

Cornerstones of Sampling of Operator Theory

David Walnut, Götz E. Pfander, Thomas Kailath

Abstract This paper reviews some results on the identifiability of classes of operators whose Kohn-Nirenberg symbols are band-limited (called *band-limited operators*), which we refer to as *sampling of operators*. We trace the motivation and history of the subject back to the original work of the third-named author in the late 1950s and early 1960s, and to the innovations in spread-spectrum communications that preceded that work. We give a brief overview of the NOMAC (Noise Modulation and Correlation) and Rake receivers, which were early implementations of spread-spectrum multi-path wireless communication systems. We examine in detail the original proof of the third-named author characterizing identifiability of channels in terms of the maximum time and Doppler spread of the channel, and do the same for the subsequent generalization of that work by Bello. The mathematical limitations inherent in the proofs of Bello and the third author are removed by using mathematical tools unavailable at the time. We survey more recent advances in sampling of operators and discuss the implications of the use of periodically-weighted delta-trains as identifiers for operator classes that satisfy Bello's criterion for identifiability, leading to new insights into the theory of finite-dimensional Gabor systems. We present novel results on operator sampling in higher dimensions, and review implications and generalizations of the results to stochastic operators, MIMO systems, and operators with unknown spreading domains.

David Walnut
George Mason University, Fairfax, Virginia, USA e-mail: dwalnut@gmu.edu

Götz E. Pfander
Jacobs University, Bremen, Germany, e-mail: g.pfander@jacobs-university.de

Thomas Kailath
Stanford University, California, USA e-mail: kailath@stanford.edu

The inversion of the traditional alphabetical ordering of authors is at the suggestion of the third author, who desires that those at the end of the alphabet get some recognition.

1 Introduction

The problem of identification of a time-variant communication channel arose in the 1950s as the problem of secure long-range wireless communications became increasingly important due to the geopolitical situation at the time. Some of the theoretical and practical advances made then are described in this paper, and more recent advances extending the theory to more general operators, and onto a more rigorous mathematical footing, known as *sampling of operators* are developed here as well.

The launching point for the theory of operator sampling is the early work of the third-named author in his Master's thesis at MIT, entitled "Sampling models for linear time-variant filters" [21], see also [24, 25], and [23] in which he reviews the identification problem for time-variant channels. The third named author as well as Bello in subsequent work [5] were attempting to understand and describe the theoretical limits of identifiability of time-variant communication channels. Section 2 of this paper describes in some detail their work and explores some of the important mathematical challenges they faced. In Section 3, we describe the more recently developed framework of operator sampling. Results addressing the problem considered by Bello are based on insights on finite dimensional Gabor systems which are presented in Section 4. Malikiosis's recent result [34] allows for the generalization of those results to a higher-dimensional setting, these are stated and proven in Section 5. We conclude the paper in Section 6 with a short summary of the sampling of operators literature, that is, of results presented in detail elsewhere.

2 Historical Remarks.

2.1 *The Cold War Origins of the Rake System.*

In 1958, Price and Green published *A Communication Technique for Multi-path Channels* in Proc. IRE [57], in which they describe a communication system called Rake, designed to solve the *multi-path problem*. When a wireless transmitter does not have line-of-sight with the receiver, the transmitted signal is reflected possibly multiple times before reaching the receiver. Reflection by stationary objects such as the ground or buildings introduces random time delays to the signal, and reflection or refraction by moving objects such as clouds, the troposphere, ionosphere, or a moving vehicle produce random frequency or Doppler shifts in the signal as well. Due to scattering and absorption, the reflected signals are randomly amplitude-attenuated too. The problem is to recover the transmitted signal as accurately as possible from the superposition of time-frequency-shifted and randomly amplitude-attenuated versions of it. Since the location and velocities of the reflecting objects change

with time, the effects of the unknown, time-variant channel must be estimated and compensated for.

Price and Green's paper [57] was the disclosure in the literature of a long-distance system of wide-band or spread-spectrum communications that had been developed in response to strategic needs related to the Cold War. This fascinating story has been described in several articles by those directly involved ([63, 64, 59, 12]). We present a summary of those remarks and of the Rake system below. The goal is to motivate the original work of the third-named author on which the theory of operator sampling is based.

In the years following World War II, the Soviet Union was exercising its power in Eastern Europe with a major point of contention being Berlin, which the Soviets blockaded in the late 1940s. This made secure communication with Berlin a top priority. As Paul Green describes it,

[T]he Battle of Berlin was raging, the Russians having isolated the city physically on land, so that the Berlin Airlift was resorted to, and nobody knew when all the communication links would begin to be subjected to heavy Soviet jamming. [12]

By 1950, with a shooting war in Korea about to begin, the Army Signal Corps approached researchers at MIT to develop secure, and reliable wireless communication with the opposite ends of the world. According to Green,

It is difficult today to recall the fearful excitement of those times. The Russians were thought to be 12 feet high in anything having to do with applying mathematics to communication problems ("all Russians were Komogorovs or Kotelnikovs")....[T]here was a huge backlog of unexploited theory lying around, and people were beginning to build digital equipment with the unheard of complexity of a hundred or so vacuum tube-based bits (!). And the money flowed. [12]

The effort was called Project Lincoln (precursor to Lincoln Laboratory). The researchers were confronted by two main problems: 1) making a communications system robust to noise and deliberate jamming, and 2) enabling good signal recovery from multiple paths.

2.2 Spread Spectrum communications and NOMAC

The technique chosen to address the first problem is an application of the notion, already well-understood and used by that time, that combatting distortions from noise and jamming can be achieved by spreading the signal over a wide frequency band. The idea of spreading the spectrum had been around for a long time [59, 69, 56] and can be found even in a now famous Hedy Lamarr-George Antheil patent of 1942 [35, 59], which introduced the concept later called "frequency hopping". The system called NOMAC (Noise Modulation and Correlation) was developed in the early 1950s and used noise like (pseudo-noise or PN) signals to achieve spectrum spreading. Detailed discussion of its history can be found in [59, 12, 70].

The huge backlog of “unexploited theory” mentioned above included the recent work of Claude Shannon on communication theory [67], of Norbert Wiener on correlation functions and least mean squares prediction and filtering [71], and recent applications of statistical decision theory to detection problems in radar and communications.

The communication problem addressed by NOMAC was to encode data represented by a string of ones and zeros into analog signals that could be electromagnetically transmitted over a noisy communication channel in a way that foiled “jamming” by enemies. The analog signals $x_1(\cdot)$ and $x_0(\cdot)$, commonly called Mark and Space, associated with the data digits 1 and 0, were chosen to be waveforms of approximate bandwidth B , and with small cross correlation. The target application was 60 wpm teletype, with 22 msec per digit (called a baud), which corresponds to a transmission rate of $1/0.022 \text{ sec} = 45 \text{ Hz}$. The transmitted signals were chosen to have a bandwidth of 10 KHz, which was therefore expected to yield a “jamming suppression ration” of $10,000/45 = 220$, or 23 db [12, 70]. The jamming ratio is often called the “correlation gain”, because the receiver structure involves cross correlation of the received signal with each of the possible transmitted signals. If the correlation with the signal $x_1(\cdot)$ is larger than the one with the signal $x_0(\cdot)$, then it is decided that the transmitted signal corresponded to the digit 1. This scheme can be shown to be optimum in the sense of minimum probability of error provided that the transmitted signals are not distorted by the communications channel and that the receiver noise is white Gaussian noise (see, for example, [17]). The protection against jamming is because unless the jammer has good knowledge of the noise like transmitted signals, any jamming signals would just appear as additional noise at the output of the crosscorrelations.

More details on the nontrivial ideas required for building a practical system can be found in the references. We may mention that the key ideas arose from three classified MIT dissertations by Basore [4], Pankowski [38], and Green [10], in fact, documents on NOMAC remained classified until 1961 [12].

A transcontinental experiment was run on a NOMAC system, but was found to have very poor performance because of the presence of multiple paths; the signals arriving at the receiver by these different paths sometimes interfere destructively. This is the phenomenon of “fading”, which causes self jamming of the system. Some improvement was achieved by adding additional circuitry and the receiver to separately identify and track the two strongest signals and combine them after phase correction; this use of time and space diversity enabled a correlation gain of 17 db, 6 db short of the expected performance. It was determined that this loss was because of the neglected weaker paths, of which there could be as many as 20 or 30. So attention turned to a system that would allow the use of all the different paths.

2.3 The Rake system

One conceptual basis for this new system was provided by the doctoral thesis of Robert Price [60], the main results of which were published in 1956 [61]. In a channel with severe multi-path the signal at the receiver is composed of large number of signals of different amplitudes and phases and so Price made the assumption that the received “signal” was a Gaussian random process. He studied the problem of choosing between the hypothesis

$$H_i : w(\cdot) = Ax_i(\cdot) + n(\cdot), \quad i = 0, 1,$$

where the random time variant linear communication channel A is such that the $\{Ax_i(\cdot)\}$ are Gaussian processes. In this case, the earlier cross correlation detection scheme makes no sense, because the “signal” arriving at the receiver is not deterministic but is a sample function of a random process, which is not available to the receiver because it is corrupted by the additive noise. Price worked out the optimum detection scheme and then ingeniously interpreted the mathematical formulas to conclude that the new receiver forms least mean-square estimates of the $\{Ax_i(\cdot)\}$ and then crosscorrelates the $w(\cdot)$ against these estimates. In practice of course, one does not have enough statistical information to form these estimates and therefore more heuristic estimates are used and this was done in the actual system that was built. The main heuristic, from Wiener’s least mean-square smoothing filter solution and earlier insights, is that one should give greater weight to paths with higher signal-to-noise ratio.

So Price and Green devised a new receiver structure comprised of a delay line of length 3 ms intervals (the maximum expected time spread in their channel), with 30 taps spaced every 1/10 KHz, or 100 μ s. This would enable the capture of all the multi-path signals in the channel. Then the tap gains were made proportional to the strength of the signal received at that tap. Since a Mark/Space decision was only needed every 22 ms (for the transmission rate of 60 wpm), and since the fading rate of the channel was slow enough that the channel characteristics remain constant over even longer than 22 ms, tap gains could be averaged over several 3 ms intervals. The new system was called “Rake”, because the delay line structure resembled that in a typical garden rake!

Trials showed that this scheme worked well enough to recover the 6 db loss experienced by the NOMAC system. The system was put into production and was successfully used for jam-proof communications between Washington DC and Berlin during the “Berlin crisis” in the early 60s.

HF communications is no longer very significant, but the Rake receiver has found application in a variety of problems such as sonar, the detection of underground nuclear explosions, and planetary radar astronomy (pioneered by Price and Green, [11, 58]) and currently it is much used in mobile wireless communications. It is interesting to note that the eight racks of equipment

needed to build the Rake system in the 1960s is now captured in a small integrated circuit chip in a smart phone!

However the fact that the Rake system did not perform satisfactorily when the fading rates of the communication channel were not very slow led MIT professor John Wozencraft, (who had been part of the Rake project team at Lincoln Lab) to suggest in 1957 (even before the open 1958 publication of the Rake system) to his new graduate student Thomas Kailath a fundamental study of linear time-variant communication channels and their identifiability for his Masters thesis. While linear time-variant linear systems had begun to be studied at least as early as 1950 (notably by Zadeh [72]), in communication systems there are certain additional constraints, notably limits on the bandwidths of the input signal and the duration of the channel memory. So a more detailed study was deemed to be worthwhile.

2.4 *Kailath's Time-Variant Channel Identification Condition*

In the paper [21], the author considers the problem of measuring a channel whose characteristics vary rapidly with time. He considers the dependence of any theoretical channel estimation scheme on how rapidly the channel characteristics change and concludes that there are theoretical limits on the ability to identify a rapidly changing channel. He models the channel A as a linear time-variant filter and defines

$A(\lambda, t)$ = response of A , measured at time t to a unit impulse input at time $t - \lambda$.

$A(\lambda, t)$ is one form of the time-variant impulse response of the linear channel that emphasizes the role of the “age” variable λ . The channel response to an input signal $x(\cdot)$ is

$$Ax(t) = \int A(\lambda, t) x(t - \lambda) d\lambda.$$

An impulse response $A(\lambda, t) = A(\lambda)$ represents a time-invariant filter. Further, the author states

Therefore the rate of variation of $A(\lambda, t)$ with t , for fixed λ , is a measure of the rate of variation of the filter. It is convenient to measure this variation in the frequency domain by defining a function \mathcal{A}

$$\mathcal{A}(\lambda, f) = \int_{-\infty}^{\infty} A(\lambda, t) e^{-2\pi i f t} dt$$

Then he defines

$$B = \max_{\lambda} [b - a, \text{ where } \mathcal{A}(\lambda, f) = 0 \text{ for } f \notin [a, b]].$$

While symmetric support is assumed in the paper, this definition makes clear that non-rectangular regions of support are already in view. Additionally, he defines the memory as the maximum time-delay spread in response to an impulse of the channel as

$$L = \max_t [\min_{\lambda'} \text{ such that } A(\lambda, t) = 0, \lambda \geq \lambda'].$$

In short, the assumption in the continuation of the paper is that

$$\text{supp } \mathcal{A}(\lambda, f) \subseteq [0, L] \times [-W, W]$$

where $W = B/2$. The function $\mathcal{A}(\lambda, f)$ is often called the *spreading function* of the channel. He then asks under what assumptions on L and $B = 2W$ can such a channel be measured? In the context of the Rake system, this translates to the question of whether there are limits on the rate of variation of the filter that can assure that the measurement filter can be presumed to be effective.

The author's assertion is that as long as $BL \leq 1$, then a "simple measurement scheme" is sufficient.

We have assumed that the bandwidth of any "tap function", $A_\lambda(\cdot) [= A(\lambda, \cdot)]$, is limited to a frequency region of width B , say a low-pass region $(-W, W)$ for which $B = 2W$. Such band-limited taps are determined according to the Sampling theorem, by their values at the instants $i/2W$, $i = 0, \pm 1, \pm 2, \dots$

If the memory, L , of the filter, $A(\lambda, t)$ is less than $1/2W$ these values are easily determined: we put in unit impulses to $A(\lambda, t)$ at instants $0, 1/2W, 2/2W, \dots, T$, and read off from the responses the desired values of the impulse response $A(\lambda, t)$. [...] If $L \leq 1/2W$, the responses to the different input impulses do not interfere with one another and the above values can be unambiguously determined.

In other words, sufficiently dense samples of the tap functions can be obtained by sending an impulse train $\sum_n \delta_{n/2W}$ through the channel. Indeed,

$$A\left(\sum_n \delta_{n/2W}\right)(t) = \sum_n \int A(\lambda, t) \delta_{n/2W}(t - \lambda) d\lambda = \sum_n A(t - n/2W, t).$$

Evaluating the operator response at $t = \lambda_0 + n_0/2W$, $n_0 \in \mathbb{Z}$, we obtain

$$\begin{aligned} A\left(\sum_n \delta_{n/2W}\right)(\lambda_0 + n_0/2W) &= \sum_n A(\lambda_0 + (n_0 - n)/2W, \lambda_0 + n_0/2W) \\ &= A(\lambda_0, \lambda_0 + n_0/2W) \end{aligned}$$

since $L \leq 1/2W$ implies that $A(\lambda_0 + (n_0 - n)/2W, \lambda_0 + n_0/2W) = 0$ if $n \neq n_0$. In short, for each λ , the samples $A(\lambda, \lambda + n/2W)$ for $n \in \mathbb{Z}$ can be recovered.

