Exposure of Impact of Cyclostationarity Exploitation on Volterra System Identification

Davide Mattera
Dipartimento di Ingegneria Elettrica
e delle Tecnologie dell'Informazione
Università di Napoli Federico II
Via Claudio, 21, 80125 Napoli, Italy
email: mattera@unina.it

Abstract

This article is devoted to a historical reconstruction of the most important contributions leading to methods of nonlinear time-invariant system identification based on the Volterra system model, the hurdles faced in this work, and the solutions provided by extension and generalization of this methodology from stationary to cyclostationary system inputs, specifically for applications in which system input test signals are under the control of the experimentalist.

I. Introduction

Within the general field of nonlinear modeling involving the use of Volterra system models, the purpose of this article is to describe those contributions for which the cyclostationarity properties of the input and output signals are exploited to advantage for achieving the modeling task. In order to enable the reader to appreciate the value of exploiting cyclostationarity, the presentation begins with an in-depth general reconstruction of key previous contributions, dating back to 1950, to the field including both thematic contributions and their historical interconnections.

A. History of the Development of Modeling of Nonlinear Time-Invariant Systems in Engineering

This subsection is mainly historical in its nature, but it is believed that it is important to have a general view of how approaches to nonlinear modeling evolved over time to fully appreciate the step forward made with the recognition of the advantages offered by cyclostationarity exploitation. In the period of Volterra's life between the year 1882 (when he still was a student of mathematics with the famous Italian mathematician Ulisse Dini) and the year 1887 (when, still very young, he was appointed to the position of full professor in one of the most important Italian Universities), he made three important contributions to the field of functional analysis published in an article [1] written in Italian:

- The first contribution lies in recognizing the need for extending the ordinary concept of *function* on the basis of the relevance for science of such an extended mathematical tool: "in many issues of Physics and Mathematics, and in the integration of partial differential equations, the need arises to consider quantities that depend *on all the values* that one or more functions of one variable assume in given intervals"[1]. Volterra does not introduce a new name for this novel concept, but he does remark on the novelty of this concept: "As it is easy to comprehend, the extension of the concept of function, which we are talking of here, differs essentially from the ordinary one of function of function[2]."), and he remarks that his extension stems directly from the same Dirichlet's abstract definition

[1] Here translated in english from [1]
[2] Meaning composition of two functions

of function. Hadamard later coined the term *functional* to express this concept, though he gave Volterra ample written and public recognition for its introduction.

- The second contribution lies in a first study of the property of continuity and differentiability of such extended functions, an issue that is at the basis of the mathematical properties of functionals and that, however, in [1] is required for the third contribution.
- The third contribution lies in a first and important attempt to provide a general analytical relation between what we today call the input of a system and its output; this attempt is based on the use of the Taylor series expansion and provides what is today a well-known input-output relation. This third contribution has been attributed to Volterra by naming this input-output relation after him.

In the Engineering literature, when Volterra's contributions are discussed, a second contribution [2] is often included. It consists of a multidimensional extension of the concept introduced in his first contribution [1] with reference to the case in which the input signal is two-dimensional or three-dimensional, an extension that Volterra derives on the basis of a geometrical view of the physical phenomena where two- and three-dimensional signals are seen as lines of the plane and of the space.

Following his first contribution, Volterra devoted much of his research to the use of his series expansion for expressing the solutions of some systems of integro-differential equations, which comprise his contribution to what is today called *system representation*. This contribution consisted of determining the properties of the system of integro-differential

equations whose solution can be expressed by his series. The results of his successive extensions of the conditions under which the Volterra series can be used are reported in a paper [3] dated 1910, written in French, and in a collection of seminars [4], dated 1925, written in Spanish and translated into English [5] a few years later in 1930, and republished [6] in the USA in 1948. Along this line of research, an important contribution was made by the French mathematician Frechet who expanded the mathematical conditions under which the Volterra representation is valid. This extension was important because Volterra relied on the validity of the Taylor expansion while Frechet used the convergence of polynomials to any nonlinear function in a compact set.

Because of this line of research, many European mathematicians were aware of the possibility of using the Volterra series to express the solutions of systems of general integro-differential equations when the Second World War broke out and had many consequences regarding scientific research.

The history of nonlinear Volterra modeling is strongly related to that which is today called *system theory* in Engineering, which is addressed next. For a long time, the word *system* was used to express the system of equations that describe physical phenomena. With the passage of time, the term *physical system* also came to be utilized, especially in Physics, when the system of differential equations is used to create a mathematical model of a real mechanism. In engineering applications, the terms *circuit*, *machine*, and *device* came to be used instead to specify the particular physical object.

Once the impact of World War II on scientific research in Europe receded, collaboration among scientists from multiple countries in Europe developed and ultimately led to the abstract conceptualization of various complex engineering objects as systems governed by a unified representation as it is conceived of today in Engineering study. But this was not an easy transition and did not really settle in and encompass the thinking in both Europe and the USA until the early 1960s, 80 years beyond Volterra's introduction of his fundamentally important system model.

To gain insight into the kinds of difficulties that had to be circumvented, the philosophical contribution [7] of the renowned American mathematician, Norbert Wiener, is recommended for reading. This contribution played a seminal role in extending engineering systems theory and method to the study of biological systems. Written jointly with two American scientists, it starts by firmly expressing the need for a unified view of diverse phenomena by providing an abstract description of what is now known as a *black-box* system description: "Given any object, relatively abstracted from its surroundings for study, the behavioristic approach consists in the examination of the output of the object and of the relations of this output to the input. By output is meant any change produced in the surroundings by the object. By input, conversely, is meant any event external to the object that modifies this object in any manner." They use the expression "behavioristic study of natural events" to express their preference for what is today known as the *input-output* system description, as illustrated with the quotation "the behavioristic method of study omits the specific structure and

the intrinsic organization of the object". They distinguish the proposed approach from the traditional approach for which an *entity* is studied for the purpose of determining its "intrinsic organization", "structure", and "properties", and for which "the relations between the object and the surroundings are relatively incidental". The traditional approach came to be called *functional analysis* and the traditional description of systems of equations, came to be called the *input-state-output* description of a system. This brief discussion explains the original reluctance to use the non-popular term *system* for naming the novel input-output concept.

