of the HySt estimator. Theorem 3 below states the main result.

**Theorem 3:** If  $w(t)$  and  $x(t)$  have continuous first and third derivatives in  $\mathcal{T}_B$  respectively then

$$\lim_{N \rightarrow \infty} N^5 \sigma_{HySt,N}^2(f) = \sigma_{HySt,lim}^2(f), \quad (59)$$

where  $\sigma_{HySt,lim}^2(f) = \frac{2}{45} H^5 \int_{\mathcal{T}} \frac{w^2(t) [x^{(2)}(t)]^2}{g^5(t)} dt$ .

Theorem 3 implies that uniform convergence rate for signals with nonzero second derivative cannot be faster than  $N^{-5}$ . By combining this observation with Theorem 2 we conclude that the uniform convergence rate of HySt estimators is exactly  $N^{-5}$ .

*Proof of Theorem 3:* We start with applying the mean value theorem to  $\lambda(t, f)$  and noting that for any  $f$  and  $t \in \mathcal{T}_{N,l}$  there exists  $\bar{t} \in \mathcal{T}_{N,l}$  such that

$$\lambda(t, f) = \lambda(c_{N,l}, f) + \lambda^{(1)}(\bar{t}, f)(t - c_{N,l}). \quad (60)$$

By substituting (60) in (50) and using (51), we express the estimation error by

$$\mathcal{E}_{N,l}(f) = \Gamma_{N,l}(f) + \Lambda_{N,l}(f), \quad (61)$$

where

$$\Gamma_{N,l}(f) \triangleq \left[ \frac{\Delta_{N,l}^3}{12} - \frac{\Delta_{N,l}}{2} z_l(\tau_l) \right] \lambda(c_{N,l}, f) x^{(2)}(c_{N,l}) \quad (62)$$

and

$$\Lambda_{N,l}(f) \triangleq \left[ \int_{\mathcal{T}_{N,l}} z_{N,l}(t) (t - c_{N,l}) \lambda^{(1)}(\bar{t}, f) dt - \Delta_{N,l} z_{N,l}(\tau_{N,l}) (\tau_{N,l} - c_{N,l}) \lambda^{(1)}(\bar{\tau}_{N,l}, f) \right] \frac{x^{(2)}(c_{N,l})}{2} + \bar{\chi}_l. \quad (63)$$

Since

$|\Gamma_{N,l}(f)| \leq \left[ \frac{\Delta_{N,l}^3}{12} + \frac{\Delta_{N,l}}{2} z_{N,l}(\tau_{N,l}) \right] |\lambda(c_{N,l}, f)| |x^{(2)}(c_{N,l})|$ , then by deploying (5), (7), (28) and (54), we get  $|\Gamma_{N,l}(f)| \leq \frac{5}{24} \Delta_{N,l}^3 x_{2,max} \leq N^{-3} \frac{45}{8} \frac{H^3 x_{2,max}}{g_{min}^3}$  and conclude

$$\Gamma_{N,l}(f) = O(N^{-3}). \quad (64)$$

Similarly,  $|\Lambda_{N,l}(f)| \leq \left[ \int_{\mathcal{T}_{N,l}} z_{N,l}(t) |t - c_{N,l}| |\lambda^{(1)}(\bar{t}, f)| dt + \Delta_{N,l} z_l(\tau_{N,l}) |\tau_{N,l} - c_{N,l}| |\lambda^{(1)}(\bar{\tau}_{N,l}, f)| \right] \frac{x^{(2)}(c_{N,l})}{2} + \bar{\chi}_l$  and since  $|\tau_{N,l} - c_{N,l}| \leq 0.5\Delta_{N,l}$ , we get from (28) and (56)  $|\Lambda_{N,l}(f)| \leq \frac{5}{64} \Delta_{N,l}^4 \lambda_{1,max} x_{2,max} \leq N^{-4} \frac{H^4}{g_{min}^4} \left( \frac{405}{64} \lambda_{1,max} x_{2,max} + \frac{135}{16} x_{3,max} \right)$ . Thus,

$$\Lambda_{N,l}(f) = O(N^{-4}). \quad (65)$$

Additionally,  $E\{|\mathcal{E}_{N,l}(f)|^2\} = E\{|\Gamma_{N,l}(f)|^2\} + E\{|\Lambda_{N,l}(f)|^2\} + 2\text{Re}[\Gamma_{N,l}^*(f)\Lambda_{N,l}(f)]$ . Utilizing (64) and (65), we obtain  $E\{|\mathcal{E}_{N,l}(f)|^2\} = E\{|\Gamma_{N,l}(f)|^2\} + O(N^{-7})$ . The variance (43) is  $\sigma^2\{\hat{X}_{HySt,N}(f)\} = \sum_{l=0}^{L_N-1} E\{|\Gamma_{N,l}(f)|^2\} + O(N^{-6})$ . Substituting this in (59) yields

$$\sigma_{HySt,lim}^2(f) = \lim_{N \rightarrow \infty} N^5 \sum_{l=0}^{L_N-1} E\{|\Gamma_{N,l}(f)|^2\} \quad (66)$$