The described Kailath sounding procedure is depicted in Figure 1. In this visualization, we plot the kernel $\kappa(s, t) = A(t - s, t)$ of the operator A , that is,

$$Ax(t) = \int A(\lambda, t) x(t - \lambda) d\lambda = \int A(t - s, t) x(s) ds = \int \kappa(t, s) x(s) ds.$$

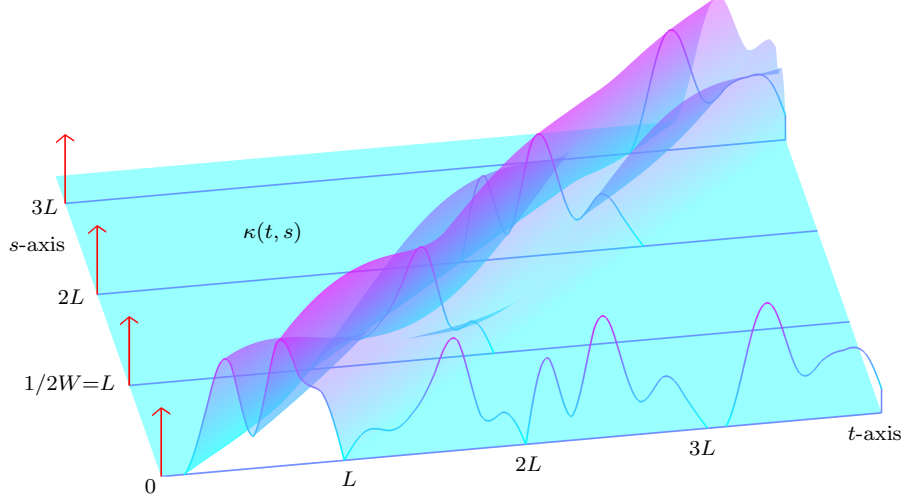


Fig. 1: Kailath sounding of A with $\text{supp } \mathcal{A}(\lambda, f) \subseteq [0, L] \times [-W, W]$ and $L = 1/2W$. The kernel $\kappa(t, s)$ is displayed on the (t, s) plane, the impulse train $\sum_n \delta_{n/2W}(s)$ on the s -axis, and the output signal $Ax(t) = A(\sum_n \delta_{n/2W})(t) = \sum_n A(t - n/2W, t) = \sum_n \kappa(t, n/2W)$. The sample values of the tab functions $A_\lambda(t) = A(\lambda, t) = \kappa(t, t - \lambda)$ can be read off $Ax(t)$.

2.5 Necessity of Kailath's Condition for Channel Identification.

For the “simple measurement scheme” to work, $BL \leq 1$ is sufficient but could be restrictive.

We need, therefore, to devise more sophisticated measurement schemes. However, we have not pursued this question very far because for a certain class of channels we can show that the condition

$$L \leq 1/2W, \text{ i.e. }, BL \leq 1$$

is necessary as well as sufficient for unambiguous measurement of $A(\lambda, t)$. The class of channels is obtained as follows: We first assume that there is a bandwidth constraint on the possible input signals to $A(\lambda, t)$, in that the signals are restricted to $(-W_i, W_i)$ in frequency. We can now determine a filter $A_{W_i}(\lambda, t)$ that is equivalent to $A(\lambda, t)$ over the bandwidth $(-W_i, W_i)$, and find necessary and sufficient

conditions for unambiguous measurement of $A_{W_i}(\lambda, t)$. If we now let $W_i \rightarrow \infty$, this condition reduces to condition (1), viz: $L \leq 1/2W$. Therefore, condition (1) is valid for all filters $A(\lambda, t)$ that may be obtained as the limit of band-limited channels. This class includes almost all filters of physical interest. The argument is worked out in detail in Ref. 6¹ but we give a brief outline here.

The class of operators in view here can be described as limits (in some unspecified sense) of operators whose impulse response $A(\lambda, t)$ is bandlimited to $[-W_i, W_i]$ in λ for each t and periodic with period $T > 0$ in t for each λ . Here, T is assumed to have some value larger than the maximum time over which the channel will be operated. We could take it as the duration of the input signal to the channel.

The restriction to input signals bandlimited $(-W_i, W_i)$ indicates that it suffices to know the values of $A(\lambda, t)$ or $\mathcal{A}(\lambda, f)$ for a finite set of values of λ : $\lambda = 0, 1/2W_i, 2/2W_i, \dots, L$, assuming for simplicity that L is a multiple of $1/2W_i$. Therefore, we can write

$$A(\lambda, t) = \sum_n A(n/2W_i, t) \operatorname{sinc}_{W_i}(\lambda - n/2W_i),$$

where $\operatorname{sinc}_{W_i}(t) = \sin(2\pi W_i t)/(2\pi W_i t)$ so that as $W_i \rightarrow \infty$, $\operatorname{sinc}_{W_i}(t)$ becomes more concentrated at the origin.

Also, T -periodicity in t allows us to write

$$A(\lambda, t) = \sum_k A(\lambda, k/T) e^{2\pi i k t / T},$$

so that combining gives

$$A(\lambda, t) = \sum_n \sum_k A(n/2W_i, k/T) \operatorname{sinc}_{W_i}(\lambda - n/2W_i) e^{2\pi i k t / T}.$$

Based on the restriction to bandlimited input signals which are T periodic, we have obtained a representation of A which is neither compactly supported in λ nor bandlimited in t . However, the original restriction that

$$\operatorname{supp} \mathcal{A}(\lambda, f) \subseteq [0, L] \times [-W, W]$$

motivates the assumption that we are working with finite sums, viz.

$$A(\lambda, t) = \sum_{n/2W_i \in [0, L]} \sum_{k/t \in [-W, W]} A(n/2W_i, k/T) \operatorname{sinc}_{W_i}(\lambda - n/2W_i) e^{2\pi i k t / T}.$$

This is how the author obtains the estimate that there are at most $(2W_i L + 1)(2WT + 1)$ degrees of freedom in any impulse response A in the given class.

For any input signal $x(t)$ bandlimited to $[-W_i, W_i]$, the output will be bandlimited to $[-W - W_i, W + W_i]$. Specifically,

¹ Ref. 6 is [21].

$$\begin{aligned}
Ax(t) &= \int A(\lambda, t) x(t - \lambda) d\lambda \\
&= \sum_{n/2W_i \in [0, L]} \sum_{k/t \in [-W, W]} A(n/2W_i, k/T) e^{2\pi i kt/T} \\
&\quad \int x(t - \lambda) \operatorname{sinc}_{W_i}(\lambda - n/2W_i) d\lambda \\
&= \sum_{n/2W_i \in [0, L]} \sum_{k/t \in [-W, W]} A(n/2W_i, k/T) e^{2\pi i kt/T} \\
&\quad (x * \operatorname{sinc}_{W_i})(t - n/2W_i).
\end{aligned}$$

Since $e^{2\pi i kt/T} (x * \operatorname{sinc}_{W_i})(t - n/2W_i)$ is bandlimited to $[-W_i, W_i] + (k/T)$ for $k/T \in [-W, W]$, it follows that $Ax(t)$ is bandlimited to $[-W - W_i, W + W_i]$.

If we restrict our attention to signals $x(t)$ time-limited to $[0, T]$, the output signal $Ax(t)$ will have duration $T + L$, and $Ax(\cdot)$ will be completely determined by its samples at $\frac{n}{2(W + W_i)} \in [0, T + L]$, from which we can identify $2(T + L)(W + W_i) + 1$ degrees of freedom.

In order for identification to be possible, the number of degrees of freedom of the output signal must be at least as large as the number of degrees of freedom of the operator, i.e.

$$\begin{aligned}
2W_i T + 2W_i L + 2WT + 2WL + 1 &= \\
2(T + L)(W_i + W) + 1 &\geq (2WT + 1)(2W_i L + 1) \\
&= 2WT + 2W_i L + 1 + 4W_i WTL
\end{aligned}$$

which reduces ultimately to

$$\frac{1}{1 - 1/(2W_i T)} \geq 2WL = BL.$$

That is, BL needs to be strictly smaller than 1 in the approximation while $BL = 1$ may work in the limiting case $W_i \rightarrow \infty$ (and/or $T \rightarrow \infty$).

This result got a lot of attention because it corresponded with experimental evidence that Rake did not function well when the condition $BL < 1$ was violated. It led to the designation of “underspread” and “overspread” channels for which BL was less than or greater than 1.

2.6 Some Remarks on Kailath's Results

This simple argument is surprising, particularly in light of the fact that the author obtained a deep result in time-frequency analysis with none of the tools of modern time-frequency analysis at his disposal. He very deftly uses the extremely useful engineering “fiction” that the dimension of the space

of signals essentially bandlimited to $[-W, W]$ and time-limited to $[0, T]$ is approximately $2WT$. The then recent papers of Landau, Slepian and Pollak [30, 31], which are mentioned explicitly in [21], provided a rigorous mathematical framework for understanding the phenomenon of essentially simultaneous band- and time-limiting. While the existence of these results lent considerable mathematical heft to the argument, they were not incorporated into a fully airtight mathematical proof of his theorem.

In the proof we have used a degrees-of-freedom argument based on the sampling theorem which assumes strictly bandlimited functions. This is an unrealistic assumption for physical processes. It is more reasonable to call a process band (or time) limited if some large fraction of its energy, say 95%, is contained within a finite frequency (or time) region. Recent work by Landau and Slepian has shown the concept of approximately $2TW$ degrees of freedom holds even in such cases. This leads us to believe that our proof of the necessity of the $BL \leq 1$ condition is not merely a consequence of the special properties of strictly band-limited functions. It would be valuable to find an alternative method of proof.

While Kailath's Theorem is stated for channel operators whose spreading functions are supported in a rectangle, it is clear that the later work of Bello [5] was anticipated and more general regions were in view. This is stated explicitly.

We have not discussed how the bandwidth, B is to be defined. There are several possibilities: we might take the nonzero f -region of $\mathcal{A}(\lambda, f)$; or use a "counting" argument. We could proceed similarly for the definition of L . As a result of these several possibilities, the value 1, of the threshold in the condition $BL \leq 1$ should be considered only as an order of magnitude value.

...constant and predictable variations in B and L , due for example to known Doppler shifts or time displacements, would yield large values for the absolute values of the time and frequency spreadings. However such predictable variations should be subtracted out before the BL product is computed; *what appears to be important is the area covered in the time- and frequency-spreading plane rather than the absolute values of B and L .* (emphasis added)

The reference to "counting" as a definition of bandwidth clearly indicates that essentially arbitrary regions of support for the operator spreading function were in view here, and that a necessity argument relying on degrees of freedom and not the shape of the spreading region was anticipated. The third-named author did not pursue the measurement problem studied in his MS thesis because he went on in his PhD dissertation to study the optimum (in the sense of minimum probability of error) detector scheme of which Rake is an intelligent engineering approximation. See [22, 23, 25].

The mathematical limitations of the necessity proof in [21] can be removed by addressing the identification problem directly as a problem on infinite-dimensional space rather than relying on finite-dimensional approximations to the channel. This approach also avoids the problem of dealing with simultaneously time and frequency-limited functions. In this way, the proof can be made completely mathematically rigorous. This approach is described in Section 3.2.

2.7 Bello's time-variant Channel Identification Condition

Kailath's Theorem was generalized by Bello in [5] along the lines anticipated in [21]. Bello's argument follows that of [21] in its broad outlines but with some significant differences. Bello clearly anticipates some of the technical difficulties that have been solved more recently by the authors and others and which have led to the general theory of operator sampling.

Continuing with the notation of this section, Bello considers channels with spreading function $\mathcal{A}(\lambda, f)$ supported in a rectangle $[0, L] \times [-W, W]$. If L and W are all that is known about the channel, then Kailath's criterion for measurability requires that $2WL \leq 1$. Bello considers channels for which $2WL$ may be greater than 1 but for which

$$S_A = |\text{supp } \mathcal{A}(\lambda, f)| \leq 1$$

and argues that this is the most appropriate criterion to assess measurability of the channel modeled by A .

In order to describe Bello's proof we will fix parameters $T \gg L$ and $W_i \gg W$ and following the assumptions earlier in this section, assume that inputs to the channel are time-limited to $[0, T]$ and (approximately) bandlimited to $[-W_i, W_i]$. Under this assumption, Bello considers the spreading function of the channel to be approximated by a superposition of point scatterers, viz.

$$\mathcal{A}(\lambda, f) = \sum_n \sum_k A_{n,k} \delta(f - (k/T)) \delta(\lambda - (n/2W_i)).$$

Hence the response of the channel to an input $x(\cdot)$ is given by

$$\begin{aligned} Ax(t) &= \iint x(t - \lambda) e^{2\pi i f(t - \lambda)} \mathcal{A}(\lambda, f) d\lambda df \\ &= \sum_n \sum_k A_{n,k} x(t - (n/2W_i)) e^{2\pi i (k/T)(t - (n/2W_i))}. \end{aligned} \tag{1}$$

Note that this is a continuous-time Gabor expansion with window function $x(\cdot)$ (see, e.g., [13]). By standard density results in Gabor theory, the collection of functions $\{x(t - (n/2W_i)) e^{2\pi i (k/T)(t - (n/2W_i))}\}$ is overcomplete as soon as $2TW_i > 1$. Consequently, without further discretization, the coefficients $A_{n,k}$ are in principle unrecoverable. Taking into consideration support constraints on \mathcal{A} , we assume that the sums are finite, viz.

$$\left(\frac{n}{2W_i}, \frac{k}{T} \right) \in \text{supp } \mathcal{A}.$$

Hence determining the channel characteristics amounts to finding $A_{n,k}$ for those pairs (n, k) . It should be noted that for a given spreading function

$\mathcal{A}(\lambda, f)$ for which $\text{supp } \mathcal{A}$ is a Lebesgue measurable set, given $\epsilon > 0$, there exist T and W_i sufficiently large that the number of such (n, k) is no more than $2S_A W_i T(1 + \epsilon)$. On the other hand, for a given T and W_i , there exist spreading functions $\mathcal{A}(\lambda, f)$ with arbitrarily small non-convex S_A for which the number of nonzero coefficients $A_{n,k}$ can be large. For example, given T and W_i , S_A could consist of rectangles centered on the points $(n/(2W_i), k/T)$ with arbitrarily small total area.

By sampling, (1) reduces to a discrete, bi-infinite linear system, viz.

$$Ax\left(\frac{p}{2W_i}\right) = \sum_n \sum_k A_{n,k} x\left(\frac{p-n}{2W_i}\right) e^{2\pi i \frac{k}{T} \left(\frac{p-n}{2W_i}\right)} \quad (2)$$

for $p \in \mathbb{Z}$. Note that (2) is the expansion of a vector in a discrete Gabor system on $\ell^2(\mathbb{Z})$, a fact not mentioned by Bello, and of which he was apparently unaware. Specifically, defining the translation operator \mathcal{T} and the modulation operator \mathcal{M} on ℓ^2 by

$$\mathcal{T}x(n) = x(n-1), \quad \text{and} \quad \mathcal{M}x(n) = e^{\pi i n / (TW_i)} x(n), \quad (3)$$

(2) can be rewritten as

$$Ax\left(\frac{p}{2W_i}\right) = \sum_n \sum_k (\mathcal{T}^n \mathcal{M}^k x)(p) A_{n,k}. \quad (4)$$

Since there are only finitely many nonzero unknowns in this system, Bello's analysis proceeds by looking at finite sections of (4) and counting degrees of freedom.

Necessity. Following the lines of the necessity argument in [21], we note that there are at least $2(T+L)(W+W_i)$ degrees of freedom in the output vector $Ax(t)$, that is, at least that many independent samples of the form $Ax(p/2W_i)$, and as observed above, no more than $2S_A W_i T(1 + \epsilon)$ nonzero unknowns $A_{n,k}$. Therefore, in order for the $A_{n,k}$ to be determined in principle, it must be true that

$$2W_i T(1 + \epsilon) S_A \leq 2(T+L)(W+W_i)$$

or

$$S_A \leq \frac{(T+L)(W+W_i)}{W_i T(1 + \epsilon)}.$$

Letting $T, W_i \rightarrow \infty$ and $\epsilon \rightarrow 0$, we arrive at $S_A \leq 1$.