In Wiener's view, the existing Volterra contributions were included inside a general approach that tries to obtain a description of the entity by describing its internal organization and by resorting to its structure and to the properties of its parts. His important contribution to the development of science and engineering lies in circumventing this approach by providing a new method for describing the given object; however, by circumventing the previous approach, he also ignored those parts of the existing literature that could have helped him in developing a general input-output system description, as explained below.

The reason that Norbert Wiener, in his 1943 paper, could challenge the traditional approach to reconstructing the behavior of an object lies in a crucial innovation, which he had introduced, that led to an alternative approach for determining a mathematical description of a given device. This alternative approach first appears in a wartime technical report [8], dated 1942, where he states that two important contributions are given. The first contribution is a first rediscovery of the Volterra contribution: with reference to a specific input-output relation

in a specific nonlinear device (the considered device is an RLC circuit whose resistance is described by a zero-memory quadratic characteristic), he derives its input-output relation which lies within the general Volterra framework. This first contribution is not particularly important because it is a rediscovery of a known concept but, more importantly, because it is a much more limited achievement than that which came before: whereas Volterra and Frechet had shown that the Volterra input-output relation could be used to represent a generality of physical systems and they had made contributions for years increasing the generality of the class of systems that admit this representation, Wiener shows only that the representation holds for a specific electrical device.

In contrast, that which he refers to as the second contribution of the report is a major breakthrough in the history of science and engineering. Prior to Wiener's work, mathematicians attempting to expand the class of integro-differential equations for which the Volterra system description of the solution is valid had to assume the parameters in the integro-differential equations being investigated were all known. But Wiener gave little attention to this. Instead, he departed significantly from prior work by focusing on particular inputs of the system that can facilitate determination of the input-output relation of a completely unknown system. He wrote [8]: "So far we have said nothing about $v(t)$ [the input signal], but we are now ready to make use of the fact that $v(t)$ is a random voltage. This will constitute the second step of the paper and will be accomplished by taking the averages of the random voltages in accordance with known formulas. In these formulas the average is

taken with respect to the parameter α which in going from 0 to 1 runs through all Brownian motions." After deriving the formulas that express the average of $v_1(t)$ [the system output] and $v_1^2(t)$, and after stating that it is possible to evaluate higher-order moments of $v_1(t)$ as well, he wrote: "An average of much importance is that of $v_1(t)v_1(t+\sigma)$: this average is called the auto-correlation coefficient, and its Fourier transform gives the frequency distribution of the square of the voltage. ... The method above, of first solving for the voltage across part of the circuit in terms of the entire voltage, and then getting statistical averages, is clearly quite general".

The method conceived by Wiener and first reported in [8] gave birth to the discipline called system identification. Wiener immediately understood the importance of constructing signal processing devices that could implement the time averages on which his approach to system identification relied as well as the importance of obtaining identification formulas that could be implemented with the available signal processing devices. However, only later did Wiener's group become aware of the importance of the relationship between system representation and system identification. This impeded progress in Wiener's research group for nearly 20 years until long-existing results on Volterra representations were discovered and integrated into Wiener's approach, which was based on time averages of inputs and outputs. With reference to the practical aspects of constructing signal processing devices useful for implementing the time averages, Wiener was committed to providing concrete contributions. With substantial contributions from a previous Ph.D. student of Wiener's, Y.W. Lee, who

returned to MIT in 1946, a new instrument was constructed that was able to implement the statistical averages required by Wiener's method. According to the description later given by Wiener himself in [9]: "We have called this instrument the autocorrelator. Many people at M.I.T., and Lee in particular, have reduced the design of autocorrelators to a surprising degree of perfection". Wiener compares it to the Michelson's interferometer: "Michelson invented an instrument called the interferometer, which was the most delicate machine ever invented for the study of the spectrum of light, and which enabled him to carry out such a seemingly impossible task as the determination of the angles subtended at the earth by some of the fixed stars. . . The analogy between the interferometer and the autocorrelator is deep and significant, and the earlier work of Michelson has given us a whole language for the reading of the results presented by such machines." The results obtained are a major step in human capability for modeling physical phenomena and are described by Wiener as follows: "When the crude original records of brain waves are transformed by the autocorrelator, we obtain a picture of remarkable clarity and significance, quite unlike the illegible confusion of the crude records which have gone into the machine." And, still in the year 1956, when he writes his autobiography, he is aware that many steps ahead need to be taken: "We are at the very beginning of our work in this field, but we have great hopes of what it will offer for the future."

Starting in 1946, and continuing for about two decades, a number of PhD students, under the supervision of Lee, contributed first to the creation of a new instrument for performing

time averages of products of signals, then to its use for nonlinear system identification. In the meantime, already in his 1948 book [10], Wiener closes the gap between (1) his relatively abstract 1942 formulas, which use the statistical expectation over the parameter α describing the system input, and (2) the practical alternative of using only time averages over a single time series. He achieves this by taking advantage of the ergodic properties of the specific input signal that he uses at the system input. With this advance, he creates a solid basis for what becomes the ubiquitous system identification procedure of using time averages, but only for the specific type of input signal that he proposed. The first application of this method of system identification is reported in the 1950 publications [11], [12] where Lee's hardware for calculating the needed correlations is used.