By using (62) and (3) we get  $E\{|\Gamma_{N,l}(f)|^2\} = E\left\{\left[\frac{\Delta_{N,l}^3}{12} - \frac{\Delta_{N,l}}{2} z_{N,l}(\tau_{N,l})\right]^2\right\} |\lambda(c_{N,l}, f)|^2 [x^{(2)}(c_{N,l})]^2 = E\left\{\left[\frac{\Delta_{N,l}^3}{12} - \frac{\Delta_{N,l}}{2} z_{N,l}(\tau_{N,l})\right]^2\right\} w^2(c_{N,l}) [x^{(2)}(c_{N,l})]^2$ . It follows from (55)

that  $E\left\{\frac{\Delta_{N,l}}{2} z_{N,l}(\tau_{N,l})\right\} = \frac{\Delta_{N,l}^3}{12}$ . Therefore:  $E\left\{\left[\frac{\Delta_{N,l}^3}{12} - \frac{\Delta_{N,l}}{2} z_{N,l}(\tau_{N,l})\right]^2\right\} = \sigma^2 \left\{\frac{\Delta_{N,l}}{2} z_{N,l}(\tau_{N,l})\right\}$ . We note that  $\sigma^2 \left\{\frac{\Delta_{N,l}}{2} z_{N,l}(\tau_{N,l})\right\} = E\left\{\left[\frac{\Delta_{N,l}}{2} z_{N,l}(\tau_{N,l})\right]^2\right\} - \left(\frac{\Delta_{N,l}^3}{12}\right)^2$ , where  $E\left\{\left[\frac{\Delta_{N,l}}{2} z_{N,l}(\tau_{N,l})\right]^2\right\} = \frac{\Delta_{N,l}}{4} \int_{\mathcal{T}_{N,l}} z_{N,l}^2(t) dt$ . By (52), we get:  $E\left\{\left[\frac{\Delta_{N,l}}{2} z_{N,l}(\tau_{N,l})\right]^2\right\} = \frac{\Delta_{N,l}^6}{120}$ , and hence  $E\left\{\left[\frac{\Delta_{N,l}^3}{12} - \frac{\Delta_{N,l}}{2} z_{N,l}(\tau_{N,l})\right]^2\right\} = \frac{1}{720} \Delta_{N,l}^6$ . This implies

$$E\{|\Gamma_{N,l}(f)|^2\} = \frac{1}{720} \Delta_{N,l}^6 w^2(c_{N,l}) [x^{(2)}(c_{N,l})]^2 \quad (67)$$

By substituting (67) in (66) and using (17), we get

$$\begin{aligned} \lim_{N \rightarrow \infty} N^5 \sigma_{HySt,N}^2(f) &= \\ \frac{1}{720} \lim_{N \rightarrow \infty} N^5 \sum_{l=0}^{L_N-1} \Delta_{N,l}^6 w^2(c_{N,l}) [x^{(2)}(c_{N,l})]^2 &= \\ \frac{H^5}{720} \lim_{N \rightarrow \infty} \frac{N^5}{L_N^5} \sum_{l=0}^{L_N-1} \frac{w^2(c_{N,l}) [x^{(2)}(c_{N,l})]^2}{g^5(t_{h,l})} \Delta_{N,l} &= \\ \lim_{N \rightarrow \infty} N^5 / L_N^5 = 32. \quad \text{Hence by Riemann integral} &= \\ \lim_{N \rightarrow \infty} N^5 \sigma_{HySt,N}^2(f) = \frac{2}{45} H^5 \int_{\mathcal{T}} \frac{w^2(t) [x^{(2)}(t)]^2}{g^5(t)} dt, & \quad \text{which} \\ \text{completes the proof of Theorem 3.} & \end{aligned}$$