Sufficiency. Considering a section of the system (4) based on the assumption that $\text{supp } \mathcal{A} \subseteq [0, L] \times [-W, W]$, the system has approximately $2W_i(T+L)$ equations in $(2W_i T)(2WL)$ unknowns. Since L and $2W$ are simply the dimensions of a rectangle that encloses the support of \mathcal{A} , $2WL$ may be quite large and independent of S_A . Hence the system will not in general be solvable.

However by assuming that $S_A < 1$, only approximately $S_A(2W_iT)$ of the $A_{n,k}$ do not vanish and the system reduces to one in which the number of equations is roughly equal to the number of unknowns. In this case it would be possible to solve (4) as long as the collection of appropriately truncated vectors $\{\mathcal{T}^n \mathcal{M}^k x : A_{n,k} \neq 0\}$ forms a linearly independent set for some vector x .

In his paper, Bello was dealing with independence properties of discrete Gabor systems apparently without realizing it, or at least without stating it explicitly. Indeed, he argues in several different ways that a vector x that produces a linearly independent set should exist, and intriguingly suggests that a vector consisting of ± 1 should exist with the property that the Grammian of the Gabor matrix corresponding to the section of (4) being considered is diagonally dominant.

The setup chosen below to prove Bello's assertion leads to the consideration of a matrix whose columns stem from a Gabor system on a finite-dimensional space, not on a sequence space.

3 Operator Sampling

The first key contribution of operator sampling is the use of frame theory and time-frequency analysis to remove assumptions of simultaneous band- and time-limiting, and also to deal with the infinite number of degrees of freedom in a functional analytic setting (Section 3.1). A second key insight is the development of a "simple measurement scheme" of the type used by the third-named author but that allows for the difficulties identified by Bello to be resolved. This insight is the use of periodically-weighted delta-trains as measurement functions for a channel. Such measurement functions have three distinct advantages.

First, they allow for the channel model to be essentially arbitrary and clarify the reduction of the operator identification problem to a finite-dimensional setting without imposing a finite dimensional model that approximates the channel. Second, it combines the naturalness of the simple measurement scheme described earlier with the flexibility of Bello's idea for measuring channels with arbitrary spreading support. Third, it establishes a connection between identification of channels and finite-dimensional Gabor systems and allows us to determine windowing vectors with appropriate independence properties.

In Section 3.1, we introduce some operator-theoretic descriptions of some of the operator classes that we are able to identify, and discuss briefly different ways of representing such operators. Such a discussion is beneficial in several ways. First, it contains a precise definition of identifiability, which comes into play when considering the generalization of the necessity condition for so-called overspread channels (Section 3.2). Second, we can extend the necessity

condition to a very large class of inputs. In other words, we can assert that in a very general sense, no input can identify an overspread channel. Third, it allows us to include both convolution operators and multiplication operators (for which the spreading functions are distributions) in the operator sampling theory. The identification of multiplication operators via operator sampling reduces to the classical sampling formula, thereby showing that classical sampling is a special case of operator sampling. In Section 3.2 we present a natural formalization of the original necessity proof of [21] (Section 2.5) to the infinite-dimensional setting, which involves an interpretation of the notion of an under-determined system to that setting. Finally, in Section 3.3 we present the scheme given first in [42, 49] for the identification of operator classes using periodically-weighted delta trains and techniques from modern time-frequency analysis.

3.1 Operator classes and operator identification

We formally consider an arbitrary operator as a *pseudodifferential operator* represented by

$$Hf(x) = \int \sigma_H(x, \xi) \widehat{f}(\xi) e^{2\pi i x \xi} d\xi, \quad (5)$$

where $\sigma_H(x, \xi) \in L^2(\mathbb{R}^2)$ is the *Kohn-Nirenberg* (KN) symbol of H . The *spreading function* $\eta_H(t, \nu)$ of the operator H is the *symplectic Fourier transform* of the KN symbol, viz.

$$\eta_H(t, \nu) = \iint \sigma_H(x, \xi) e^{-2\pi i(\nu x - \xi t)} dx d\xi \quad (6)$$

and we have the representation

$$Hf(x) = \iint \eta_H(t, \nu) \mathcal{T}_t \mathcal{M}_\nu f(x) d\nu dt \quad (7)$$

where $\mathcal{T}_t f(x) = f(x - t)$ is the *time-shift operator* and $\mathcal{M}_\nu f(x) = e^{2\pi i \nu x} f(x)$ is the *frequency-shift operator*.

This is identical to the representation given in [21] where $\eta_H(t, \nu) = \mathcal{A}(\nu, t)$, see Section 2.4.

To see more clearly where the spreading function arises in the context of communication theory, we can define the *impulse response* of the channel modeled by H , denoted $h_H(x, t)$, by

$$Hf(x) = \int h_H(x, t) f(x - t) dt.$$

Note that if h_H were independent of x , then H would be a convolution operator and hence a model for a time-invariant channel. In fact, with $\kappa_H(x, t)$ being the *kernel* of the operator H ,

$$Hf(x) = \int \kappa_H(x, t) f(t) dt \quad (8)$$

$$= \int h_H(x, t) f(x - t) dt \quad (9)$$

$$= \iint \eta_H(t, \nu) e^{2\pi i \nu(x-t)} f(x - t) d\nu dt \quad (10)$$

$$= \int \sigma_H(x, \xi) \widehat{f}(\xi) e^{2\pi i x \xi} d\xi, \quad (11)$$

where

$$\begin{aligned} h_H(x, t) &= \kappa_H(x, x - t) \\ &= \int \sigma_H(x, \xi) e^{2\pi i \xi t} d\xi, \\ &= \int \eta_H(t, \nu) e^{2\pi i \nu(x-t)} d\nu. \end{aligned} \quad (12)$$

With this interpretation, the maximum support of $\eta_H(t, \nu)$ in the first variable corresponds to the maximum spread of a delta impulse sent through the channel and the maximum support of $\eta_H(t, \nu)$ in the second variable corresponds to the maximum spread of a pure frequency sent through the channel.

Since we are interested in operators whose spreading functions have small support, it is natural to define the following operator classes, called *operator Paley-Wiener spaces* (see [41]).

Definition 1. For $S \subseteq \mathbb{R}^2$, we define the operator Paley-Wiener spaces $OPW(S)$ by

$$OPW(S) = \{H \in \mathcal{L}(L^2(\mathbb{R}), L^2(\mathbb{R})) : \text{supp } \eta_H \subseteq S, \|\sigma_H\|_{L^2} < \infty\}.$$

Remark 1. In [41, 43], the spaces $OPW^{p,q}(S)$, $1 \leq p, q < \infty$, were considered, where L^2 -membership of σ_H is replaced

$$\|\sigma_H\|_{L^{p,q}} = \left(\int \left(\int |\sigma_H(x, \xi)|^q d\xi \right)^{p/q} dx \right)^{1/p}$$

with the usual adjustments made when either $p = \infty$ or $q = \infty$. $OPW^{p,q}(S)$ is a Banach space with respect to the norm $\|H\|_{OPW^{p,q}} = \|\sigma_H\|_{L^{p,q}}$. Note that if S is bounded, then $OPW^{\infty,\infty}(S)$ consists of all bounded operators whose spreading function is supported on S . In fact, the operator norm is

then equivalent to the $OPW^{\infty,\infty}(S)$ norm, where the constants depend on S [28].

The general definition is beneficial since it also allows the inclusion of convolution operators with kernels whose Fourier transforms lie in $L^q(\mathbb{R})$ ($OPW^{\infty,q}(\mathbb{R})$) and multiplication operators whose multiplier is in $L^p(\mathbb{R})$ ($OPW^{p,\infty}(\mathbb{R})$).

The goal of operator identification is to find an input signal g such that each operator H in a given class is completely and stably determined by Hg . In other words, we ask that the operator $H \mapsto Hg$ be continuous and bounded below on its domain. In our setting, this translates to the existence of $c_1, c_2 > 0$ such that

$$c_1 \|\sigma_H\|_{L^2} \leq \|Hg\|_{L^2} \leq c_2 \|\sigma_H\|_{L^2}, \quad H \in OPW(S). \quad (13)$$

This definition of identifiability of operators originated in [26]. Note that (13) implies that the mapping $H \mapsto Hg$ is *injective*, that is, that $Hg = 0$ implies that $H \equiv 0$, but is not equivalent to it. The inequality (13) adds to injectivity the assertion that H is also stably determined by Hg in the sense that a small change in the output Hg would correspond to a small change in the operator H . Such stability is also necessary for the existence of an algorithm that will reliably recover H from Hg . In this scheme, g is referred to as an *identifier* for the operator class $OPW(S)$ and if (13) holds, we say that *operator identification* is possible.

In trying to find an explicit expression for an identifier, we use as a starting point the “simple measurement scheme” of [21], in which g is a delta train, viz. $g = \sum_n \delta_{nT}$ for some $T > 0$. In the framework of operator identification the channel measurement criterion in [21] takes the following form [26, 42, 41].

Theorem 1. *For $H \in OPW([0, T] \times [-\Omega/2, \Omega/2])$ with $T\Omega \leq 1$, we have*

$$\|H \sum_{k \in \mathbb{Z}} \delta_{kT}\|_{L^2(\mathbb{R})} = T \|\sigma_H\|_{L^2},$$

and H can be reconstructed by means of

$$\kappa_H(x+t, x) = \chi_{[0,T]}(t) \sum_{n \in \mathbb{Z}} \left(H \sum_{k \in \mathbb{Z}} \delta_{kT} \right)(t+nT) \frac{\sin(\pi T(x-n))}{\pi T(x-n)} \quad (14)$$

where $\chi_{[0,T]}(t) = 1$ for $t \in [0, T]$ and 0 elsewhere and with convergence in the L^2 norm and uniformly in x for every t .

As was observed earlier, the key feature of this scheme is that the spacing of the deltas in the identifier is sufficiently large so as to allow the response of the channel to a given delta to “die out” before the next delta is sent. In other words, the parameter T must exceed the time-spread of the channel. On the other hand, the rate of change of the channel, as measured by its

bandwidth Ω , must be small enough that its impulse response can be recovered from “samples” of the channel taken T time units apart. In particular, the samples of the impulse response T units apart can be easily determined from the output. In the general case considered by Bello, in which the spreading support of the operator is not contained in a rectangle of unit area, this intuition breaks down.

Specifically, suppose that we consider the operator class $OPW(S)$ where $S \subseteq [0, T_0] \times [-\Omega_0/2, \Omega_0/2]$ and $T_0\Omega_0 \gg 1$ but where $|S| < 1$. Then sounding the channel with a delta train of the form $g = \sum_n \delta_{nT_0}$ would severely *undersample* the impulse response function. Simply increasing the sampling rate, however, would produce overlap in the responses of the channel to deltas close to each other. An approach to the undersampling problem in the literature of classical sampling theory is to sample at the low rate transformed versions of the function, chosen so that the interference of the several undersampled functions can be dealt with. This idea has its most classical expression in the Generalized Sampling scheme of Papoulis [39]. Choosing shifts and constant multiples of our delta train results in an identifier of the form $g = \sum_n c_n \delta_{nT}$ where the weights (c_n) have period P (for some $P \in \mathbb{N}$) and $T > 0$ satisfies $PT > T_0$.

If g is discretely supported (for example, a periodically-weighted delta-train), then we refer to operator identification as *operator sampling*. The utility of periodically-weighted delta trains for operator identification is a cornerstone of operator sampling and has far-reaching implications culminating in the developments outlined in Sections 5 and 6.

3.2 Kailath’s necessity proof and operator identification

In Section 2.5 we presented the proof of the necessity of the condition $BL \leq 1$ for channel identification as given in [21]. The argument consisted of finding a finite-dimensional approximation of the channel H , and then showing that, given any putative identifier g , the number of degrees of freedom present in the output Hg must be at least as large as the number of degrees of freedom in the channel itself. For this to be true in any finite-dimensional setting, we must have $BL < 1$ and so in the limit we require $BL \leq 1$. In essence, if $BL > 1$, we have a linear system with fewer equations than unknowns which necessarily has a nontrivial nullspace. The generalization of this notion to the infinite-dimensional setting is the basis of the necessity proof that appears in [26]. In this section, we present an outline of that proof, and show how the natural tool for this purpose once again comes from time-frequency analysis.

To see the idea of the proof, assume that $BL > 1$ and for simplicity let $S = [-\frac{L}{2}, \frac{L}{2}] \times [-\frac{B}{2}, \frac{B}{2}]$. The goal is to show that for any sounding signal s in

an appropriately large space of distributions², the operator $\Phi_s: OPW(S) \rightarrow L^2(\mathbb{R})$, $H \mapsto Hs$, is not stable, that is, it does not possess a lower bound in the inequality (13).

First, define the operator $E: l_0(\mathbb{Z}^2) \rightarrow OPW(S)$, where $l_0(\mathbb{Z}^2)$ is the space of finite sequences equipped with the l^2 norm, by

$$E(\sigma) = E(\{\sigma_{k,l}\}) = \sum_{k,l} \sigma_{k,l} \mathcal{M}_{k\lambda/L} \mathcal{T}_{l\lambda/B} P \mathcal{T}_{-l\lambda/B} \mathcal{M}_{-k\lambda/L}$$

where $1 < \lambda$ is chosen so that $1 < \lambda^4 < BL$ and where P is a time-frequency localization operator whose spreading function $\eta_P(t, \nu)$ is infinitely differentiable, supported in S , and identically one on $[-\frac{L}{2\lambda}, \frac{L}{2\lambda}] \times [-\frac{B}{2\lambda}, \frac{B}{2\lambda}]$. It is easily seen that the operator E is well-defined and has spreading function

$$\eta_{E(\sigma)}(t, \nu) = \eta_P(t, \nu) \sum_{k,l} \sigma_{k,l} e^{2\pi i(k\lambda t/L - l\lambda\nu/B)}.$$

By construction, it follows that for some constant c_1 , $\|E(\sigma)\|_{OPW(S)} \geq c_1 \|\sigma\|_{l^2(\mathbb{Z}^2)}$, for all σ , and that for any distribution s , Ps decays rapidly in time and in frequency.

Next define the Gabor analysis operator $C_g: L^2(\mathbb{R}) \rightarrow l^2(\mathbb{Z}^2)$ by

$$C_g(s) = \{\langle s, \mathcal{M}_{k\lambda^2/L} \mathcal{T}_{l\lambda^2/B} g \rangle\}_{k,l \in \mathbb{Z}}$$

where $g(x) = e^{-\pi x^2}$. A well-known theorem in Gabor theory asserts that $\{\mathcal{M}_{k\alpha} \mathcal{T}_{l\beta} g\}_{k,l \in \mathbb{Z}}$ is a Gabor frame for $L^2(\mathbb{R})$ for every $\alpha\beta < 1$ ([33, 66, 65]). Consequently C_g satisfies, for some $c_2 > 0$, $\|C_g(s)\|_{l^2(\mathbb{Z}^2)} \geq c_2 \|s\|_{L^2(\mathbb{R})}$ for all s , since $\lambda^2/L \cdot \lambda^2/B = \lambda^4/B\lambda < 1$.

For any s , consider the composition operator

$$C_g \circ \Phi_s \circ E: l_0(\mathbb{Z}^2) \rightarrow l^2(\mathbb{Z}^2).$$

The crux of the proof lies in showing that this composition operator is not stable, that is, it does not have a lower bound. Since C_g and E are both bounded below, it follows that Φ_s cannot be stable. Since $s \in S'_0(\mathbb{R})$ was arbitrary, this completes the proof.