During the 1960s and beyond, following Wiener's work, his mathematical model characterized in terms of the parameter alpha (describing Brownian motion, a random process that he had studied for decades) for his input signal is left behind, and replaced with the Russian Mathematician's, A. N. Kolmogorov's, model of a stochastic process and G. D. Birkhoff's Ergodic Theorem, and it becomes common practice in communications engineering to use other random input signals for linear system identification and to invoke the ergodic theorem to justify the use of time averages instead of the expected values used in the underlying theory for calculating auto- and cross-correlation functions. At that same time, chapters in two books appeared [13], [14] and showed that the abstraction of the Kolmogorov Stochastic Process model and the Birkhoff Ergodic Theorem were unnecessary, because there existed a

probabilistic theory based entirely on time averages that was dual to the more abstract theory of ergodic stationary stochastic processes and was entirely adequate for system identification based on cross correlation measurements. However, there was much enthusiasm about the mathematical rigor of Kolmogorov's Stochastic Process Models, and these two important chapters quickly faded into the background. It was not until the mid-1980s that a book-length comprehensive theory of statistical signal processing, especially spectral analysis, by Gardner [15] (see also [16]) was developed and shown to provide a basis for more concrete theories of statistical signal processing in general, based on time averages. This theory included not only stationary random signals but also a relatively new class of signals referred to as cyclostationary. This new theory provided a basis for a huge jump forward in system identification capability. This topic is returned to in the sequel.

In the 1950s, Wiener's group at M.I.T. began considering the extension of system identification from linear to nonlinear systems. In a series of papers, they initiate an approach to the identification of nonlinear systems. The earliest contribution appears to be the technical report by Ikehara [17] dated December 1951, which is mainly devoted to showing that the replacement of the statistical average (expected value) with the time average is possible also in nonlinear system identification. Prior to this contribution, the use of the term *system* is minimal within Wiener's group. Ikehara states " *on the assumption that the current-voltage function of the nonlinear element and the system transfer function are given . . ." and "the current response of the system to a unit voltage impulse . . .*", but five years later the term

system is fully used throughout the technical report [18] based on the 1956 PhD thesis by Bose (later published in similar form as [19]). Similarly, Wiener in his 1958 book [20] starts using this term, opening the way for its common usage in engineering thereafter. In these two documents, [19] and [20], a general form of input-output relation that is much more powerful than that used in Wiener's 1942 technical report is constructed. However, it is not the Volterra relation, and no study is carried out to determine conditions under which it can represent a nonlinear system, although its similarity with the Volterra relation suggests that it is sufficiently general. Volterra and Frechet are not mentioned in these two documents, suggesting that Wiener was not aware of these earlier contributions.

The first work from the M.I.T. group that mentions Volterra's contributions is Brilliant's technical report [21] dated March 1958, which is based on his Ph.D. thesis. However, the contribution by Volterra that Brilliant cites is the paper [3], which is written in French while neither Volterra's original contribution [1] nor his English contribution [5], [6] were cited. The fact that Brilliant's contribution was mainly devoted to determining the condition(s) for which the Volterra's representation is valid suggests that he was aware of the limitations of the Wiener representation, particularly its strict requirement that the input signal have a Gaussian probability distribution. Finally, in the 1959 M.I.T. technical report [22] by George, Volterra's derivation of his nonlinear system model is recognized, and his paper written in English [5] is cited. The core of the identification formula, later more fully developed by the M.I.T. group, is already present in George's report; moreover, now-standard results in modern

descriptions of nonlinear systems are given. But in the 1960 Ph.D. thesis of Zames [23], which contains one of the first modern descriptions of nonlinear systems, an identification formula for Volterra system identification is still not provided. Similarly, the 1961 PhD Thesis by Schetzen does not mention the Volterra system representation. However, in a technical report dated August 1963, the group of Lee and Schetzen at M.I.T. recognizes the need to adapt the existing identification methods (proposed in the previous year for estimating the class of systems, called Wiener systems) in order to make them viable within the Volterra system framework. This recognition of the superior basis in terms of system characterization, developed in decades of earlier research on the Volterra model, led to the name Volterra-Wiener system, which first appeared in print in [24] and was later used by Schetzen in [25] and Rugh [26].

Outside the research group at M.I.T., use of the Volterra model during the 1950s came primarily from European researchers whose work focused on nonlinear system representation, not on identification. Published work recognizing Volterra's contribution include Barrett [27], [28], [29], [30] and Smets [31]. The end of the pioneering period of research on nonlinear systems is defined by the landmark 1962 paper by Zadeh [32] who clearly establishes the formalization of what a system is and what the problem of system identification is, separate from the problems of system characterization and system classification. This paper clearly identifies the relative merits of the various past contributions, leading to a clear statement of what is known about both the solved and unsolved problems at

that time. A major contribution of Zadeh was to show that system identification needs to be carried out with reference to a system class for which the problem of system characterization has already been completed, and he identifies specifically the Volterra model. The natural result of this clear statement of the state of the art led thereafter to the conversion of the identification formula developed for the Wiener model to a formula for the Volterra model. Following Zadeh's contribution, the Volterra system gained general acceptance as a useful tool for nonlinear system representation and progress was made toward determining methods for its identification, and in particular identification methods based on the averages (correlations) of its input and output signals.

II. Basic definitions regarding Volterra systems

In order to continue our discussion regarding the history of Volterra system identification methods and to insert the contributions deriving from the processing of cyclostationary signals, we first need to introduce some technical definitions and general concepts; this section is introduced to just describe them.

A. Volterra nonlinear input-output representation

For a time-invariant Volterra system, the input-output relation can be expressed as

$$\begin{aligned} y(t) &= h_0 + \int_{\mathcal{R}} h_1(\tau)x(t-\tau)d\tau + \int_{\mathcal{R}^2} h_2(\tau_1,\tau_2)x(t-\tau_1)x(t-\tau_2)d\tau_1 d\tau_2 + \ldots = \\ &= h_0 + \sum_{n=1}^{\infty} \mathcal{H}_n[x(t)] \end{aligned} \tag{1}$$

This expression constitutes a generalization to functionals of the general Taylor expression for ordinary functions, first introduced by Volterra.

As a recognition to Volterra's contributions, this representation has taken the name of Volterra series representation and the functions $h_n(\tau_1, \ldots, \tau_n)$ are called Volterra kernels of the system. Moreover, the general term

$$\mathcal{H}_n[x(t)] = \int_{\mathcal{R}^n} h_n(\tau_1, \ldots, \tau_n) x(t-\tau_1) \ldots x(t-\tau_n) d\tau_1 \ldots d\tau_n \tag{2}$$

is called the Volterra operation of order n. The Fourier transformations of the Volterra kernels $H_n(f_1, \ldots, f_n)$ are called the Volterra transfer functions.