## V. NUMERICAL EXAMPLES

The numerical examples presented in this section compare the performances of ToRa, StSa, AnSt and HySt estimators. We consider the following signal  $x(t)$  comprising two spectral components centred around 2kHz and 70kHz

$$x(t) = A_1 \text{sinc}(B(t-d)) \cos(2\pi f_1(t-d)) + A_2 \cos(2\pi f_2 t) \quad (68)$$

where  $A_1 = 10^5$ ,  $B = 0.5\text{kHz}$ ,  $f_1 = 2\text{kHz}$ ,  $d = 7.5\text{ms}$ ,  $A_2 = 5 \times 10^3$  and  $f_2 = 70\text{kHz}$ . The length of the observation window  $\mathcal{T}$  is  $H = 15\text{ms}$ . Its Fourier Transform  $X(f)$  defined by (1) is calculated with the use of the Hanning window  $w(t) = 0.5 - 0.5 \cos(2\pi t/H)$ . It will be estimated using each of the four DASP methods discussed in this paper. For each of the three stratified estimators, we used  $g(t) = 1$ . Fig. 1 shows the magnitude of the target Fourier Transform  $X(f)$ .

In the first experiment we estimate the MSE defined by (2) by averaging the squared errors obtained from 1000 independent simulations. The results for ToRa, StSa, AnSt and HySt methods against the number of signal samples  $N$  are shown in Fig. 2, separately for frequencies 2, 70 and 160 kHz. These plots reveal that the MSE of HySt and ToRa estimators are nearly insensitive to the frequency for which the error is estimated. As previously explained, this is attributed to the fact that these estimators' pointwise and uniform convergence rates

are identical. On the other hand, for StSa and AnSt, the accelerated convergence rates become visible once the number of collected samples  $N$  is sufficiently large for the considered frequency. Hence, the estimation errors as functions of  $N$  are frequency-sensitive. Fig. 2a shows that at 2kHz, when  $N < 10^4$ , the estimators that use stratification exhibit quality similar to each other and notably better than that of ToRa. At higher frequencies, HySt significantly outperforms all other approaches. The plots for 70kHz presented in Fig. 2b show that when  $N < 1800$  the error for StSa and AnSt estimation is better aligned with that of ToRa than with their fast decay rates of  $N^{-3}$  and  $N^{-5}$ , respectively. Only when  $N$  exceeds 1800, the accelerated rates of StSa and AnSt become visible. Their sluggish behavior and inferior performance comparing to HySt become even more profound when the examined frequency  $f$  is further increased. The results for  $f = 160\text{kHz}$  in Fig. 2c show that the accelerated convergence of StSa and AnSt start when  $N > 3200$ . We also note that for  $N = 10,000$ , the AnSt and HySt exhibit the same performance at  $f = 2\text{kHz}$ . However, HySt outperforms AnSt by 20dB at 70kHz and by approximately 30dB at 160kHz.

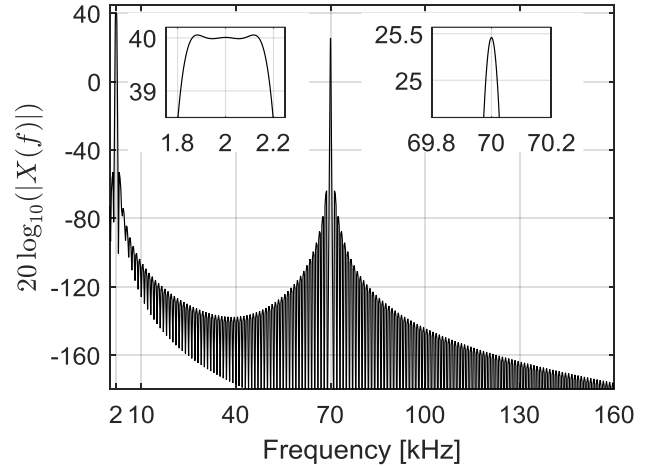


Fig. 1. Magnitude of the Fourier Transform of  $x(t)$  defined by (68). The insets show details of the results in the neighborhoods of 2kHz and 70kHz.

In the second experiment, we set the number of signal samples to  $N = 400$  (401 in the case of HySt) and run ten independent simulations for each of the four methods. The magnitudes of the estimated FT for ToRa, StSa and AnSt are shown in Fig. 3, whilst Fig. 4 presents the results for HySt. This experiment illustrates the opportunities and difficulties in detecting spectral components of the analyzed signals when using DASP-based FT estimators. The monitored frequency range is confined to  $[0, f_{max}]$ . Since there is no theoretical upper limit above which DASP approaches stop working, the frequency  $f_{max}$  is arbitrarily chosen as  $f_{max} = 160\text{kHz}$ . Examination of the depicted results confirms that all ten displayed estimates produced by the four estimators consistently reveal the presence of the low-frequency component centered at 2kHz. The differences between the estimates in the neighborhood of this frequency are bigger for the slowly converging ToRa than for the faster counterparts StSa, AnSt and HySt. However, at  $f = 70\text{kHz}$ , only HySt estimate exposes the presence of the second spectral