To complete this final step we examine the canonical bi-infinite matrix representation of the above defined composition of operators, that is, the matrix $M = (m_{k',l',k,l})$ that satisfies

$$(C_g \circ \Phi_s \circ E(\sigma))_{k',l'} = \sum_{k,l} m_{k',l',k,l} \sigma_{k,l}.$$

² $S'_0(\mathbb{R})$, the dual space of the Feichtinger algebra $S_0(\mathbb{R})$ [13], or $\mathcal{S}'(\mathbb{R})$, the space of tempered distributions [43]. These spaces are large enough to contain weighted infinite sums of delta distributions.

It can be shown that M has the property that for some rapidly decreasing function $w(x)$,

$$|m_{k',l',k,l}| \leq w(\max\{|\lambda k' - k|, |\lambda l' - l|\}). \quad (15)$$

The proof is completed by the following Lemma. Its proof can be found in [26] and generalizations can be found in [55].

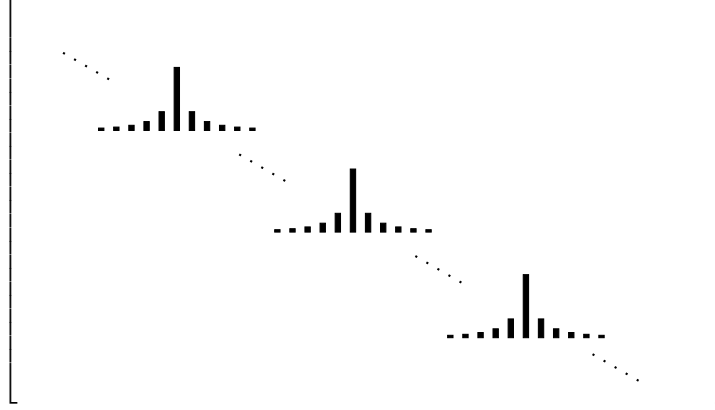


Fig. 2: A $1/\lambda$ -slanted matrix M . The matrix is dominated by entries on a slanted diagonal of slope $1/\lambda$.

Lemma 1. *Given $M = (m_{j',j})_{j',j \in \mathbb{Z}^2}$. If there exists a monotonically decreasing function $w: R_0^+ \rightarrow R_0^+$ with $w = O(x^{-2-\delta})$, $\delta > 0$, and constants $\lambda > 1$ and $K_0 > 0$ with $|m_{j',j}| < w(\|\lambda j' - j\|_\infty)$ for $\|\lambda j' - j\|_\infty > K_0$, then M is not stable.*

Intuitively, this result asserts that a bi-infinite matrix whose entries decay rapidly away from a skew diagonal behaves like a finite matrix with more rows than columns (see Figure 2). Such a matrix will always have a nontrivial nullspace. In the case of an infinite matrix what can be shown is that at best its inverse will be unbounded.

We can make a more direct connection from this proof to the original necessity argument in [21] in the following way. If we restrict our attention to sequences $\{\sigma_{k,l}\}$ with a fixed finite support of size say N , then the image of this subspace of sequence space under the mapping E is an N -dimensional subspace of $OPW(S)$. The operator P is essentially a time-frequency localization operator. This fact is established in [26] and follows from the rapid decay of the Fourier transform of η_P . Since η_P itself is concentrated on a rectangle of area BL/λ^2 , its Fourier transform will be concentrated on a rectangle of area λ^2/BL . From this it follows that for σ as described above, the operator

$E(\sigma)$ essentially localizes a function to a region in the time-frequency plane of area $N(\lambda^2/BL)$.

Considering now the Gabor analysis operator C_g , we observe that the Gaussian $g(x)$ essentially occupies a time-frequency cell of area 1, and that this function is shifted in the time-frequency plane by integer multiples of $(\lambda^2/B, \lambda^2/L)$. Hence to “cover” a region in the time-frequency plane of area $N(\lambda^2/BL)$ would require only about

$$\frac{N(\lambda^2/BL)}{\lambda^4/BL} = \frac{N}{\lambda^2}$$

time-frequency shifts. So roughly speaking, in order to resolve N degrees of freedom in the operator $E(\sigma_{k,l})$, we have only $N/\lambda^2 < N$ degrees of freedom in the output of the operator $E(\sigma_{k,l})s$.

3.3 Identification of operator Paley-Wiener spaces by periodically weighted delta-trains

Theorem 1 is based on arguments outlined in Section 2.4 and applies only to $OPW(S)$ if S is contained in a rectangle of area less than or equal to one. In the following, we will develop the tools that allow us to identify $OPW(S)$ for any compact set S of Lebesgue measure less than one.

In our approach we discretize the channel by covering the spreading support S with small rectangles of fixed sidelength, which we refer to as a *rectification* of S . As long as the measure of S is less than one, it is possible to do this in such a way that the total area of the rectangles is also less than one. This idea seems to bear some similarity to Bello’s philosophy of sampling the spreading function on a fixed grid but with one fundamental difference. Bello’s approach is based on replacing t and x by samples, thereby approximating the channel. For a better approximation, sampling on a finer grid is necessary, which results in a larger system of equations that must be solved. In our approach, as soon as the total area of the rectification is less than one, the operator modeling the channel is completely determined by the discrete model. Once this is achieved, identification of the channel reduces to solving a single linear system of equations at each point.

Given parameters $T > 0$ and $P \in \mathbb{N}$, we assume that S is rectified by rectangles of size $T \times \Omega$, where $\Omega = 1/(TP)$, such that the total area of the rectangles is less than one. Given a period- P sequence $c = (c_n)_{n \in \mathbb{Z}}$, we then define the *periodically weighted delta-train* g by $g = \sum_{n \in \mathbb{Z}} c_n \delta_{nT}$. The goal of this subsection is to describe the scheme by which a linear system of P equations in a priori P^2 unknowns can be derived by which an operator $H \in OPW(S)$ can be completely determined by $Hg(x)$. In this sense, the

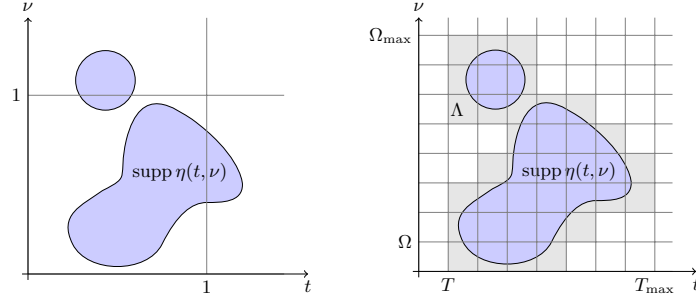


Fig. 3: A set not satisfying Kailath's condition is rectified with $1/(T\Omega) = P \in \mathbb{N}$, the rectification has area ≤ 1 , $\Omega_{\max} \leq 1/T$, and $T_{\max} \leq 1/\Omega$.

“degrees of freedom” in the operator class $OPW(S)$, and that of the output function $Hg(x)$ are precisely defined and can be effectively compared.

The basic tool of time-frequency analysis that makes this possible is the *Zak transform* (see [13]).

Definition 2. The non-normalized Zak Transform is defined for $f \in \mathcal{S}(\mathbb{R})^3$, and $a > 0$ by

$$Z_a f(t, \nu) = \sum_{n \in \mathbb{Z}} f(t - an) e^{2\pi i a n \nu}.$$

$Z_a f(t, \nu)$ satisfies the quasi-periodicity relations

$$Z_a f(t + a, \nu) = e^{2\pi i a \nu} Z_a f(t, \nu)$$

and

$$Z_a f(t, \nu + 1/a) = Z_a f(t, \nu).$$

$\sqrt{a} Z_a$ can be extended to a unitary operator from $L^2(\mathbb{R})$ onto $L^2([0, a] \times [0, 1/a])$.

A somewhat involved but elementary calculation yields the following (see [50] and Section 7.1).

Lemma 2. Let $T > 0$, $P \in \mathbb{N}$, $c = (c_n)$, and g be given as above. Then for all $(t, \nu) \in \mathbb{R}^2$, and $p = 0, 1, \dots, P-1$,

$$\begin{aligned} & e^{-2\pi i \nu T p} (Z_{TP} \circ H)g(t + Tp, \nu) \\ &= \Omega \sum_{q, m=0}^{P-1} (T^q M^m c)_p e^{-2\pi i \nu T q} \eta_H^Q(t + Tq, \nu + m/TP). \end{aligned} \quad (16)$$

Here \mathcal{T} and \mathcal{M} are the translation and modulation operators given in Definition 3, and $\eta_H^Q(t, \nu)$ is the *quasiperiodization* of η_H ,

³ $\mathcal{S}(\mathbb{R})$ denotes the Schwartz class of infinitely-differentiable, rapidly-decreasing functions.

$$\eta_H^Q(t, \nu) = \sum_k \sum_\ell \eta_H(t + kTP, \nu + \ell/T) e^{-2\pi i \nu kTP} \quad (17)$$

whenever the sum is defined.

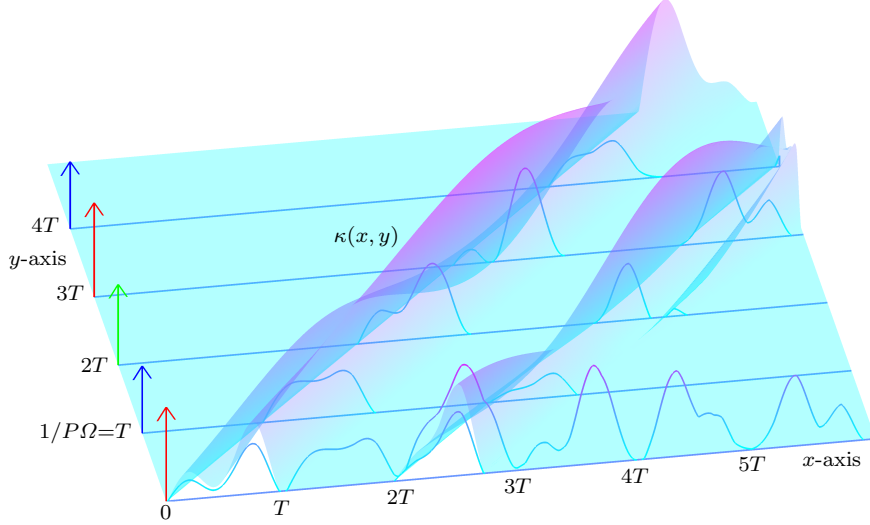


Fig. 4: Channel sounding of $OPW([0, 2/3] \times [-1/4, 1/4] \cup [4/3, 2] \times [-1/2, 1/2])$ using a P -periodically weighted delta train g . The kernel $\kappa(x, y)$ takes values on the (x, y) -plane, the sounding signal g , a weighted impulse train, is defined on the y -axis, and the output signal $Hg(x) = \int \kappa(x, y)g(y)dy$ is displayed on the x -axis. Here, the sample values of the tab functions $h(x, t) = \kappa(x, t - x)$ are not easily read of the response $Hg(x)$ as, for example for $x \in [2T, 3T] = [4/3, 2]$ we have $Hg(x) = 0.7\kappa(x, 0) + 0.6\kappa(x, 2T) = 0.7h(x, x) + .6h(x, 2T - x)$. In detail, we have $g = \dots + 0.7\delta_{-2} + 0.5\delta_{-4/3} + 0.6\delta_{-2/3} + 0.7\delta_0 + 0.5\delta_{2/3} + 0.6\delta_{4/3} + 0.7\delta_2 + 0.5\delta_{8/3} + \dots$, so $P = 3$, $T = 2/3$, $\Omega = 1/PT = 1/2$, $c_n = 0.7$ if $n \bmod 3 = 0$, $c_n = 0.5$ if $n \bmod 3 = 1$, $c_n = 0.6$ if $n \bmod 3 = 2$.

Under the additional simplifying assumption that the spreading function $\eta_H(t, \nu)$ is supported in the large rectangle $[0, TP] \times [0, 1/T]$, and by restricting (16) to the rectangle $[0, T] \times [0, 1/(TP)]$, we arrive at the $P \times P^2$ linear system

$$\mathbf{Z}_{Hg}(t, \nu)_p = \sum_{q, m=0}^{P-1} G(c)_{p, (q, m)} \boldsymbol{\eta}_H(t, \nu)_{(q, m)} \quad (18)$$

where

$$\mathbf{Z}_{Hg}(t, \nu)_p = (Z_{TP} \circ H)g(t + pT, \nu) e^{-2\pi i \nu pT}, \quad (19)$$

$$\boldsymbol{\eta}_H(t, \nu)_{(q, m)} = \Omega \eta_H(t + qT, \nu + m/TP) e^{-2\pi i \nu qT} e^{-2\pi i qm/P}, \quad (20)$$

and where $G(c)$ is a finite Gabor system matrix (23). If (18) can be solved for each $(t, \nu) \in [0, T] \times [0, 1/(TP)]$, then the spreading function for an operator H can be completely determined by its response to the periodically-weighted delta-train g .

As anticipated by Bello, two issues now become relevant. (1) We require that $\text{supp } \eta_H$ occupy no more than P of the shifted rectangles $[0, T] \times [0, 1/(TP)] + (qT, k/(TP))$, so that (18) has at least as many equations as unknowns. This forces $|\text{supp } \eta_H| \leq 1$. (2) We require that c be chosen in such a way that the $P \times P$ system formed by removing the columns of $G(c)$ corresponding to vanishing components of $\boldsymbol{\eta}_H$ is invertible. That such c exist is a fundamental cornerstone of operator sampling and is the subject of the next section.

Based on the existence of c such that any set of P columns of $G(c)$ form a linearly independent set, we can prove the following [49].

Theorem 2. *For $S \subseteq (0, \infty) \times \mathbb{R}$ compact with $|S| < 1$, there exists $T > 0$ and $P \in \mathbb{N}$, and a period- P sequence $c = (c_n)$ such that $g = \sum_n c_n \delta_{nT}$ identifies $OPW(S)$. In particular, there exist period- P sequences $b_j = (b_{j,k})$, and integers $0 \leq q_j, m_j \leq P-1$, for $0 \leq j \leq P-1$ such that*

$$\begin{aligned} h(x, t) &= e^{-\pi i t/T} \sum_k \sum_{j=0}^{P-1} [b_{j,k} Hg(t - (q_j - k)T) \\ &\quad e^{2\pi i m_j(x-t)/PT} \phi((x-t) + (q_j - k)T) r(t - q_j T)] \end{aligned} \quad (21)$$

where $r, \phi \in \mathcal{S}(\mathbb{R})$ satisfy

$$\sum_{k \in \mathbb{Z}} r(t + kT) = 1 = \sum_{n \in \mathbb{Z}} \widehat{\phi}(\gamma + n/PT), \quad (22)$$

where $r(t)\widehat{\phi}(\gamma)$ is supported in a neighborhood of $[0, T] \times [0, 1/PT]$, and where the sum in (21) converges unconditionally in L^2 and for each t uniformly in x .

Equation (21) is a generalization of (14) which is easily seen by choosing $\phi(x) = \sin(\pi PTx)/(\pi PTx)$ and $r(t)$ to be the characteristic function of $[0, T]$.