The term h_0 represents the output of the system resulting from the null input ($x(t) \equiv 0$). The equation (2) can be considered an n-dimensional convolution integral. For $n = 1$ we obtain the linear component of the system response

$$\mathcal{H}_1[x(t)] = \int_{\mathcal{R}} h_1(\tau) x(t-\tau) d\tau_1 \tag{3}$$

where $h_1(\tau)$ represents the impulsive response of the linear component of the overall system. For $n = 2$ (2) represents the quadratic component (or second-order component) of the system representation

$$\mathcal{H}_2[x(t)] = \int_{\mathcal{R}^2} h_2(\tau_1, \tau_2) x(t-\tau_1) x(t-\tau_2) d\tau_1 d\tau_2 \tag{4}$$

where $h_2(\tau_1, \tau_2)$ is the second-order Volterra kernel that represents the impulsive response (in two dimensions) of the quadratic component of the overall system and, analogously, for successive values of n.

Analogously to a linear system that is completely characterized by its impulsive response, a nonlinear system that can be represented by the Volterra series is completely characterized by its Volterra kernels.

The Volterra kernels are usually assumed to be symmetric functions of their variables, i.e., each of the $n!$ possible permutations among the variables $\tau_1, \ldots, \tau_n$ leaves $h_n(\tau_1, \ldots, \tau_n)$ unchanged; for example, in the case $n = 2$, this means that $h_2(\tau_1, \tau_2) = h_2(\tau_2, \tau_1)$.

Alternatively, in order to reduce the number of terms in the representation, the Volterra kernel can be assumed to be triangular[3]. When no assumption is introduced (neither symmetry nor triangularity), an infinite number of equivalent representations exist. For this reason, one of these two assumptions is usually introduced.

B. Discrete-time systems

The input-output relation (1), when discrete-time systems are considered, becomes:

$$\begin{aligned} y(n) &= h_0 + \sum_{j_1} h_1(j_1)x(n-j_1) + \sum_{j_1,j_2} h_2(j_1,j_2)x(n-j_1)x(n-j_2) + \ldots = \\ &= h_0 + \sum_{i=1}^{\infty} \mathcal{H}_i[x(n)] \end{aligned} \tag{5}$$

with

$$\mathcal{H}_i[x(n)] = \sum_{j_1,\ldots,j_i} h_i(j_1,\ldots,j_i)x(n-j_1)\ldots x(n-j_i). \tag{6}$$

If only the first N sums in (5) are not identically zero, then we call this a finite-order Volterra model. In this case the problem of approximating the behavior of the system by means of

[3] Given n distinct values $\tau_1, \tau_2, \ldots, \tau_n$ the symmetric functions assume the same value in any permutation of the point $(\tau_1, \tau_2, \ldots, \tau_n)$ while the triangular functions are nonnull only in just one of such permutations, typically just in that specific point (among the $n!$ permutations) with increasing or decreasing ordering of the values of the components. The definition is analogously extended to the case in which the components τ_i for $i = 1, 2, \ldots, n$ are non distinct.

a Volterra series expansion is simpler. When the discrete-time system has finite memory L, then the input belongs to a finite closed subset of $\mathcal{R}^L$, then, from the Stone-Weierstrass theorem it follows that every continuous system (i.e., one for which small input variations imply small output variations) can be expressed [33] in a finite-order Volterra discrete-time expansion (with finite-dimension input-signal space), also in the case for which we adopt the most rigorous approximation criterion, that of uniform approximation. Infinite-memory systems results similar to those obtained in [34] with reference to continuous systems do not seem to be present in the literature.

If the discrete-time model is obtained from uniform time-sampling of continuous-time signals, not much attention has been dedicated to the proper sampling procedure, i.e., to the conditions that guarantee the existence of a discrete-time system representation that correctly relates the sampled versions of the input and output signals. For example, it is often not noted that, for a low-pass input signal and finite-order analog Volterra system, the higher the system order is, the larger the output-signal bandwidth becomes and, therefore, a larger sampling frequency is needed to guarantee the existence of a discrete-time Volterra system representation. It is interesting to note that the complexity of the representation in Volterra series increases with the quality of required approximation, in the sense that a larger number of terms of the Volterra kernels are needed. When the sampling interval is reduced, the memory of the discrete-time representation correspondingly increases; however, the consequent oversampling of the Volterra kernels may enable a reduction in the number

of significant terms in the Volterra input-output representation.

C. Alternative representations in comparison with the Volterra series

Thanks to its direct and natural interpretation as an extension of the Taylor series to functionals, the Volterra series was the first model adopted for mathematically representing the behavior of nonlinear physical systems. However, other approaches for nonlinear system representation have been introduced such as empirical approximation, descriptive function, input-state-output representation; in particular, some authors [35], [36], [37], [38] have shown the relation between bilinear models and Volterra systems.

Other series expansions, different from the Volterra series, have been introduced for representing the nonlinear input-output system relation such as those of Cameron-Martin [39], Fourier, Taylor-Cauchy [40], and Wiener [20], [41], [42]. In particular, the Wiener series is often used in identification methods when the input signal is Gaussian. Methods that can operate with a different series expansion can be also used to identify the Volterra kernels; in particular, the relation between the kernels of the Wiener series and those of the Volterra series are well known.

Some interesting descriptions of a nonlinear system resort to the interconnections of dynamical linear systems (L) and zero-memory nonlinearities (N). The most popular type of system in this class is the cascade system: a linear system with memory followed by a nonlinear zero-memory element in cascade with another linear system with memory (denoted by LNL), also sometimes called the general model. A subclass of this type, that of the LN

kind, takes the name of Wiener model (the name is due to the use of such a model in his first 1942 contribution [8]); another of the type NL is called Hammerstein model. The cascaded block model can be useful when the order of the Volterra system, which is needed in order to achieve a reasonable approximation, is too large to be managed in practice. However, it can be seen as prior information regarding the Volterra model, which should be used by the identification procedure in order to improve its quality.