component. The estimation errors for ToRa, StSa and AnSt approaches are so big that they mask this component. By scrutinizing Fig. 2c and making crude analyses, it could be argued that ToRa, StSa and AnSt need around  $N = 1000$  signal samples to reveal the spectral component at 70 kHz. But in order to match the quality of HySt estimation at  $f = 70\text{kHz}$ ,

ToRa needs around 10,000 samples while StSa and AnSt need 3000 samples. For comparison, if uniform sampling was used to estimate the FT then the smallest number of collected samples that allows avoiding aliasing up to 160kHz is 4800. This is approximately twelve times more than what was used by the HySt estimator. If the FT were to be estimated with the same accuracy in a frequency range stretching beyond 160kHz the HySt approach can deliver these results without taking additional samples. However, uniform sampling would require increasing the density of samples and therefore collecting more data.

## VI. FINAL REMARKS AND CONCLUSIONS

We introduced the HySt approach for alias-free (DASP) FT estimation and showed that it outperforms by various measures its predecessors that tackle the same problem. In this section, we briefly discuss selected topics that could be of interest to potential users of HySt and other DASP approaches.

All four DASP estimators use “multiply-and-accumulate” process to estimate the FT. Therefore, once the complex-valued multipliers for the collected samples are known, each of these four methods takes  $2N$  multiplications and  $2N - 2$  additions to obtain the estimate at a single frequency point. It has been demonstrated that the HySt method often needs less signal samples to match or exceed the performance of the other three existing approaches. This potentially makes HySt the most computationally-efficient DASP solution for FT estimation. Similarly to ToRa, StSa and AnSt, the HySt multipliers can be calculated as soon as the weighting function  $w(t)$  and the sampling instants are known. The integrals needed in (30) and (32) to calculate the HySt multipliers  $\alpha_{N,l}(f)$  and  $\gamma_{N,l}(f)$  are independent from the random sampling instants and, thereby, can be pre-calculated even if these random instants are selected in real-time. Consequently, the workload related to incorporating the effect of the random sampling instants on the multipliers is more or less the same for all four DASP approaches.

Each of the four DASP Fourier transform estimates considered in this paper can be represented by  $\hat{X}_N(f) = X(f) + \Delta X_N(f)$ , where  $\Delta X_N(f)$  is a zero-mean random variable whose variance uniformly converges to zero when  $N$  goes to infinity. Thus, for a sufficiently large  $N$ , the estimated FT can be made arbitrarily similar to the  $X(f)$  target. This observation helps addressing the questions about the frequency resolution of DASP estimators as well as their ability of detecting weak spectral components in the analyzed signals. Such features are ultimately determined by  $X(f)$  defined by (1) rather than by the choice of a specific DASP method. The situation complicates when  $N$  is small and the increased variance of  $\Delta X_N(f)$  adds noise-like spectrum to  $X(f)$ . As a result, any less distinct features of  $X(f)$  can be obscured and even made invisible in  $\hat{X}_N(f)$ . In the second numerical example, the 70kHz component of the analyzed signal could not be detected by any DASP method, apart from the HySt estimator. For  $N \cong 400$ , only the variance of the HySt estimator was small enough to reveal specific features of  $X(f)$  at this frequency.