4 Linear Independence Properties of Gabor Frames

4.1 Finite Gabor Frames

Definition 3. Given $P \in \mathbb{N}$, let $\omega = e^{2\pi i/P}$ and define the *translation operator* \mathcal{T} on $(x_0, \dots, x_{P-1}) \in \mathbb{C}^P$ by

$$\mathcal{T}x = (x_{P-1}, x_0, x_1, \dots, x_{P-2}),$$

and the *modulation operator* \mathcal{M} on \mathbb{C}^P by

$$\mathcal{M}x = (\omega^0 x_0, \omega^1 x_1, \dots, \omega^{P-1} x_{P-1}).$$

Given a vector $c \in \mathbb{C}^P$ the *finite Gabor system with window c* is the collection $\{\mathcal{T}^q \mathcal{M}^p c\}_{q,p=0}^{P-1}$. Define the *full Gabor system matrix* $G(c)$ to be the $P \times P^2$ matrix

$$G(c) = [D_0 W_P \mid D_1 W_P \mid \dots \mid D_{P-1} W_P] \quad (23)$$

where D_k is the diagonal matrix with diagonal

$$\mathcal{T}^k c = (c_{P-k}, \dots, c_{P-1}, c_0, \dots, c_{P-k-1}),$$

and W_P is the $P \times P$ Fourier matrix $W_P = (e^{2\pi i n m / P})_{n,m=0}^{P-1}$.

Remark 2. (1) For $0 \leq q, p \leq P-1$, the $(q+1)$ st column of the submatrix $D_p W_P$ is the vector $\mathcal{M}^p \mathcal{T}^q c$ where the operators \mathcal{M} and \mathcal{T} are as in Definition 3. This means that each column of the matrix $G(c)$ is a unimodular constant multiple of an element of the finite Gabor system with window c , namely $\{e^{-2\pi i p q / P} \mathcal{T}^q \mathcal{M}^p c\}_{q,p=0}^{P-1}$.

(2) Note that the finite Gabor system defined above consists of P^2 vectors in \mathbb{C}^P which form an overcomplete tight frame for \mathbb{C}^P [32]. For details on Gabor frames in finite dimensions, see [32, 29, 9] and the overview article [54].

(3) Note that we are abusing notation slightly by identifying a vector $c \in \mathbb{C}^P$ with an P -periodic sequence $c = (c_n)$ in the obvious way.

Definition 4. [8] The *Spark* of an $M \times N$ matrix F is the size of the smallest linearly dependent subset of columns, i.e.,

$$\text{Spark}(F) = \min\{\|x\|_0 : Fx = 0, x \neq 0\}$$

where $\|x\|_0$ is the number of nonzero components of the vector x . If $\text{Spark}(F) = M+1$, then F is said to have *full Spark*. $\text{Spark}(F) = k$ implies that any collection of fewer than k columns of F is linearly independent.

4.2 Finite Gabor frames are generically full Spark

The existence of Gabor matrices with full Spark has been addressed in [32, 34]. The results in these two papers are as follows.

Theorem 3. [32] *If $P \in \mathbb{N}$ is prime then there exists a dense, open subset of $c \in \mathbb{C}^P$ such that every minor of the Gabor system matrix $G(c)$ is nonzero. In particular, for such c , $G(c)$ has full Spark.*

Theorem 4. [34] *For every $P \in \mathbb{N}$, there exists a dense, open subset of $c \in \mathbb{C}^P$ such that the Gabor system matrix $G(c)$ has full Spark.*

The goal of this subsection is to outline the proof of Theorems 3 and 4. We will adopt some of the following notation and terminology of [34].

Let $P \in \mathbb{N}$ and let M be an $P \times P$ submatrix of $G(c)$. For $0 \leq \kappa < P$ let ℓ_κ be the number of columns of M chosen from the submatrix $D_\kappa W_P$ of (23). While the vector $\ell = (\ell_\kappa)_{\kappa=0}^{P-1}$ does not determine M uniquely, it describes the matrix M sufficiently well for our purposes. Define M_κ to be the $P \times \ell_\kappa$ matrix consisting of those columns of M chosen from $D_\kappa W_P$. Given the *ordered partition* $B = (B_0, B_1, \dots, B_{P-1})$ where $\{B_0, B_1, \dots, B_{P-1}\}$ forms a partition of $\{0, \dots, P-1\}$, and where for each $0 \leq \kappa < P$, $|B_\kappa| = \ell_\kappa$, let $M_\kappa(B_\kappa)$ be the $\ell_\kappa \times \ell_\kappa$ submatrix of M_κ whose rows belong to B_κ . Then $\det(M) = \prod \det(M_\kappa(B_\kappa))$ where the product is taken over all such ordered partitions B . This formula is called the *Lagrange expansion* of the determinant.

Each ordered partition B corresponds to a permutation on \mathbb{Z}_P as follows. Define the *trivial partition* $A = (A_0, A_1, \dots, A_{P-1})$ by

$$A_j = \left\{ \sum_{i=0}^{j-1} \ell_i, \left(\sum_{i=0}^{j-1} \ell_i \right) + 1, \dots, \left(\sum_{i=0}^j \ell_i \right) - 1 \right\}$$

so that $A_0 = [0, \ell_0 - 1]$, $A_1 = [\ell_0, \ell_0 + \ell_1 + 1]$, \dots , $A_{P-1} = [\ell_0 + \dots + \ell_{P-2}, P - 1]$. Then given $B = (B_0, B_1, \dots, B_{P-1})$ there is a permutation $\sigma \in S_P$ such that $\sigma(A_\kappa) = B_\kappa$ for all κ . This σ is unique up to permutations that preserve A , that is, up to $\tau \in S_P$ such that $\tau(A_\kappa) = A_\kappa$ for all κ . Call such a permutation *trivial* and denote by Γ the subgroup of S_P consisting of all trivial permutations. Then the ordered partitions B of \mathbb{Z}_P can be indexed by equivalence classes of permutations $\sigma \in S_P/\Gamma$.

The key observation is that $\det(M)$ is a homogeneous polynomial in the P variables c_0, c_1, \dots, c_{P-1} and we can write

$$\det(M) = \sum_{\sigma \in S_P/\Gamma} a_\sigma C^\sigma \quad (24)$$

where the monomial C^σ is given by

$$C^\sigma = \prod_{\kappa=0}^{P-1} \prod_{j \in \sigma(A_\kappa)} c_{(j-\kappa) \pmod P}.$$

If it can be shown that this polynomial does not vanish identically then we can choose a dense, open subset of $c \in \mathbb{C}^P$ for which $\det(M) \neq 0$. Since there are only finitely many $P \times P$ submatrices of $G(c)$ it follows that there is a dense, open subset of c for which $\det(M) \neq 0$ for all M , and we conclude that, for these c , $G(c)$ has full Spark.

Following [34], we say that a monomial C^{σ_0} *appears uniquely* in (24) if for every $\sigma \in S_P/\Gamma$ such that $\sigma \neq \sigma_0$, $C^\sigma \neq C^{\sigma_0}$. Therefore, in order to show that the polynomial (24) does not vanish identically, it is sufficient to show that (1) there is a monomial C^σ that appears uniquely in (24) and (2) the coefficient a_σ of this monomial does not vanish.

Obviously, whether or not (24) vanishes identically does not depend on how the variables c_i are labelled. More specifically, if the variables are renamed by a cyclical shift of the indices, viz., $c_i \mapsto c_{(i+\gamma) \bmod P}$ for some $0 \leq \gamma < P$, then

$$\det(M)(c_{\gamma+1}, \dots, c_{P-1}, c_0, \dots, c_\gamma) = \pm \det(M')(c_0, \dots, c_{P-1})$$

where M' is an $P \times P$ submatrix described by the vector

$$\ell' = (\ell_{\gamma+1}, \dots, \ell_{P-1}, \ell_0, \dots, \ell_\gamma).$$

4.2.1 The lowest index monomial

In [32], a monomial referred to in [34] as the *lowest index (LI) monomial* is defined that has the required properties when P is prime. In order to see this, note first that each coefficient a_σ in the sum (24) is the product of minors of the Fourier matrix W_P and since P is prime, Chebotarev's Theorem says that such minors do not vanish [68]. More specifically,

$$a_\sigma C^\sigma = \pm \prod_{\kappa=0}^{P-1} \det(M_\kappa(\sigma(A_\kappa)))$$

and for each κ , the columns of M_κ are columns of W_P where each row has been multiplied by the same variable c_j and $M_\kappa(\sigma(A_\kappa))$ is a square matrix formed by choosing ℓ_κ rows of M_κ . Hence for each κ , $\det(M_\kappa(\sigma(A_\kappa)))$ is a monomial in c with coefficients a constant multiple of a minor of W_P . Since a_σ is the product of those minors, it does not vanish.

Note moreover that each submatrix $M_\kappa(\sigma(A_\kappa))$ is an $\ell_\kappa \times \ell_\kappa$ matrix, so that $\det(M_\kappa(\sigma(A_\kappa)))$ is the sum of a multiple of the product of $\ell_\kappa!$ diagonals of $M_\kappa(\sigma(A_\kappa))$. Hence $a_\sigma C^\sigma$ is the sum of multiples of the product of $\prod_{\kappa=0}^{P-1} \ell_\kappa!$ generalized diagonals of M .

We define the LI monomial as in [32] as follows. If M is 1×1 , then $\det(M)$ is a multiple of a single variable c_j and we define the LI monomial, p_M by $p_M = c_j$. If M is $d \times d$, let c_j be the variable of lowest index appearing in M . Choose any entry of M in which c_j appears, eliminate the row and column containing that entry, and call the remaining $(d-1) \times (d-1)$ matrix M' . Define $p_M = c_j p_{M'}$. It is easy to see that the monomial p_M is independent of the entry of M chosen at each step. In order to show that the LI monomial appears uniquely in (24), we observe as in [32] that the number of diagonals

in $\det(M)$ that correspond to the LI monomial is $\prod_{\kappa=0}^{P-1} \ell_{\kappa}!$. Since this is also the number of generalized diagonals appearing in the calculation of each $\det(M_{\kappa}(\sigma(A_{\kappa})))$, it follows that this monomial appears only once. The details of the argument can be found in Section 7.2. Note that because P is prime, this argument is valid no matter how large the matrix M is. In other words, M does not have to be an $P \times P$ submatrix in order for the result to hold. Consequently, given $k < P$ and M an arbitrary $P \times k$ submatrix of $G(c)$, we can form the $k \times k$ matrix M_0 by choosing k rows of M in such a way that the LI monomial of M_0 contains at most only the variables c_0, \dots, c_{k-1} . This observation leads to the following theorem for matrices with arbitrary Spark.

Theorem 5. [50] *If $P \in \mathbb{N}$ is prime, and $0 < k < P$, there exists an open, dense subset of $c \in \mathcal{C}^k \times \{0\} \subseteq \mathbb{C}^P$ with the property that $\text{Spark}(G(c)) = k+1$.*

This result has implications for relating the capacity of a time-variant communication channel to the area of the spreading support, see [50].

4.2.2 The consecutive index monomial

In [34], a monomial referred to as the *consecutive index (CI) monomial* is defined that has the required properties for any $P \in \mathbb{N}$. The CI monomial, C^I , is defined as the monomial corresponding to the identity permutation in S_P/Γ , that is, to the equivalence class of the trivial partition $A = (A_0, A_1, \dots, A_{P-1})$. Hence

$$C^I = \prod_{\kappa=0}^{P-1} \prod_{j \in A_{\kappa}} c_{(j-\kappa) \bmod P}.$$

For each κ , the monomial appearing in $\det(M_{\kappa}(A_{\kappa}))$, $\prod_{j \in A_{\kappa}} c_{(j-\kappa) \bmod P}$, consists of a product of ℓ_{κ} variables c_j with consecutive indices modulo P .

That $a_I \neq 0$ follows from the observation that for each κ , $\det(M_{\kappa}(A_{\kappa}))$ is a monomial whose coefficient is a nonzero multiple of a Vandermonde determinant and hence does not vanish (for details, see [34]). The proof that C^I appears uniquely in (24) amounts to showing that, with respect to an appropriate cyclical renaming of the variables c_i , the CI monomial uniquely minimizes the quantity $\Lambda(C^{\sigma}) = \sum_{i=0}^{P-1} i^2 \alpha_i$, where α_i is the exponent of the variable c_i in C^{σ} . An abbreviated version of the proof of this result as it appears in [34] is given in Section 7.3.

As a final observation, we quote the following corollary that provides an explicit construction of a unimodular vector c such that $G(c)$ has full Spark.

Corollary 1. [34] *Let $\zeta = e^{2\pi i/(P-1)^4}$ or any other primitive root of unity of order $(P-1)^4$ where $P \geq 4$. Then the vector*

$$c = (1, \zeta, \zeta^4, \zeta^9, \dots, \zeta^{(P-1)^2})$$

generates a Gabor frame for which $G(c)$ has full Spark.

5 Generalizations of operator sampling to higher dimensions

The operator representations (5), (6), and (7) hold verbatim for higher dimensional variables $x, \xi, t, \nu \in \mathbb{R}^d$. In this section, we address the identifiability of

$$OPW(S) = \{H \in \mathcal{L}(L^2(\mathbb{R}^d), L^2(\mathbb{R}^d)) : \text{supp } \mathcal{F}_s \sigma_H \subseteq S, \|\sigma_H\|_{L^2} < \infty\}$$

where $S \subseteq \mathbb{R}^{2d}$.

Looking at the components of the multidimensional variables separately, Theorem 1 easily generalizes as follows.

Theorem 6. *For $H \in OPW(\prod_{\ell=1}^d [0, T_\ell] \times \prod_{\ell=1}^d [-\Omega_\ell/2, \Omega_\ell/2])$ with $T_\ell \Omega_\ell \leq 1$, $\ell = 1, \dots, d$, we have*

$$\|H \sum_{k_1 \in \mathbb{Z}} \dots \sum_{k_d \in \mathbb{Z}} \delta_{(k_1 T_1, \dots, k_d T_d)}\|_{L^2(\mathbb{R})} = T_1 \dots T_d \|\sigma_H\|_{L^2},$$

and H can be reconstructed by means of

$$\begin{aligned} \kappa_H(x + t, x) &= \chi_{\prod_{\ell=1}^d [0, T_\ell]}(t) \sum_{n_1 \in \mathbb{Z}} \dots \sum_{n_d \in \mathbb{Z}} \\ &\quad \left(H \sum_{k_1 \in \mathbb{Z}} \dots \sum_{k_d \in \mathbb{Z}} \delta_{(k_1 T_1, \dots, k_d T_d)} \right) (t + (n_1 T_1, \dots, n_d T_d)) \\ &\quad \frac{\sin(\pi T_1 (x_1 - n_1))}{\pi T_1 (x_1 - n_1)} \dots \frac{\sin(\pi T_d (x_d - n_d))}{\pi T_d (x_d - n_d)} \end{aligned}$$

with convergence in the L^2 norm.

In the following, we address the situation where S is not contained in a set $\prod_{\ell=1}^d [0, T_\ell] \times \prod_{\ell=1}^d [-\Omega_\ell/2, \Omega_\ell/2]$ with $T_\ell \Omega_\ell \leq 1$, $\ell = 1, \dots, d$. For example, $S = [0, 1] \times [0, 2] \times [0, \frac{1}{4}] \times [0, 1] \subseteq \mathbb{R}^4$ of volume $\frac{1}{2}$ is not covered by Theorem 6.

To give a higher dimensional variant of Theorem 2, we shall denote pointwise products of finite and infinite length vectors k and T by $k \star T$, that is, $k \star T = (k_1 T_1, \dots, k_d T_d)$ for $k, T \in \mathbb{C}^d$. Similarly, $k/T = (k_1/T_1, \dots, k_d/T_d)$.