The Uryson model, introduced in [43], consists of many Hammerstein models in parallel; also more general models, more complex than these, have been introduced. Among them, the large set of neural networks have gained much interest in the scientific literature after the first experiment carried out by Rosenblatt on the LN Wiener model. The experiments by Rosenblatt dated 1964, in the years when the Volterra model was gaining its position as the main approach to nonlinear modeling, showed that it is not so important to specify the nonlinearity but a great representation capability is obtained in the LN cascade by setting the nonlinearity fixed to a sharp sigmoidal characteristics and adapting the coefficients of the linear part.

Such representation capability is verified by showing that the nonlinear system behavior reconstructed through a proper processing of a subset of the available input-output signals is close to the ideal and unknown nonlinear system also when operating on the remaining part of the measured input-output signals, without an analytical verification of the formal correspondence between the ideal system and the reconstructed one. In such an approach,

the search for the coefficient values of the linear part resorted to the inductive criterion of empirical error minimization, where the adjective "empirical" refers to the use of a given set of measured data and "error minimization" means for which the parameters of the linear part are adaptively modified in order to minimize an error function that expresses the quality of the approximation of the current system on the available input-output data.

Actually, the error function could be expressed by resorting to input-output signal averages and carrying out its minimization by using the result of such a processing, though the adaptive algorithms do not pass for the time averages. The advantage of the considered approach lies in its capability of obtaining acceptable results also when the number ℓ of available data is much smaller than that necessary to guarantee the convergence of the time averages. On the other hand such methods are not able to provide acceptable performance when the value of the output signal-to-noise ratio SNR_y is too small.

A line of research, much of which was contributed by Wiener, had indicated that the LN model (where N is a fixed sigmoidal zero-memory nonlinearity and L is a linear processing structure) can represent the behavior of the brain neuron. This motivated the name *perceptron* attributed to the Rosenblatt model and the introduction of nonlinear models, usually named neural networks, where linear systems with memory and nonlinear zero-memory subsystems are interconnected in layered structures that resemble the brain structure and makes the dependence of the output signals on the system parameters highly nonlinear.

More than twenty years passed from the time of the Rosenblatt experiments before (in

the year 1986) a new algorithm for adaptive reconstruction from the available data, of the parameters of the considered layered nonlinear system was proposed. The proposed algorithm is able to determine a local minimum of the cost function but its performance is heavily dependent on many important parameters. A procedure for assigning them specific values in terms of the given data is not available. Many experiments have demonstrated the modeling capability of this approach but the empirical dependence of the modeling procedure on many parameters whose specification depends on a trial-and-error procedure constitutes a strong limit on the utility of the procedure.

The difference between the classical identification method and the neural networks approach remains so large that the comparison of the relative merits of the two approaches often remains only qualitatively stated in that a general setting with which they can be quantitatively compared is missing. Neural networks remain important in scenarios where the construction of a mathematical model from data is important also in implicit form, i.e., if the determined model, though involved and unreadable, is sufficiently accurate and the output corresponding to a given input can be obtained by software calculations in a reasonable time. The procedure is coherent with available data but it can be designed only after a long set of trials, and without a specific guarantee that it will continue to show coherence with data when its behavior is verified on new data. The procedure is particularly useful in a scenario with many input signals and small amount of available data, and, therefore, when sufficient time is available in order to continue to test the constructed model and improve its refinement.

When a good implicit model is available in terms of a neural network model, it may be interesting to derive an equivalent Volterra model but, however, the diffuse utilization of sigmoidal zero-memory nonlinear systems with the neural model makes it impossible to find a low-order Volterra model, so such equivalent models have not been investigated.

The highly nonlinear dependence of the system output on the model parameters in the neural networks approach limits the capability of the theoretical analysis and implies the need to resort to trial-and-error approaches. For this reason, a few years after 1986, when such theoretical difficulties became clear, an important property of the Volterra model was recognized. This property is the linear dependence of the output signal on the model parameters. Instead of using the specific Volterra nonlinear expansion, it started to be considered the possibility to use a two stage structure, named generalized linear models, where in the first stage there is a fixed nonlinear processing of the input signal generating a certain number of different intermediate signals and in the second stage there is a linear processing of such intermediate signals. The possible choices of the first stage include the Volterra products but we can consider also other choices more suited for modeling sharp nonlinearities.

The trial-and-error approach often dominates the choice of the first stage but some attempts have been made to introduce a slow (i.e., dependent on the processing of large amount of data) adaptation of the first stage. This is probably the most interesting attempt to compensate for Volterra limitations in synthetically modeling systems with sharp nonlinear behavior.

The most important choice for the first stage of the generalized linear models is the kernel model where the number of signal outputs of the first stage is as large as the number ℓ of available data, i.e., such a number is chosen to be high in order to increase the modeling capability in the presence of limited data. In any case this approach has often been used when the amount of available data is small and the available measurements indicate the output SNR is high.

We are not aware of successful use of the generalized linear model in a scenario characterized by a large amount of data and noisy input measurements. In fact, as for the identification procedures proposed, the perceptron and the neural networks and those proposed for the generalized linear models are not able to provide acceptable performance when the output signal measurements are noisy.

D. General Issues in System Identification

For the general system identification problem, the following kinds of identification methods need to be considered:

- Full control of the system input: in this case the signal at the system input can be chosen by the system-identification-instrument designer under consideration of the available identification methods for the input-signal candidates being considered; alternative methods should be compared on the basis of their predicted or measured performance, depending on which of these approaches is possible for the given system and the selected input. Two subcases need to be distinguished for consideration: real-valued

input signals and complex-valued input signals (e.g., complex low-pass representation of real bandpass signals); the case of real signals is dominant because physical systems (in contrast to mathematically prescribed systems) cannot be driven by complex-valued input signals. However, advantages do accrue when complex input signals are possible.