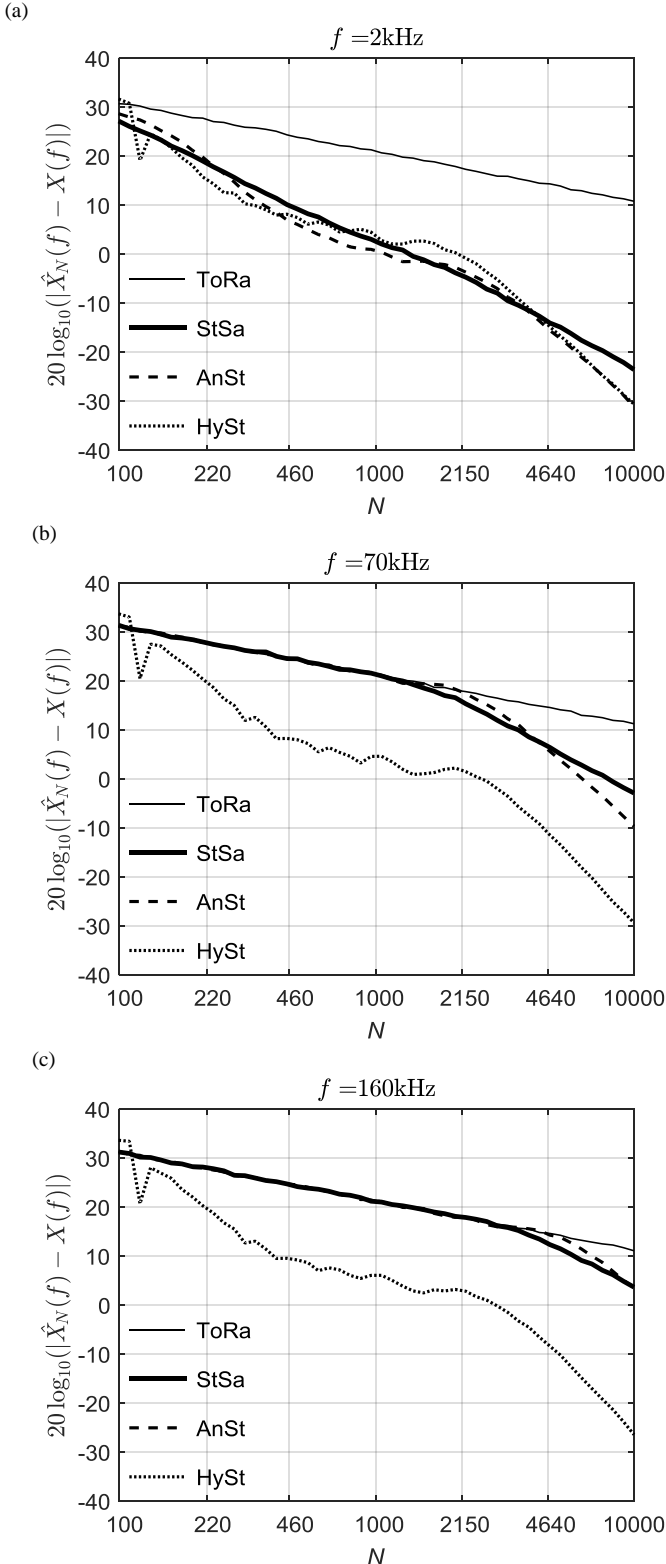


Fig. 2. MSE of the FT estimation as a function of the number of collected signal samples  $N$  for ToRa, StSa, AnSt and HySt at 2, 70 and 160kHz.