Theorem 7. *If $S \subseteq (0, \infty)^d \times \mathbb{R}^d$ is compact with $|S| < 1$ then $OPW(S)$ is identifiable. Specifically, there exist $T_1, \dots, T_d > 0$ and pairwise relatively prime natural numbers P_1, \dots, P_d such that*

$$S \subseteq \prod_{\ell=1}^d [0, P_\ell T_\ell] \times \prod_{\ell=1}^d [-1/(2T_\ell), 1/(2T_\ell)],$$

and a sequence $c = (c_n) \in \ell^\infty(\mathbb{Z}^d)$ which is P_ℓ periodic in the ℓ -th component n_ℓ such that $g = \sum_{n \in \mathbb{Z}^d} c_n \delta_{n \star T}$ identifies $OPW^2(S)$. In fact, for such g there exists for each $j \in J = \prod_{\ell=1}^d \{0, 1, \dots, P_\ell - 1\}$ a sequences $b_j = (b_{j,k})$ which is P_ℓ periodic in k_ℓ and $2d$ -tuples $(q_j, m_j) \in J \times J$ with

$$h(x, t) = e^{-\pi i \sum_{\ell=1}^d t_\ell / T_\ell} \sum_{k \in \mathbb{Z}^d} \sum_{j \in J} [b_{j,k} Hg(t - (q_j - k) \star T) e^{2\pi i m_j \cdot ((x-t)/P \star T)} \phi((x-t) + (q_j - k) \star T) r(t - q_j \star T)]. \quad (25)$$

The functions $r, \phi \in \mathcal{S}(\mathbb{R}^d)$ are assumed to satisfy

$$\sum_{k \in \mathbb{Z}^d} r(t + k \star T) = 1 = \sum_{n \in \mathbb{Z}^d} \hat{\phi}(\gamma + (n/P \star T), \quad (26)$$

and $r(t) \hat{\phi}(\gamma)$ is supported in a neighborhood of $\prod_{\ell=1}^d [0, T_\ell] \times \prod_{\ell=1}^d [0, 1/P_\ell T_\ell]$. The sum in (25) converges unconditionally in L^2 and for each t uniformly in x .

This result follows from adjusting the proof of Theorem 7 to the higher dimensional setting. For example, it will employ the Zak transform

$$Z_{T \star P} f(t, \nu) = \sum_{n \in \mathbb{Z}^d} f(t - n \star P \star T) e^{2\pi i \nu \cdot (P \star T)},$$

where $P = (P_1, \dots, P_d)$. We are then led again to a system of linear equations of the form

$$\mathbf{Z}_{Hg}(t, \nu)_p = \sum_{q \in J} \sum_{m \in J} G(c)_{p, (q, m)} \boldsymbol{\eta}_H(t, \nu)_{(q, m)} \quad (27)$$

where as before

$$\mathbf{Z}_{Hg}(t, \nu)_p = (Z_{T \star P} \circ H)g(t + p \star T, \nu) e^{-2\pi i \nu \cdot p \star T},$$

$$\boldsymbol{\eta}_H(t, \nu)_{(q, m)} = (T_1 P_1 \dots T_d P_d)^{-1} \eta_H(t + q \star T, \nu + (m/T \star P)) e^{-2\pi i \nu \cdot (q \star T)} e^{-2\pi i q \cdot (m/P)},$$

and where $G(c)$ is now a multidimensional finite Gabor system matrix similar to (23).

In order to show that the spreading function for operator H can be completely determined by its response to the periodically-weighted d -dimensional delta-train g , we need to show that (27) can be solved for each $(t, \nu) \in \prod_{\ell=1}^d [0, T_\ell] \times \prod_{\ell=1}^d [0, 1/(T_\ell P_\ell)]$ if $c \in \mathbb{C}^{P_1 \times \dots \times P_d}$ is chosen appropriately.

To see that a choice of c is possible, observe that the product group $\mathbb{Z}_{P_1} \times \dots \times \mathbb{Z}_{P_d}$ is isomorphic to the cyclic group $\mathbb{Z}_{P_1 \dots P_d}$ since the P_ℓ are chosen

pairwise relatively prime. Theorem 4 applied to the cyclic group $\mathbb{Z}_{P_1 \dots P_d}$ guarantees the existence of $\tilde{c} \in \mathbb{C}^{P_1 \dots P_d}$ so that the Gabor system matrix $G(\tilde{c})$ is full spark. We can now define $c \in \mathbb{C}^{P_1 \times \dots \times P_d}$ by setting

$$c_{n_1, \dots, n_d} = \tilde{c}_{n_1 + n_2 P_1 + n_3 P_1 P_2 + \dots + n_d P_1 \dots P_{d-1}}, \quad n = (n_1, \dots, n_d) \in J$$

and observe that $G(c)$ is simply a rearrangement of $G(\tilde{c})$, hence, $G(c)$ is full spark.

6 Further results on operator sampling

The results discussed in this paper are discussed in detail in [24, 5, 26, 42, 41] and [50]. The last listed article contains the most extensive collection of operator reconstruction formulas, including extensions to some $OPW(S)$ with S unbounded. Moreover, some hints on how to use parallelograms to rectify a set S for operator sampling efficiently are given.

A central result in [50] is the classification of all spaces $OPW(S)$ that are identifiable for a given $g = \sum_{n \in \mathbb{Z}} c_n \delta_{nT}$ for c_n being P -periodic.

The papers [43, 41] address some functional analytic challenges in operator sampling, and [28] focuses on the question of operator identification if we are restricted to using more realizable identifiers, for example, truncated and modified versions of g , namely, $\tilde{g}(t) = \sum_{n=0}^N c_n \varphi(t - nT)$. The problem of recovering parametric classes of operators in $OPW(S)$ is discussed in [2, 3].

In the following, we briefly review literature that address some other directions in operator sampling.

6.1 Multiple Input Multiple Output

A Multiple Input Multiple Output (MIMO) channel \mathbf{H} with N transmitters and M receivers can be modeled by an $N \times M$ matrix whose entries are time-varying channel operators $H_{mn} \in OPW(S_{mn})$. For simplicity, we write $\mathbf{H} \in OPW(\mathbf{S})$. Assuming that the operators H_{mn} are independent, a sufficient criterion for identifiability is given by $\sum_{n=1}^N |S_{mn}| \leq 1$ for $m = 1, \dots, M$. Conversely, if for a single m , $\sum_{n=1}^N |S_{mn}| > 1$, then $OPW(\mathbf{S})$ is not identifiable by any collection s_1, \dots, s_N of input signals [47, 40].

A somewhat dual setup was considered in [20]. Namely, a Single Input Single Output (SISO) channel with S being large, say $S = [0, M] \times [-N/2, N/2]$ with $N, M \geq 2$. As illustrated above, $OPW([0, M] \times [-N/2, N/2])$ is not identifiable, but if we are allowed to use MN (infinite duration) input signals g_1, \dots, g_{MN} , then $H \in OPW([0, M] \times [-N/2, N/2])$ can be recovered from the MN outputs Hg_1, \dots, Hg_{MN} .

6.2 Irregular Sampling of Operators

The identifier $g = \sum_{n \in \mathbb{Z}} c_n \delta_{nT}$ is supported on the lattice $T\mathbb{Z}$ in \mathbb{R} . In general, for stable operator identification, choosing a discretely supported identifier is reasonable, indeed, in [28] it is shown that identification for $OPW(S)$ in full requires the use of identifiers that neither decay in time nor in frequency. (Recovery of the characteristics of H during a fixed transmission band and a fixed transmission interval can be indeed recovered when using Schwartz class identifiers [28].)

In irregular operator sampling, we consider identifiers of the form $g = \sum_{n \in \mathbb{Z}} c_n \delta_{\lambda_n}$ for nodes λ_n that are not necessarily contained in a lattice. If such g identifies $OPW(S)$, then we refer to $\text{supp } g = \{\lambda_n\}$ as a sampling set for $OPW(S)$, and similarly, the *sampling rate* of g is defined to be

$$D(g) = D(\text{supp } g) = D(\Lambda) = \lim_{r \rightarrow \infty} \frac{n^-(r)}{r}$$

where

$$n^-(r) = \inf_{x \in \mathbb{R}} \# \{ \Lambda \cap [x, x+r] \}$$

assuming that the limit exists [20, 50].

To illustrate a striking difference between irregular sampling of functions and operators, note that \mathbb{Z} is a sampling set for $OPW([0, 1] \times [-\frac{1}{2}, \frac{1}{2}])$ as well as for the Paley Wiener space $PW([-\frac{1}{2}, \frac{1}{2}])$, but the distribution $g = c_0 \delta_{\lambda_0} + \sum_{n \in \mathbb{Z} \setminus \{0\}} c_n \delta_n$ does not identify $OPW([0, 1] \times [-\frac{1}{2}, \frac{1}{2}])$, regardless of our choice of c_n and $\lambda_0 \neq 0$. This shows that, for example, Kadec's $\frac{1}{4}$ th theorem does not generalize to the operator setting [19].

In [50] we give with $D(g) = D(\Lambda) \geq B(S)$ a necessary condition on the (operator) sampling rate based on the bandwidth $B(S)$ of $OPW(S)$ which is defined as

$$B(S) = \sup_{t \in \mathbb{R}} |\text{supp } \eta(t, \nu)| = \left\| \int_{\mathbb{R}} \chi_S(\cdot, \nu) d\nu \right\|_{\infty}. \quad (28)$$

Here, χ_S denotes the characteristic function of S . This quantity can be interpreted as the maximum vertical extent of S and takes into account gaps in S . Moreover, in [50] we discuss the goal of constructing $\{\lambda_n\}$ of small density, and/or large gaps in order to reserve time-slots for information transmission. Results in this direction can be interpreted as giving bounds on the capacity of a time-variant channel in $OPW(S)$ in terms of $|S|$ [50].

Finally, we give in [50] an example of an operator class $OPW(S)$ that cannot be identified by any identifier of the form $g = \sum_{n \in \mathbb{Z}} c_n \delta_{nT}$ with $T > 0$ and periodic c_n , but requires coefficients that form a bounded but non-periodic sequence. In this case, S is a parallelogram and $B(S) = D(g)$ (see Figure 5)

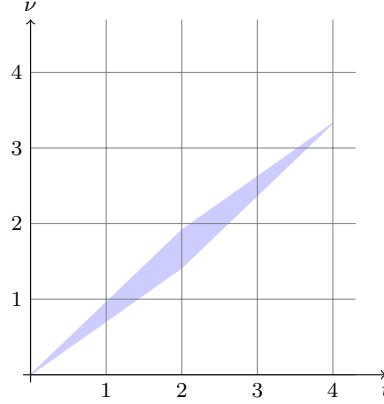


Fig. 5: The operator class $OPW^2(S)$ with $S = (2, 2; \sqrt{2}, \sqrt{2}+1/2)[0, 1]^2$ whose area equals 1 and bandwidth equals $1/2$ is identifiable by a (non-periodically) weighted delta train with sampling density $1/2$. It is not identifiable using a periodically-weighted delta train.

6.3 Sampling of $OPW(S)$ with unknown S .

In some applications, it is justified to assume that the set S has small area, but its shape and location are unknown. If further S satisfies some basic geometric assumptions that guarantee that S is contained in $[0, TP] \times [-1/2T, 1/2T]$ and only meets few rectangles of the rectification grid $[kT, (k+1)T] \times [q/TP, (q+1)/TP]$, then recovery of S and, hence, an operator in $OPW(S)$ is possible [50, 16].

The independently obtained results in [50, 16] employ the same identifiers $g = \sum_{n \in \mathbb{Z}} c_n \delta_{\lambda_n}$ as introduced above. Operator identification is therefore again reduced to solving (18), that is, the system of P linear equations

$$\mathbf{Z}(t, \nu) = G(c)\boldsymbol{\eta}(t, \nu) \quad (29)$$

for the vector $\boldsymbol{\eta}(t, \nu) \in \mathbb{C}^{P^2}$ for $(t, \nu) \in [0, T] \times [-1/2TP, 1/2TP]$. While the zero components of $\boldsymbol{\eta}(t, \nu)$ are not known, the vector is known to be very sparse. Hence, for fixed (t, ν) , we can use the fact that $G(c)$ is full spark and recover $\boldsymbol{\eta}(t, \nu)$ if it has at most $P/2$ nonzero entries. Indeed, assume $\boldsymbol{\eta}(t, \nu)$ and $\tilde{\boldsymbol{\eta}}(t, \nu)$ solve (29) and both have at most $P/2$ nonzero entries. Then $\boldsymbol{\eta}(t, \nu) - \tilde{\boldsymbol{\eta}}(t, \nu)$ has at most P nonzero entries and the fact that $G(c)$ is full spark indicates that $G(c)(\boldsymbol{\eta}(t, \nu) - \tilde{\boldsymbol{\eta}}(t, \nu)) = 0$ implies $\boldsymbol{\eta}(t, \nu) - \tilde{\boldsymbol{\eta}}(t, \nu) = 0$.

Clearly, under the geometric assumptions alluded to above, the criterion that at most $P/2$ rectangles in the grid are met can be translated to the unknown area of S has measure less than or equal to $1/2$.

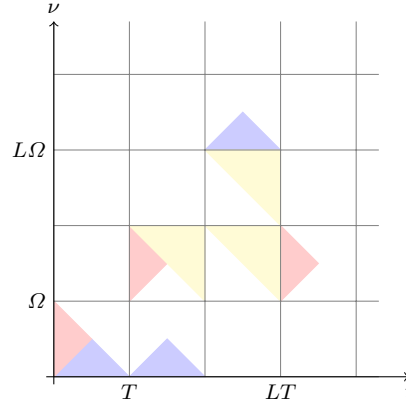


Fig. 6: For S the union of the colored sets, $OPW(S)$ is identifiable even though $7 > 3$ boxes are active, implying that S cannot be rectified with $P = 3$ and $T = 1$ is not possible (see Section 3.3). Recovering η from Hg requires solving three systems of linear equations, one to recover η on the yellow support set, one to recover η on the red support set, and one to recover η on the blue support set. The reconstruction formula (21) does not apply for this set S .

In [16], this area $1/2$ criterion is improved by showing that H can be identified whenever at most $P - 1$ rectangles in the rectification grid are met by S . This result is achieved by using a joint sparsity argument, based on the assumption that for all (t, ν) , the same cells are active.

Alternatively, the “area $1/2$ ” result can be strengthened by not assuming that for all (t, ν) , the same cells are active. This requires solving (29), for $\eta(t, \nu)$ sparse, for each considered (t, ν) independently, see Figure 6 and [50].

It must be added though, that solving (29) for $\eta(t, \nu)$ being $P/2$ sparse is not possible for moderately sized P , for example for $P > 15$. If we further reduce the number of active boxes, then compressive sensing algorithms such as Basis Pursuit and Orthogonal Matching Pursuit become available, as is discussed in the following section.

6.4 Finite dimensional operator identification and compressive sensing

Operator sampling in the finite dimensional setting translates into the following matrix probing problem [45, 7, 6]. For a class of matrices $\mathcal{M} \in \mathbb{C}^{P \times P}$, find $c \in \mathbb{C}^P$ so that we can recover $M \in \mathcal{M}$ from Mc .

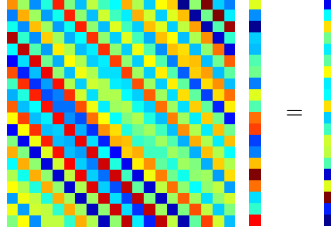


Fig. 7: The matrix probing problem: find c so that the map $\mathcal{M} \rightarrow \mathbb{C}^P$, $M \mapsto Mc$ is injective and therefore invertible.

The classes of operator considered here are of the form $M_\eta = \sum_\lambda \eta_\lambda B_\lambda$ with $B_\lambda = B_{p,q} = \mathcal{T}^p \mathcal{M}^q$, and the matrix identification problem is reduced to solving

$$\mathbf{Z} = M_\eta c = \sum_{p,q=0}^{P-1} \eta_{p,q} (\mathcal{T}^p \mathcal{M}^q c) = G(c) \boldsymbol{\eta}, \quad (30)$$

where c is chosen appropriately; this is just (29) with the dependence on (t, ν) removed.