- No control of the system input: in this case, the designer does not have the possibility of choosing the input signal and must perform system identification using the given input. However, it can be assumed that the input signal has some known characteristics but it cannot be forgotten that the values of the input signal are the result of a measurement process.

With reference to the output-signal measurements in both of these cases, and with reference to the input-signal measurements in the second case above, it should generally be assumed that a disturbance signal, such as measurement error, affects the available measurements. The acronym SNR_y is used here to denote the output signal-to-noise ratio (ratio of average signal power to average noise power) for both above cases and for the second case the acronym SNR_x is used. In addition, it is useful to designate a notation for the number of time samples of the input signal that are available for system identification, because comparisons among candidate methods need to take the value of this parameter that is used for each method into account. It is also helpful to designate a notation for the maximum memory length of the Volterra model. The symbol L is used here. Another important parameter needed in system identification is the maximum polynomial order N to be considered for use in

the Volterra model, which is typically taken to be the smallest order that is sufficient to accurately represent the system to be identified. Although the Volterra model can be used with a very large order to maximize the representation capability of the given system, in practice the model order used needs to be limited so as to control computational complexity and the impact of calculation errors. The values for the couple (N, L) for the Volterra model must be selected in accordance with some assumed knowledge of, or measurements on, the system to be identified. Although this is an important challenge in practice, it suffices for the purposes of this review article to focus on describing system identifications methods and simply identifying the known promises and limitations of these methods for purposes of making general comparisons among the methods.

III. Standard methods for Volterra system identification

The identification of the Volterra kernels $\{h_i(\tau_1, \ldots, \tau_i)\}_{i=1,\ldots,N}$, where N denotes the selected order of the Volterra system model, is a very difficult problem both because it is difficult to separate the contributions on the output of the various nonlinear kernels and because a large number of computations are required by the various identification methods.

The most robust approach to Volterra system identification left from the Wiener group is the following identification formula

$$h_N(\tau_1, \ldots, \tau_N) = \frac{1}{N!\, A^N} \ \langle y(t) x(t-\tau_1) \ldots x(t-\tau_N) \rangle \qquad \tau_i \neq \tau_j \ \text{ for } \ i \neq j \quad (7)$$

where N (assumed finite and already known) is the maximum order term in the polynomial representation of the system to be identified and the symbol $\langle \cdot \rangle$ denotes the time average,

which for a continuous signal $x(t)$ can be written as

$$\langle x(t) \rangle = \lim_{T \to +\infty} \frac{1}{2T} \int_{-T}^{T} x(t) \mathrm{d}t \tag{8}$$

The method is not explicitly stated in the form reported in (7) since also in his later contributions, Schetzen emphasizes the importance of carrying out the identification approach with reference to the class of Wiener systems, but its derivation is rather straightforward from the overall contribution in [41]. An important consequence of its being rooted in the Wiener approach to nonlinear modeling is that the authors do not notice that the two assumptions they introduce are not both necessary. The first necessary assumption is that the input signal $x(t)$ is white (and the quantity A in (7) is the constant value of its power spectral density) while the second assumption is that the density of the input signal is Gaussian. The same care taken in demonstrating the correspondence between the statistical mean and the time average shows that they believe that it is a critical assumption but they do not introduce any concern with reference to the practical issue of the possibly non-Gaussian density of the true input

signal, which should have merited[4] much more attention if it were a critical assumption. However, it is easy to verify that the identification formula (7) holds for any fraction-of-time density of a strict-sense white input signal whose odd-order moments are null (i.e., $\langle x^{2n+1}(t)\rangle = 0$ for $n \geq 0$). It is, however, important to recognize that, for continuous time, the only known forms of strict-sense white noise are Gaussian, white Poisson impulse noise, and a sum of these, whereas for a discrete time process, any amplitude distribution of an independent identically distributed sequence of random variables is a valid strict-sense white process.

The Wiener method implicitly assumes that the input signal is stationary; the great care taken with its design and with the proofs concerning the use of the time averages has a by-product consequence that $x(t)$ must be assumed to be stationary. This assumption is maintained (often implicitly) in the successive classical contributions on Volterra system

[4] At that times, in the first Wiener's contributions, it happened the opposite of what became usual in the successive decades, also in spite of the repeated Gardner's reminds on the risks deriving from an automatic straightforward replacement of the statistical averages with time averages: obliged by the available analog processing device constructed by Lee to carry out time averages, Wiener showed always aware that replacing statistical means with time averages was a critical issue and gave a mathematical justification of this replacement, establishing it on the solid basis of the Birkhoff ergodic theorem (a care motivated by the logical value of its scientific construction and carried out also despite his unfriendly personal relation with Birkhoff). It is reasonable to deduce that this initial caution used by Wiener had created a general feeling (especially among those who had not taken care to study his demonstrations) that Wiener had already completed all the demonstrations needed to carry out such replacement in any context.

identification before Gardner's introduction of the use of cyclostationary inputs to great advantage, as explained in the sequel. Moreover, when the stochastic process theoretical framework is used, even though only time averages are of interest, the process can be nonstationary as long it is asymptotically-mean stationary as defined by R.M. Gray in the mid-1980s. This guarantees that each sample path of the process is stationary in the fraction-of-time probability framework, which is all that is needed for methods based on time averages. Similarly, Gardner's methods are valid for generally nonstationary processes that are asymptotically-mean cyclostationary which guarantees its samples paths are cyclostationary in the fraction-of-time sense. Because expected values are of no interest when only time averages are used, none of the stochastic process models need to be ergodic or cycloergodic.

Since (7) allows the estimation of only the highest-order Volterra kernel, the lower-order Volterra kernels must be estimated by subtracting the contributions of the already known kernels from the system output. It is rather direct to note, however, that, because the odd-order moments are all zero, the same identification holds also for the order $N-1$

$$h_{N-1}(\tau_1,\ldots,\tau_{N-1}) = \frac{1}{(N-1)!\,A^{N-1}}\ \langle y(t)x(t-\tau_1)\ldots x(t-\tau_{N-1})\rangle \qquad \tau_i \neq \tau_j \ \text{ for } \ i \neq j \tag{9}$$

so that the subtraction of the last identified term of order, say, M in the Volterra polynomial enables the identification formula to be used for both orders $M-1$ and $M-2$.