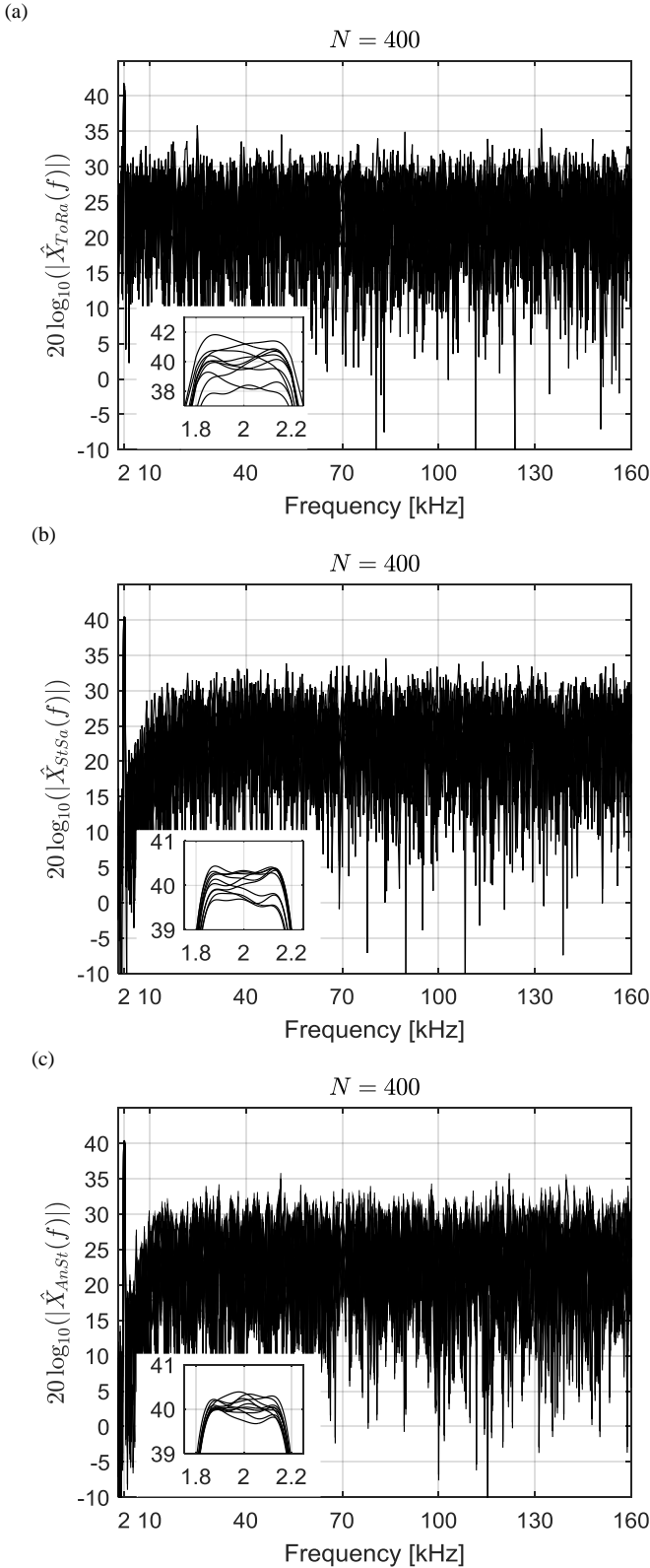


Fig. 3. The magnitude of the estimated Fourier Transform in 10 independent experiments for (a) ToRa, (b) StSa and (c) AnSt Fourier transforms. The insets show details of the results in the neighborhood of 2kHz.

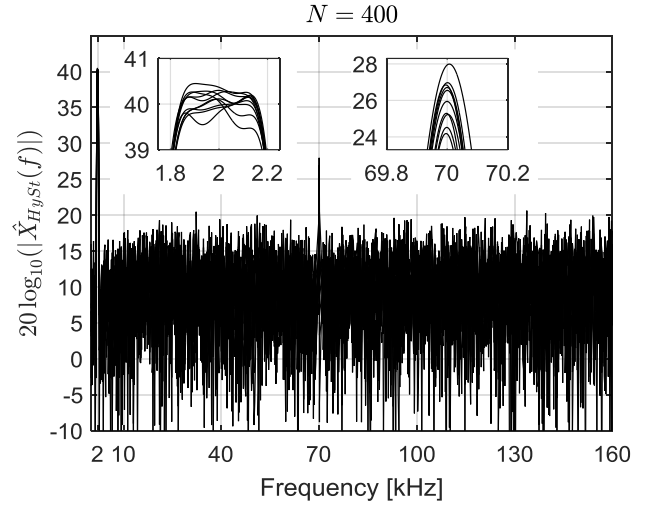


Fig. 4. The magnitude of the estimated Fourier Transform in 10 independent experiments with HySt estimator. The insets show details of the results in the neighborhoods of 2kHz and 70kHz.

The FT error estimation analyses are more complex if the signal samples are noisy. It is clear that such noise adds an extra layer of error to the estimates. Detailed analysis of how much error is added and a comparison of how different DASP approaches are affected is outside the scope of this paper.

An important question about the HySt FT estimator is whether its variance can be reduced by suitably selecting the stratifying function  $g(t)$ . The answer is yes, however, determining the optimal shape of  $g(t)$  requires solving a functional-analysis optimization problem that could be numerically difficult to tackle. A simplified closed form solution presented below minimizes the variance (59) subject to constraints (15) and  $g(t) \geq 0$ . We note that although the resultant  $g(t)$  is guaranteed to be non-negative it may not necessarily satisfy (14).

Based on the cost (59) and constraint (15), we form the following Lagrangian

$$L(g(\cdot), \xi) = \frac{2}{45} \frac{H^5}{N^5} \int_{\mathcal{T}} \frac{w^2(t) [x^{(2)}(t)]^2}{g^5(t)} dt + \xi \left[ H^{-1} \int_{\mathcal{T}} g(t) dt - 1 \right], \quad (69)$$

where  $\xi$  is the Lagrange multiplier. By equating the functional and partial derivatives of (69) with respect to  $g(t)$  and  $\xi$ , respectively, we get

$$\frac{\delta L(g(\tau), \xi)}{\delta g(t)} = -\frac{2}{9} \frac{H^5}{N^5} \frac{w^2(t) [x^{(2)}(t)]^2}{g^6(t)} + \xi H^{-1} = 0, \quad (70)$$

$$\frac{\partial L(g(\tau), \xi)}{\partial \xi} = H^{-1} \int_{\mathcal{T}} g(t) dt - 1 = 0. \quad (71)$$

It follows from (70) that  $g(t) = a^3 \sqrt{w(t) x^{(2)}(t)}$ , where  $a = H \times \sqrt[6]{\frac{2}{9\xi N^5}}$  is not fully known because of the dependence on  $\xi$ .