If $\boldsymbol{\eta}$ is assumed to be k -sparse, we arrive at the classical compressive sensing problem with measurement matrix $G(c) \in \mathbb{C}^{P \times P^2}$ which depends on $c = (c_0, c_1, \dots, c_{P-1})$. To achieve recovery guarantees for Basis Pursuit and Orthogonal Matching Pursuit, averaging arguments have to be used that yield results on the expected qualities of $G(c)$. This problem was discussed in [45, 44, 46] as well as, in slightly different terms, in [1, 18]. The strongest results were achieved in [27] by estimating Restricted Isometry Constants for c being a Steinhaus sequence. These results show that with high probability, $G(c)$ has the property that Basis Pursuit recovers $\boldsymbol{\eta}$ from $G(c) \boldsymbol{\eta}$ for every k sparse $\boldsymbol{\eta}$ as long as $k \leq C P / \log^2 P$. for some universal constant C .

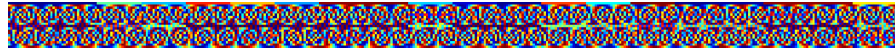


Fig. 8: Time-frequency structured measurement matrix $G(c)$ with c randomly chosen.

6.5 Stochastic operators and channel estimation

It is common that models of wireless channels and radar environments take the stochastic nature of the medium into account. In such models, the spread-

ing function $\eta(t, \nu)$ (and therefore the operator's kernel and Kohn–Nirenberg symbol) are random processes, and the operator is split into the sum of its deterministic portion, representing the mean behavior of the channel, and its zero-mean stochastic portion that represents the noise and the environment.

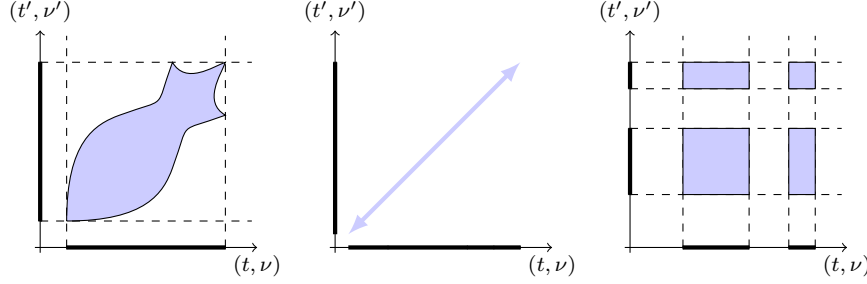


Fig. 9: Support sets of autocorrelation functions, the general case, the WSSUS case, and the tensor case.

The detailed analysis of the stochastic case was carried out in [52, 51]. One of the foci of these works lies in the goal of determining the second-order statistics of the (zero mean) stochastic process $\eta(\tau, \nu)$, that is, its so called covariance function $R(\tau, \nu, \tau', \nu') = \mathbb{E}\{\eta(\tau, \nu) \overline{\eta(\tau', \nu')}\}$. In [52, 51], it was shown that a necessary but not sufficient condition for the identifiability of $R\eta(\tau, \nu, \tau', \nu')$ from the output covariance $A(t, t') = \mathbb{E}\{Hg(t) \overline{Hg(t')}\}$ is that $R(\tau, \nu, \tau', \nu')$ is supported on a bounded set of 4-dimensional volume less than or equal to one. Unfortunately, for some sets $S \subseteq \mathbb{R}^4$ of arbitrary small measure, the respective stochastic operator Paley–Wiener space $StOPW(S)$ of operators with $R\eta$ supported on S is not identifiable; this is a striking difference to the deterministic setup where the geometry of S does not play a role at all.

In [37, 53] the special case of *wide-sense stationary operators with uncorrelated scattering*, or WSSUS operators is considered. These operators are characterized by the property that

$$R\eta(t, \nu, t', \nu') = C_\eta(t, \nu) \delta(t - t') \delta(\nu - \nu').$$

The function $C_\eta(t, \nu)$ is then called *scattering function* of H . Our results on the identifiability of stochastic operator classes allowed for the construction of two estimators for scattering functions [37, 53]. The estimator given in [53] is applicable, whenever the scattering function of H has bounded support. Note that the autocorrelation of a WSSUS operator is supported on a two dimensional plane in \mathbb{R}^4 which therefore has 4D volume 0, a fact that allows us to lift commonly assumed restrictions on the size of the 2D area of the support of the scattering function.

For details, formal definitions of identifiability and detailed statements of results we refer to the papers [37, 52, 51, 53].

7 Appendix: Proofs of Theorems.

7.1 Proof of Lemma 2

In order to see how the time-frequency shifts of c arise, we will briefly outline the calculation that leads to (16). It can be seen by direct calculation using the representation given by (7), that if $g = \sum_n \delta_{nTP}$ then $\langle Hg, s \rangle = \langle \eta_H, Z_{TP}s \rangle$ for all $s \in \mathcal{S}(\mathbb{R})$ where the bracket on the left is the L^2 inner product on \mathbb{R} and that on the right the L^2 inner product on the rectangle $[0, TP] \times [0, 1/(TP)]$. Periodizing the integral on the left gives

$$\langle \eta_H, Z_{TP}s \rangle = \int_0^{1/(TP)} \int_0^{TP} \sum_k \sum_m \eta_H(t + kTP, \nu + m/(TP)) e^{-2\pi i \nu kTP} \overline{Z_{TP}s(t, \nu)} dt d\nu.$$

Since this holds for every $s \in \mathcal{S}(\mathbb{R})$, we conclude that

$$\begin{aligned} (Z_{TP} \circ H)g(t, \nu) \\ = 1/(TP) \sum_k \sum_m \eta_H(t + kTP, \nu + m/(TP)) e^{-2\pi i \nu kTP}. \end{aligned}$$

Given $g = \sum_{n \in \mathbb{Z}} c_n \delta_{nT}$, for a period- P sequence $c = (c_n)$, and letting $n = mP - q$ for $m \in \mathbb{Z}$ and $0 \leq q < P$, we obtain

$$\begin{aligned} g &= \sum c_n \delta_{nT} = \sum_{q=0}^{P-1} \sum_{m \in \mathbb{Z}} c_{mP-q} \delta_{mPT-qT} \\ &= \sum_{q=0}^{P-1} c_{-q} \mathcal{T}_{-qT} \left(\sum_{m \in \mathbb{Z}} \delta_{mPT} \right). \end{aligned}$$

Since for $\alpha \in \mathbb{R}$, the spreading function of $H \circ \mathcal{T}_\alpha$ is $\eta_H(t - \alpha, \nu) e^{2\pi i \nu \alpha}$, we arrive at

$$\begin{aligned} (Z_{TP} \circ H)g(t, \nu) \\ = 1/(TP) \sum_{q=0}^{P-1} c_{-q} \sum_k \sum_m \eta_H(t + kTP + qT, \nu + m/(TP)) \\ e^{-2\pi i (\nu + m/(TP))qT} e^{-2\pi i \nu kTP}. \end{aligned} \quad (31)$$

Letting $m = jP + \ell$ for $j \in \mathbb{Z}$ and $0 \leq \ell < P$, we obtain

$$\begin{aligned}
& (Z_{TP} \circ H)g(t, \nu) \\
&= 1/(TP) \sum_{q=0}^{P-1} c_{-q} \sum_k \sum_j \sum_{\ell=0}^{P-1} \eta_H(t + kTP + qT, \nu + j/T + \ell/(TP)) \\
&\quad e^{-2\pi i \nu q T} e^{-2\pi i \ell q/P} e^{-2\pi i \nu k TP} \\
&= 1/(TP) \sum_{q=0}^{P-1} \sum_{\ell=0}^{P-1} (c_{-q} e^{-2\pi i \ell q/P}) e^{-2\pi i \nu q T} \eta_H^Q(t + Tq, \nu + \ell/TP).
\end{aligned}$$

Finally, replacing t by $t + pT$ for $p = 0, 1, \dots, P-1$, and changing indices by replacing q by $q - p$, we obtain

$$\begin{aligned}
& (Z_{TP} \circ H)g(t + pT, \nu) \\
&= 1/(TP) \sum_{q=0}^{P-1} \sum_{\ell=0}^{P-1} (c_{-q} e^{-2\pi i \ell q/P}) e^{-2\pi i \nu q T} \eta_H^Q(t + (q + p)T, \nu + \ell/TP) \\
&= 1/(TP) \sum_{q=0}^{P-1} \sum_{\ell=0}^{P-1} (c_{-(q-p)} e^{-2\pi i \ell (q-p)/P}) \\
&\quad e^{-2\pi i \nu (q-p) T} \eta_H^Q(t + qT, \nu + \ell/TP).
\end{aligned}$$

The observation that $(\mathcal{T}^q \mathcal{M}^m c)_p = c_{p-q} e^{2\pi i m(p-q)/P}$ completes the proof.

7.2 Proof of Theorem 3

To see why this is true, define $\mu(M)$ to be the number of diagonals of M whose product is a multiple of p_M , and proceed by induction on the size of the matrix M . If M is 1×1 then the result is obvious. Suppose that M is $n \times n$ and that it is described by the vector $\ell = (\ell_0, \dots, \ell_{P-1})$. Assuming without loss of generality that the variable of smallest index in p_M with a nonzero exponent is c_0 , there is a row of M in which the variable c_0 appears ℓ_j times for some index j . Choose one of these terms and delete the row and column in which it appears. Call the remaining matrix M' . The vector ℓ describing M' is $(\ell_0, \dots, \ell_{j-1}, \ell_j - 1, \ell_{j+1}, \dots, \ell_{P-1})$, and is independent of which term was chosen from the given row to form M' . By the construction of the LI monomial, $p_M = c_0 p_{M'}$ and by the induction hypothesis

$$\mu(M') = \ell_0! \cdots \ell_{j-1}! (\ell_j - 1)! \ell_{j+1}! \cdots \ell_{P-1}!.$$

Since there are ℓ_j ways to choose a term from the given row to produce M' we have that

$$\mu(M) = \ell_j \mu(M') = \ell_0! \cdots \ell_{j-1}! \ell_j (l_j - 1)! \ell_{j+1}! \cdots \ell_{P-1}! = \prod_{\kappa=0}^{P-1} \ell_{\kappa}!$$

which was to be proved.

Since each term $a_{\sigma} C^{\sigma}$ in (24) is made up of a sum of precisely this many terms, it follows that exactly one of these terms is a multiple of the LI monomial. Alternatively, we can think of the LI monomial as the one corresponding to the $\sigma \in S_P/\Gamma$ that minimizes the functional $\Lambda_0(C^{\sigma}) = \sum_{i=0}^{L-1} i^2 H(\alpha_i)$ where α_i is the exponent of c_i in C^{σ} and where $H(\alpha_i) = 0$ if $\alpha_i = 0$ and 1 otherwise.

Because by Chebotarev's Theorem, $a_{\sigma} \neq 0$ for all σ the proof works for any square submatrix M , no matter what size. This gives us Theorem 3.

7.3 Proof of Theorem 4

We first need to assert the existence of a cyclical renumbering of the variables such that with respect to the new trivial partition $A' = (A'_{\kappa})_{\kappa=0}^{P-1}$, the CI monomial is given by

$$C^I = \prod_{\kappa=0}^{P-1} \prod_{j \in A'_{\kappa}} c_{j-\kappa}$$

in other words, if $j \in A'_{\kappa}$ then $0 \leq j - \kappa < P$. Note first that since $\min(A'_{\kappa}) = \sum_{i=0}^{\kappa-1} \ell'_i$ for all κ , $j \in A'_{\kappa}$ implies that $j \geq \sum_{i=0}^{\kappa-1} \ell'_i$. Therefore, it will suffice to find a $0 \leq \gamma < P$ such that for all κ , $\sum_{i=0}^{\kappa-1} \ell'_i - \kappa \geq 0$ so that $j - \kappa \geq \sum_{i=0}^{\kappa-1} \ell'_i - \kappa \geq 0$.

Let $0 \leq \gamma < P$ be such that the quantity $\sum_{i=0}^{\gamma-1} \ell_i - \gamma$ is minimized, let

$$\ell' = (\ell'_i)_{i=0}^{L-1} = (\ell_{(i+\gamma) \bmod P})_{i=0}^{P-1},$$

and let $A' = (A'_{\kappa})_{\kappa=0}^{P-1}$ be the corresponding trivial partition. Now fix κ and assume that $\kappa + \gamma \leq P$. Then

$$\begin{aligned} \sum_{i=0}^{\kappa-1} \ell'_i - \kappa &= \sum_{i=0}^{\kappa-1} \ell_{(i+\gamma)} - \kappa \\ &= \left(\sum_{i=0}^{\kappa+\gamma-1} \ell_i - (\kappa + \gamma) \right) - \left(\sum_{i=0}^{\gamma-1} \ell_i - \gamma \right) \\ &\geq 0 \end{aligned}$$

since the second term in the difference is minimal. If $\kappa + \gamma \geq P + 1$ then remembering that $\sum_{i=0}^{P-1} \ell_i = L$

$$\begin{aligned}
\sum_{i=0}^{\kappa-1} \ell'_i - \kappa &= \sum_{i=0}^{\kappa-1} \ell_{(i+\gamma) \bmod P} - \kappa \\
&= \sum_{i=\gamma}^{P-1} \ell_i + \sum_{i=0}^{\kappa+\gamma-P-1} \ell_i - \kappa \\
&= \sum_{i=0}^{P-1} \ell_i - \sum_{i=0}^{\gamma-1} \ell_i + \sum_{i=0}^{\kappa+\gamma-P-1} \ell_i - \kappa \\
&= \left(\sum_{i=0}^{(\kappa+\gamma-P)-1} \ell_i - (\kappa + \gamma - P) \right) - \left(\sum_{i=0}^{\gamma-1} \ell_i - \gamma \right) \\
&\geq 0.
\end{aligned}$$

In order to complete the proof, we must show that $\Lambda(C^\sigma) \geq \Lambda(C^I)$ for all $\sigma \in S_P/\Gamma$ with equality holding if and only if σ is trivial. This will follow by direct calculation together with the following lemma which follows from a classical result on rearrangements of series ([14], Theorems 368, 369). This result is Lemma 3.3 in [34].

First, however, we adopt the following notation. For $0 \leq n < P$, let $b_n = \kappa$ if $n \in A_\kappa$. With this notation, given $\sigma \in S_P/\Gamma$,

$$C^\sigma = \prod_{n=0}^{P-1} c_{(\sigma(n) - b_n) \bmod P}$$

and under the above assumptions,

$$C^I = \prod_{n=0}^{P-1} c_{(n - b_n)}.$$

Moreover,

$$\begin{aligned}
\Lambda(C^\sigma) &= \sum_{i=0}^{P-1} i^2 \alpha_i \\
&= \sum_{i=0}^{P-1} i^2 (\#\{n : (\sigma(n) - b_n) \bmod P = i\}) \\
&= \sum_{i=0}^{P-1} ((\sigma(n) - b_n) \bmod P)^2.
\end{aligned}$$

Lemma 3. *Given two finite sequences of real numbers (α_n) and (β_n) defined up to rearrangement, the sum*

$$\sum_n \alpha_n \beta_n$$

is maximized when α and β are both monotonically increasing or monotonically decreasing. Moreover, if for every rearrangement α' of α ,

$$\sum_n \alpha'_n \beta_n \leq \sum_n \alpha_n \beta_n$$

then α and β are similarly ordered, that is, for every j, k ,

$$(\alpha_j - \alpha_k)(\beta_j - \beta_k) \geq 0.$$

In particular, for every $\sigma \in S_P$,

$$\sum_{n=0}^{P-1} n b_n \geq \sum_{n=0}^{P-1} \sigma(n) b_n$$

with equality holding if and only if σ is trivial.

Proof. The first part of the lemma is simply a restatement of Theorems 368 and 369 of [14]. To prove the second part, note first that b_n is a non-decreasing sequence and in particular is constant on each A_κ . Theorem 368 in [14] states that a sum of the form $\sum_{n=0}^{P-1} \sigma(n) b_n$ is maximized when $\sigma(n)$ is monotonically increasing, which proves the given inequality. Since b_n is constant on each A_κ , it follows that if σ is trivial, then we have equality.