We have started by presenting the most robust identification procedure developed by the

M.I.T. group but in previous works a great deal of theoretical effort had been done by Wiener to avoid the critical assumption in (7), i.e., that $\tau_i \neq \tau_j$ for $i \neq j$. The final robust procedure is the result of the practical approach to laboratory experiments of the group of Lee and allowed them to considerably simplify the identification procedure.

Schetzen [41] states that "*in practice, we can use (7) to determine the kernels as close as desired to those points at which two or more τ's are equal, and so it normally is not necessary to determine the kernels exactly at those points.*" The processing device to which Schetzen is referring to is an analog device, though implemented with some digital electronics; in particular, notice that, also with analog processing, the kernel $h_N(\tau_1, \ldots, \tau_N)$ is evaluated at a finite number of values, so that the kernel is evaluated from a finite number of samples. However, in comparison with discrete-time processing, the sampling of the kernel arguments can be non-uniform and the values of the delays $\tau_1, \tau_2, \ldots, \tau_N$ can be rather arbitrarily chosen.

However, one should consider that the condition $\tau_i \neq \tau_j$ should be more precisely written as $|\tau_i - \tau_j| \geq \tau_{\min}$ where $\tau_{\min}$ is the smallest positive value of τ for which the following relation[5] holds:

$$\langle x(t)x(t-\tau)\rangle \ll \langle x^2(t)\rangle \tag{11}$$

[5] When the standard assumption of white Gaussian signal is replaced by that of strict-sense white signal, the condition (11) is replaced by the following one:

$$\left\langle x^{2n+1}(t)x^k(t-\tau)\right\rangle \ll \left\langle x^{2n+k+1}(t)\right\rangle \qquad \forall n, k \geq 0 \tag{10}$$

In practice, $\tau_{\min}$ is not zero and, therefore, the possibility of interpolating its values from those that can be estimated by using the formula (7) depends on the characteristics of the system kernels. For example, the third-order kernel of the system $y(t) = \int h(\tau)x(t-\tau)d\tau + x(t)x(t-100\tau_{\min})x(t-200\tau_{\min})$ can be estimated by using (7) but that of the system $y(t) = \int h(\tau)x(t-\tau)d\tau + x(t)x(t-\tau_{\min}/100)x(t-\tau_{\min}/200)$ cannot. Therefore, for some systems, those for which the kernel memory is shorter than the support of the autocorrelation (i.e., shorter than the inverse of the input signal bandwidth), the robust procedure cannot be used.

This identification method is specifically related to the analog description and has therefore been reported in its analog version. When we consider the discrete-time description, the relation should obviously be replaced by the following one:

$$h_N(k_1,\ldots,k_N) = \frac{1}{N!\,A^N}\ \langle y(k)x(k-k_1)\ldots x(k-k_N)\rangle \qquad\qquad k_i \neq k_j \quad \text{for} \quad i \neq j \tag{12}$$

where the symbol $\langle\cdot\rangle$ still denotes the time average, which for a discrete-time $x(k)$ can be written as

$$\langle x(k)\rangle = \lim_{K\to+\infty} \frac{1}{2K+1} \sum_{k=-K}^{K} x(k) \tag{13}$$

In this case, the condition $k_i \neq k_j \ \text{ for } \ i \neq j$ does not need further specification. However, the possibility to interpolate the diagonal kernel values (those of the main diagonal and those of the minor diagonals) from the off-diagonal ones, which can be estimated by (12), depends on the characteristics of the system. Also when the discrete-time system is derived from a

continuous-time representation of a physical system by uniform sampling with sufficiently high sampling rate, such interpolation is possible when the sampling rate is sufficiently large but the sampling rate is limited by the requirement of insuring that the sampled version of the input signal is white, so that there is no guarantee that the method in (12) can be applied in any case, without careful consideration of the time-sampling rate and the input bandwidth (assumed to be finite as it would be in practice, instead of infinite as assumed mathematically). Moreover, interpolating the diagonal kernel values from the off-diagonal ones is prone to error propagation since the estimated contributions coming from the higher-order kernels need to be removed from the overall output in order to enable identification of lower-order kernels as if each was the highest order kernel.

The problem of estimating the diagonal elements was already clear to Wiener, and, in fact, an identification formula that holds also for diagonal elements had been derived by Wiener. It has the following form

$$h_N(\tau_1,\ldots,\tau_n) = \frac{1}{N!\,A^N}\left\langle \left\{ y(t) - \sum_{m=0}^{N-1} y_m(t) \right\} x((t-\tau_1)\ldots x(t-\tau_N) \right\rangle \qquad (14)$$

where the signal $y_m(t)$ ($m \in \{0,1,\ldots,N-1\}$) is the result of a lower-order cross-correlation procedure. The details for constructing the compensation terms $y_m(t)$ are popular in the literature [41]; we do not dwell on this because the removal errors resulting from estimated terms makes the method slow to converge, so that Lee's group finished (after years of vain attempts) by proposing the use of (7), and neglecting the presence of systems for which the identification of diagonal terms is important. It is noted here that the definition of Wiener's

terms $y_m(t)$ depends on the assumption that the white input signal is Gaussian. Nevertheless the literature is full of successive works that refer to the impossibility of estimating the diagonal terms by using Wiener's approach in (14). The authors of [44] showed a restriction of the set of possible systems for which the procedure makes sense when continuous-time systems are considered; moreover, they had reported the practical difficulties leading to '*large errors in the corresponding "diagonal integral"*' (the "diagonal integrals" are integrals along diagonal values of the higher-order kernels that affect the estimated lower-order kernel) so that '*for third or higher order systems, this problem of the Lee-Schetzen method cannot be easily avoided*'. They also mention the Marmarelis thesis [45] by reporting that a decrease of accuracy had been noticed in a few attempts to measure third order kernels. However, the doubts raised by the authors of [44] about "*fundamental difficulties of the Lee-Schetzen method rather than to computational errors*" seem too severe, and Wiener's proposal seems to be affected by a common problem in the identification approach: any identification formula is asymptotically exact, but only experimental results are able to provide an idea of how many measured samples are needed to make the estimation error negligible. With the technical and theoretical difficulties of the Fifties and the Sixties (the reader interested in the processing capability available in that period can read [46]), it was not possible to verify the Wiener method by carrying out an experiment with a sufficiently large number of data points and this led to growing doubts about the reasons for which an asymptotic method seemed to not converge. Nevertheless, Wiener's identification formula (14) exhibits a large variance of the

estimation error and, even with modern processing technologies [47], the problem of using it for estimating the diagonal kernel values still holds.