We calculate  $a$  by solving (71). Since  $\int_{\mathcal{T}} g(t) dt =$

$a \int_T \sqrt[3]{|w(t)x^{(2)}(t)|} dt$ , therefore  $a = \frac{H}{\int_T \sqrt[3]{|w(t)x^{(2)}(t)|} dt}$  and

the optimal  $g(t)$  is given by

$$g(t) = H \frac{\sqrt[3]{|w(t)x^{(2)}(t)|}}{\int_T \sqrt[3]{|w(t)x^{(2)}(t)|} dt}. \quad (72)$$

Since in typical cases  $x^{(2)}(t)$  is not known, a pragmatic solution is to choose  $g(t)$  that is proportional to  $\sqrt[3]{|w(t)|}$  and scaled so that (71) is satisfied.

Finally, it is important to mention that the “infinite bandwidth” of the DASP approaches could in practice be limited to a very wide but finite range of frequencies. First, if all the sampling instants are selected as multiples of some short time interval  $h$ , then, the alias-free analysis can be performed only up to  $0.5h^{-1}$  Hz. The second limiting factor, which was identified and analyzed in [23] with respect to ToRa, is the input clock jitter. All the analyses in this paper were made under the assumption that the sampling instants used in calculating the HySt multipliers are the same as when the signal samples were taken. However, any random error between them results in high-frequency bias of the FT estimates. A practical rule derived in [23] is that DASP should not be used for signal analyses above  $0.1\sigma^{-1}$  Hz, where  $\sigma$  is the standard deviation of the input-clock jitter.