If σ is not trivial then we will show that the sequences $\sigma(n)$ and b_n are not similarly ordered. Letting κ be the minimal index such that A_κ is not left invariant by σ , there exists $m \in A_\kappa$ such that $\sigma(m) \in A_\mu$ for some $\mu > \kappa$, and for some $\lambda > \kappa$ there exists $k \in A_\lambda$ such that $\sigma(k) \in A_\kappa$. Therefore, $b_m = \kappa < \lambda = b_k$ but since $\mu > \kappa$, $\sigma(m) > \sigma(k)$, and so $\sigma(n)$ and b_n are not similarly ordered.

In order to complete the proof, define $\mathcal{C}_1, \mathcal{C}_2 \subseteq \{0, \dots, P-1\}$ by $n \in \mathcal{C}_1$ if $0 \leq \sigma(n) - b_n < P$, and $n \in \mathcal{C}_2$ if $-P+1 \geq \sigma(n) - b_n < 0$ (note that always $|\sigma(n) - b_n| < P$) so that when $n \in \mathcal{C}_2$, $(\sigma(n) - b_n) \bmod P = \sigma(n) - b_n + P$. Let $\sigma'(n) = \sigma(n)$ if $n \in \mathcal{C}_1$ and $\sigma(n) + P$ if $n \in \mathcal{C}_2$, and let $(a_n)_{n=0}^{P-1}$ be an increasing sequence enumerating the set $\sigma(\mathcal{C}_1) \cup (\sigma(\mathcal{C}_2) + P)$. Therefore,

$$\begin{aligned}
\Lambda(C^\sigma) - \Lambda(C^I) &= \sum_{n=0}^{P-1} (\sigma'(n) - b_n)^2 - \sum_{n=0}^{P-1} (n - b_n)^2 \\
&= \left[\sum_{n=0}^{P-1} (\sigma'(n) - b_n)^2 - \sum_{n=0}^{P-1} (a_n - b_n)^2 \right] \\
&\quad + \left[\sum_{n=0}^{P-1} (a_n - b_n)^2 - \sum_{n=0}^{P-1} (n - b_n)^2 \right] \\
&= 2 \left[\sum_{n=0}^{P-1} a_n b_n - \sigma'(n) b_n \right] + \left[\sum_{n=0}^{P-1} (a_n - b_n)^2 - (n - b_n)^2 \right] \\
&= I + II.
\end{aligned}$$

Since a_n is increasing, $I \geq 0$ by Lemma 3, and since $a_n \geq n$ for all n , $(a_n - b_n) \geq (n - b_n) \geq 0$ so that $(a_n - b_n)^2 \geq (n - b_n)^2$ and hence $II \geq 0$. It remains to show that equality holds only if σ is trivial. If $\Lambda(C^\sigma) = \Lambda(C^I)$ then $I = II = 0$. Since $II = 0$, $\mathcal{C}_2 = \emptyset$ for if $a_n \in \sigma(\mathcal{C}_2) + P$ then $a_n > n$ and we would have $II > 0$. Since $\mathcal{C}_2 = \emptyset$, $\sigma'(n) = \sigma(n)$ so that

$$\begin{aligned}
0 &= \Lambda(C^\sigma) - \Lambda(C^I) \\
&= \sum_{n=0}^{P-1} (\sigma(n) - b_n)^2 - \sum_{n=0}^{P-1} (n - b_n)^2 \\
&= 2 \sum_{n=0}^{P-1} (n b_n - \sigma(n) b_n)
\end{aligned}$$

which by Lemma 3 implies that σ is trivial. The proof is complete.

References

1. L. Applebaum and S.D. Howard and S. Searle and R. Calderbank. Chirp sensing codes: Deterministic compressed sensing measurements for fast recovery. *Appl. Comput. Harmon. Anal.*, 26 (2): 283–290, 2008.
2. W.U. Bajwa and K. Gedalyahu and Y.C. Eldar. Identification of parametric underspread linear systems and super-resolution radar. *IEEE Trans. Signal Process.*, 59 (6): 2548–2561, 2011.
3. W.U. Bajwa and K. Gedalyahu and Y.C. Eldar. On the identification of parametric underspread linear systems. *EEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 4088–4091, 2011.
4. Bennet Lee Basore, Noise-like signals and their detection by correlation. Thesis (Sc. D.)-Massachusetts Institute of Technology, Dept. of Electrical Engineering, 1952. <http://hdl.handle.net/1721.1/34322>
5. P.A. Bello. Measurement of random time-variant linear channels. *IEEE Trans. Comm.*, 15:469–475, 1969.

6. R. Belanger-Rioux and L. Demanet. Compressed absorbing boundary conditions via matrix probing. Preprint.
7. J. Chiu and L. Demanet. Matrix probing and its conditioning. *SIAM J. Numer. Anal.* 50 (1):171–193, 2012.
8. D. Donoho and M. Elad. Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ^1 minimization. *Proc. Nat. Acad. Sci.*, 100(5): 2197–2202, 2003.
9. H. G. Feichtinger, W. Kozek, and F. Luef. Gabor analysis over finite abelian groups. *Appl. Comput. Harmon. Anal.*, 26(2):230–248, 2009.
10. P. E. Green, Correlation detection using stored signals. Thesis (Sc. D.)–Massachusetts Institute of Technology, Dept. of Electrical Engineering, 1953. <http://hdl.handle.net/1721.1/34880>
11. P. E. Green, Radar measurements of target scattering properties. in *Radar Astronomy*, J. V. Evans and T. Hagfors, eds. McGraw-Hill Book Company, New York, 1–78, 1968
12. P. Green, Early Spread-Spectrum and automatic equalization – NOMAC and Rake. in *IEEE GLOBECOM 2008: Global Telecommunications Conference, 2008*, 1–5, 2008.
13. K. Gröchenig. *Foundations of Time-Frequency Analysis*. Applied and Numerical Harmonic Analysis. Birkhäuser, Boston, MA, 2001.
14. G. L. Hardy, J. E. Littlewood, and G. Pólya, *Inequalities, Second Edition* Cambridge University Press, Cambridge, UK, 1952.
15. R. V. L. Hartley. Transmission of information. Bell System Technical Journal, 1928.
16. R. Heckel and H. Bölcskei. Identification of sparse linear operators. *IEEE Trans. Inform. Theory*, 59 (12): 7985–8000, 2013.
17. C. W. Helstrom *Statistical Theory of Signal Detection*. Pergamon Press, London, 1960.
18. M.A. Herman and T. Strohmer. High-Resolution Radar via Compressed Sensing. *IEEE Trans. Signal Processing* 57 (6): 2275–2284, 2009.
19. Y.M. Hong and G.E. Pfander. Irregular and multi-channel sampling of operators. *Appl. Comput. Harmon. Anal.*, 29 (2): 214–231, 2010.
20. Y. M. Hong and G. E. Pfander. Irregular and multi-channel sampling of operators. *Appl. Comput. Harmon. Anal.*, 29 (2) 214–231, 2010.
21. T. Kailath. Sampling models for linear time-variant filters. Technical Report 352, Massachusetts Institute of Technology, Research Laboratory of Electronics, 1959.
22. T. Kailath. Correlation detection of signals perturbed by a random channel. *IRE Trans. on Information Theory*, IT-6(3):361–366, June 1960.
23. T. Kailath. Communication via randomly varying channels. Thesis (Sc. D.)–Massachusetts Institute of Technology, Dept. of Electrical Engineering, 1961. <http://hdl.handle.net/1721.1/11319>
24. T. Kailath. Measurements on time-variant communication channels. *IEEE Trans. Inform. Theory*, 8(5):229–236, Sept. 1962.
25. T. Kailath. Time-variant communication channels. *IEEE Trans. Inform. Theory: Inform. Theory. Progress Report 1960–1963*, pages 233–237, Oct. 1963.
26. W. Kozek and G. E. Pfander. Identification of operators with bandlimited symbols. *SIAM J. Math. Anal.*, 37(3):867–888, 2006.
27. F. Krahmer and S. Mendelson and H. Rauhut. Suprema of chaos processes and the restricted isometry property. *Comm. Pure Appl. Math.*, 67 (11): 1877–1904, 2014.
28. F. Krahmer and G. E. Pfander. Localized sampling and local approximation of operators with bandlimited Kohn-Nirenberg symbols. preprint.
29. F. Krahmer, G. E. Pfander, and P. Rashkov. Uncertainty in time-frequency representations on finite abelian groups and applications. *Appl. Comput. Harmon. Anal.*, 25(2):209–225, 2008.
30. H. J. Landau, D. Slepian, and H. O. Pollak, Prolate Spheroidal Wave Functions, Fourier Analysis and Uncertainty – II. *Bell Systems Technical Journal*, 40:43–64, 1961.

31. H. J. Landau, D. Slepian, and H. O. Pollak, Prolate Spheroidal Wave Functions, Fourier Analysis and Uncertainty – II. *Bell Systems Technical Journal*, 40:65–84, 1961.
32. J. Lawrence, G. E. Pfander, and D. F. Walnut. Linear independence of Gabor systems in finite dimensional vector spaces. *J. Fourier Anal. Appl.*, 11(6):715–726, 2005.
33. Y.I. Lyubarskii. Frames in the Bargmann space of entire functions. *Adv.Soviet Math.*, 429:107–113, 1992.
34. R.-D. Malikiosis. A note on Gabor frames in finite dimensions. <http://arxiv.org/abs/1304.7709>, preprint, 2013.
35. H.K. Markey, and G. Antheil, *Secret Communication System, US 2292387*, August 11, 1942.
36. Maxim Integrated (TM) technical staff. An Introduction to Spread-Spectrum Communications, Tutorial 1890. <http://pdfserv.maximintegrated.com/en/an/AN1890.pdf>. *Maxim Integrated Products*, 2003.
37. P. Zheltov, G.E. Pfander, and O. Oktay. Scattering function estimation for overspread radar targets. To appear in *IET Signal Processing*.
38. B. J. Pankowski, Multiplexing a radio teletype system using a random carrier and correlation detection. Thesis (Sc. M.)-Massachusetts Institute of Technology, Dept. of Electrical Engineering, 1952.
39. A. Papoulis, Generalized sampling expansion. *IEEE Trans. Circuits and Systems*, 24(11):652–654, Nov. 1977.
40. G.E. Pfander. Measurement of time-varying multiple-input multiple-output channels. *Appl. Comput. Harmon. Anal.*, 24 (3) 393-401, 2008.
41. G. E. Pfander. Sampling of operators. <http://arxiv.org/abs/1010.6165>, preprint, 2010.
42. G. E. Pfander and D. Walnut. Measurement of time-variant channels. *IEEE Trans. Info. Theory*, 52(11):4808–4820, 2006.
43. G. E. Pfander and D. Walnut. Operator identification and Feichtinger’s algebra. *Sampl. Theory Signal Image Process.*, 5(2):151–168, 2006.
44. G.E. Pfander and H. Rauhut. Sparsity in time-frequency representations. *J. Fourier Anal. Appl.*, 16(2):233–260, 2010.
45. G.E. Pfander and H. Rauhut and J. Tanner. Identification of matrices having a sparse representation. *IEEE Trans. Signal Process.*, 56(11):5376–5388, 2008.
46. G.E. Pfander and H. Rauhut and J.A. Tropp. The restricted isometry property for time-frequency structured random matrices. *Probability Theory and Related Fields*, 156(3-4):707–737, 2013.
47. G. E. Pfander and D. Walnut. On the sampling of functions and operators with an application to Multiple-Input Multiple-Output channel identification. *Proceedings SPIE Vol. 6701, Wavelets XII*; Dimitri Van De Ville, Vivek K. Goyal, Manos Papadakis, Eds, pp. 67010T-1 - 67010T-14, 2007.
48. G. E. Pfander and D. Walnut. Operator Identification and Sampling. *Proceedings of the Conference on Sampling Theory and Applications*, Aix-Marseilles Université, Marseilles, France, 2009.
49. G. E. Pfander and D. Walnut. Sparse finite Gabor frames for Operator Sampling. *Proceedings of the Conference on Sampling Theory and Applications*, Jacobs University, Bremen, Germany, 2013.
50. G. E. Pfander and D. Walnut. Sampling and Reconstruction of Operators. *IEEE Trans. Info. Theory*, submitted, 2014.
51. G.E. Pfander and P. Zheltov. Identification of stochastic operators. *Appl. Comput. Harmon. Anal.*, 36(2):256 – 279, 2014.
52. G.E. Pfander and P. Zheltov. Sampling of stochastic operators. *IEEE Trans. Information Theory*, 60(4):2359 – 2372, 2014.
53. P. Zheltov and G.E. Pfander. Estimation of overspread scattering functions. Preprint.
54. G.E. Pfander. Gabor frames in finite dimensions. In P.G. Casazza and G. Kutyniok, editors, *Finite Frames: Theory and Applications*. Springer, 2013.

55. G.E. Pfander On the invertibility of rectangular bi-infinite matrices and applications in time-frequency analysis, *Linear Algebra and Applications*, 429 (1), 331-345, 2008.
56. R.I. Pickholtz, D.L. Schilling, and L.B. Milstein, Theory of Spread-Spectrum Communications – A Tutorial. *IEEE Transactions on Communications*, 30:855–884, 1982.
57. R. Price, and P. Green, A Communication Technique for Multipath Channels, *Proceedings of the IRE*, 46:555–570, 1958.
58. R. Price, Detectors for Radar astronomy. in *Radar Astronomy*, J. V. Evans and T. Hagfors, eds. McGraw-Hill Book Company, New York, 547–614, 1968
59. R. Price, Further Notes and Anecdotes on Spread-Spectrum Origins. *IEEE Transactions on Communications*, 31:85–97, 1982.
60. R. Price, Statistical theory applied to communication through multipath disturbances. *RLE Tech. Rpt. No. 266*, *Lincoln Laboratory Tech. Rpt. No. 34*, September 1953.
61. R. Price, Optimum detection of random signals in noise, with applications to scatter-multipath communication, I. *IRE Trans. on Information Theory*, IT-2:125–135, December 1956.
62. W. L. Root, and T. S. Pitcher, Some remarks on statistical detection. *IRE Trans. on Information Theory*, IT-1(3):33–38, December 1955.
63. R.A. Scholtz, The Origins of Spread-Spectrum Communications., *IEEE Transactions on Communications*, 30:822–854, 1982.
64. R.A. Scholtz, Notes on Spread-Spectrum History. *IEEE Transactions on Communications*, 31:82–84, 1983.
65. K. Seip. Density theorems for sampling and interpolation in the Bargmann-Fock space. I. *J. Reine Angew. Math.*, 429:91–106, 1992.
66. K. Seip and R. Wallstén. Density theorems for sampling and interpolation in the Bargmann-Fock space. II. *J. Reine Angew. Math.*, 429:107–113, 1992.
67. C. Shannon. Communication in the presence of noise. *Proc. IRE* 37(1): 10-21, 1949.
68. P. Steinhagen and H. W. Lenstra Jr. Chebotarëv and his density theorem. *Math. Intelligencer* 18(2):2637, 1996.
69. G. Turin, Introduction to Spread-Spectrum Antimultipath Techniques and Their Application to Urban Digital Radio. *Proceedings of the IEEE*, 68:328–353, 1980.
70. W. Ward, The NOMAC and Rake Systems. *The Lincoln Laboratory Journal*, 5:351–366, 1992.
71. N. Wiener, *Extrapolation, interpolation and smoothing of stationary time series, with engineering applications*, M. I. T. Press, 1949.
72. L. A. Zadeh, Frequency analysis of variable networks. *Proc. of IRE*, 67:291-299, March 1950.