## REFERENCES

- [1] J. D. Scargle, "Studies in Astronomical Time Series Analysis. II-Statistical Aspects of Spectral Analysis of Unevenly Spaced Data," *The Astrophysical Journal*, vol. 263, pp. 835-853, 1982.
- [2] S. Baishch and G. H. R. Bokelmann, "Spectral Analysis with Incomplete Time Series: An Example From Seismology," *Computers & Geosciences*, vol. 25, no. 7, pp. 739-750, 1999.
- [3] A. W. Liew, J. Xian, S. Wu, D. Smith and H. Yan, "Spectral Estimation in Unevenly Sampled Space of Periodically Expressed Microarray Time Series Data," *BMC Bioinformatics*, vol. 8, no. 1, pp. 137-156, 2007.
- [4] K. Kazmierczuk, W. Kozmiński and I. Zhukov, "Two-dimensional Fourier transform of arbitrarily sampled NMR data sets," *Journal of Magnetic Resonance*, vol. 179, no. 2, pp. 323-328, 2006.
- [5] E. Axell, G. Leus, E. G. Larsson and V. H. Poor, "Spectrum Sensing for Cognitive Radio: State-of-the-art and Recent Advances," *IEEE Signal Process. Mag.*, vol. 29, pp. 101-116, 2012.
- [6] G. Hattab, Ibnkahla and Mohamed, "Multiband Spectrum Access: Great Promises for Future Cognitive Radio Networks," *Proc. IEEE*, vol. 102, pp. 282-306, 2014.
- [7] R. G. Vaughan, N. L. Scott and D. R. White, "The theory of bandpass sampling," *IEEE Trans. Signal Process.*, vol. 39, pp. 1973-1984, 1991.
- [8] H. J. Landau, "Necessary density conditions for sampling and interpolation of certain entire functions," *Acta Mathematica*, vol. 117, no. 1, pp. 37-52, 1967.
- [9] R. Venkataramani and B. Yoram, "Perfect reconstruction formulas and bounds on aliasing error in sub-Nyquist nonuniform sampling of multiband signals," *IEEE Trans. Inf. Theory*, vol. 46, no. 6, pp. 2173-2183, 2000.
- [10] P. Feng and Y. Bresler, "Spectrum-blind minimum-rate sampling and reconstruction of multiband signals," in *IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP '96)*, May 1996, vol. 3, pp.1688-1691.
- [11] M. F. Duarte and Y. C. Eldar, "Structured Compressed Sensing: From Theory to Applications," *IEEE Trans. Signal Process.*, vol. 59, pp. 4053-4085, 2011.
- [12] D. L. David, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289-1306, 2006.
- [13] M. Mishali and Y. Eldar, "Blind Multiband Signal Reconstruction: Compressed Sensing for Analog Signals," *IEEE Trans. Signal Process.*, vol. 7, pp. 993 - 1009, 2009.
- [14] M. Vetterli, P. Marziliano and T. Blu, "Sampling signals with finite rate of innovation," *IEEE Trans. Signal Process.*, vol. 50, no. 6, pp. 1417-1428, 2002.
- [15] H. S. Shapiro and R. A. Silverman, "Alias-free sampling of random noise," *Journal of the Society for Industrial & Applied Mathematics*, vol. 8, no. 2, pp. 225-248, 1960.
- [16] E. Masry, "Alias free sampling: An alternative conceptualization and its applications," *IEEE Trans. Inf. Theory*, Vols. IT-24, p. 317-324, 1978.
- [17] I. Bilinskis and A. Mikelsons, *Randomized Signal Processing*, Prentice-Hall, 1992.
- [18] R. J. Martin, *Irregular Sampled Signals: Theories and Techniques for Analysis*, PhD Thesis, University College London, 1998.
- [19] F. Marvasti, *Nonuniform Sampling Theory and Practice*, New York: Kluwer Academic, 2001.
- [20] I. Bilinskis, *Digital Alias-free Signal Processing*, New York: John Wiley and Sons, 2007.
- [21] M. Al-Ani, *Multidimensional Random Sampling for Fourier Transform Estimation*, Doctoral Dissertation, University of Westminster, 2013.
- [22] P. Babu and P. Stoica, "Spectral analysis of nonuniformly sampled data – a review," *Digital Signal Processing*, p. 359-378, 2010.
- [23] A. Tarczynski and N. Allay, "Spectral Analysis of Randomly Sampled Signals: Suppression of Aliasing and Sampler Jitter," *IEEE Trans. Signal Proc.*, vol. 52, no. 12, pp. 3324-3334, December 2004.
- [24] A. Tarczynski and D. Qu, "Optimal Random Sampling for Spectrum Estimation in DASP Applications," *Int. Journal of Applied Mathematics and Computer Science*, vol. 14, no. 4, p. 463 - 469, 2005.
- [25] E. Masry, "Random Sampling of Deterministic Signals: Statistical Analysis of Fourier Transform Estimates," *IEEE Trans. Signal Process.*, vol. 54, no. 5, pp. 1750-1761, 2006.
- [26] E. Masry and A. Vadrevu, "Random Sampling Estimates of Fourier Transforms: Antithetical Stratified Monte Carlo," *IEEE Trans. Signal Process.*, vol. 57, no. 1, pp. 194 - 204, 2009.
- [27] B. I. Ahmad and A. Tarczynski, "A SARS Method for Reliable Spectrum Sensing in Multiband Communication Systems," *IEEE Trans. on Signal Processing*, vol. 59, no. 12, pp. 6008 - 6020, 2011.
- [28] A. Antoniou, *Digital Signal Processing: Signals, Systems, and Filters*, McGraw-Hill Professional, 2005.
- [29] F. J. Harris, "On the Use of Windows for Harmonic Analysis with the Discrete Fourier Transform," *Proc. IEEE*, vol. 66, no. 1, pp. 51-83, January 1978.



**Andrzej Tarczynski** is a Reader in Signal Processing and Control Systems at the University of Westminster, London, UK. He received his MEng in Electrical Engineering and PhD in Automatic Control from the Warsaw University of Technology, Poland in 1979 and 1986 respectively. Subsequently he worked in academia (Warsaw University of Technology) and industry (Computex Ltd., POSTINFO Ltd.). In 1991 he joined the University of Westminster, where he has occupied a number of academic and managerial positions, including Director of Research and Knowledge Transfer in the School of Electronics and Computer Science. His research interests include alias-free signal processing, optimization techniques and algorithms for signal and system analysis, and control systems.



**Bashar I. Ahmad** received the B.Eng. (Hons.) in Electronic Engineering and Ph.D. in statistical signal processing degrees from the University of Westminster, U.K, in 2007 and 2011, respectively. He is currently a Senior Research Associate in the Signal Processing and Communications Laboratory, Engineering Department, Cambridge University, U.K. Prior to joining the University of Cambridge, he was a postdoctoral researcher in the Signal Processing and Communications group at Imperial College London, U.K. His research interests include statistical signal processing, multi-modal human computer interactions, sub-Nyquist signal processing, sampling theory and cognitive radio.