The convergence problem of the identification formula was emphasized by Wiener's choice to try to estimate the kernels of a system representation that depended on the same assumption regarding its input (to be white and Gaussian); and, different authors of the period, rather than concentrating on a system representation independent of the system input properties, tried to generalize the approach by creating a set of possible representations of the same system which depended on the input properties. Ogura [48], Segall and Kailath [49], and Brillinger [50] repeated the Wiener approach for Poisson processes. Klein and Yasui [51] reach the extreme of this trend by defining the Fourier kernels, i.e., a set of kernels useful to specify the behavior of the same system, each useful for specific characteristics of the input signal. The name Fourier is used to specify the generality of the approach, capable of finding many possible representations of considered system, suited to the specific characteristics of the input signal.

Therefore, also during the Seventies, there still remained a gap between studies that emphasize the generality of the Volterra representation and the identification approaches that simplify their task by using other representation with kernels, more easily identified than the Volterra kernels but dependent on the same characteristics of the signal input. This loss of interest in the estimation of Volterra kernels, rather than true technical difficulties, resulted in failure to solve the identification problem for the diagonal values of the Volterra

kernels. The problem was so neglected that even the methods that appeared in the scientific literature which solved it were ignored. It is instructive to read how Schetzen concludes the first part of [41] where he acknowledges the generality of the Volterra system representation: "*If the order of the system is not finite, then, generally, the contribution of each Volterra operator cannot be separated from the total system response. For this case, there is no exact method for the measurement of the system Volterra kernels, and various approximation techniques must be used.*" In synthesis, up to the Seventies, the general view was that a general representation was too difficult to identify, and therefore, attention was dedicated only to a special representation, suited to the statistical properties of the input signal. After this change in direction occurred, the identification of general representations became a neglected problem and some useful contributions that appeared in the literature were practically never mentioned in the successive literature.

A. Non-Gaussian white input signals

Since the identification formula (12) still holds for any strict-sense white non-Gaussian signal, the considered extension to non-Gaussian inputs is trivial. Also in this case, the problem that remains to be solved is the identification of the diagonal kernel values for the Volterra representation. But, as just stated, rather than solving this open problem, many researchers focused on only identification of specific representations dependent on the input characteristics.

Some researchers investigated the identification problem for particular statistics of the

white input signal; there is a list of contributions that follows in this line. However, the slowness in resolving the problem of the diagonal values in kernel estimation has also fundamentally slowed research regarding proper choice of the time-series distribution for applications in which the experimentalist can choose the input signal. There is no complete comparison of the performance associated with each possible choice of the input distribution; however, the existing literature suggests that abandonment of the Gaussian input signals is associated with improved performance of the system identification procedures.

The solution of the problem of the diagonal kernel values in correlation-based methods is returned to in the sequel.

B. Correlation methods for colored input Gaussian signals

There are basically two main contributions in this line of work; they arrived in about the same period. Of the two contributions, the first one arrived gradually; it can be found formalized in [41] but the main basic steps can be found already in Schetzen's 1961 thesis report [52]. In fact, in [41] Schetzen showed how to apply his method (relative to the case of white Gaussian signal) when the input signal is Gaussian but not white. The power spectral density $S_x(j2\pi f)$ is factored as follows: $S_x(j2\pi f) = S_x^+(j2\pi f)S_x^-(j2\pi f)$ where $S_x^+(j2\pi f)$ is the complex conjugate of $S_x^-(j2\pi f)$ and all the poles and zeros of $S_x^+(s)$ possess negative real components. The method utilizes the whitening filter so that, with reference to the figure 1, defined a first filter with frequency response equal to the inverse of $S_x^+(j2\pi f)$, a nonlinear system $S^{'}$ can be considered whose input, denoted by $z(t)$, is white and Gaussian.

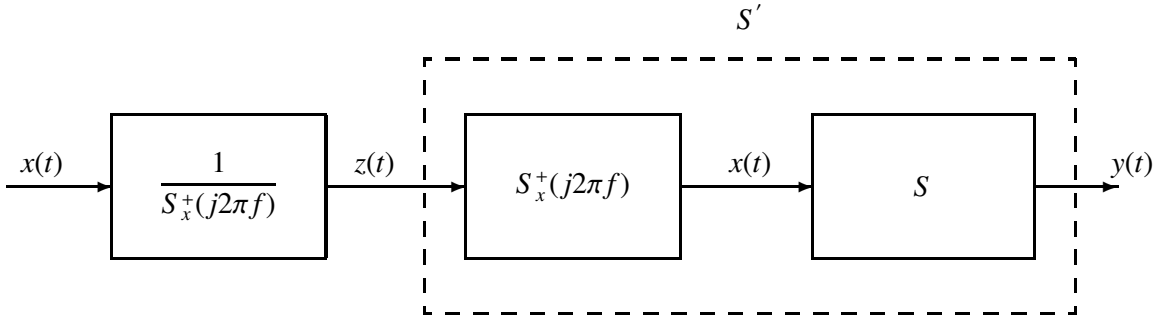

Fig. 1. The basic decomposition proposed by Schetzen for Gaussian non-white input signal.

This system S' can be therefore identified with the procedure already developed for white Gaussian input signals, and then the impact of the coloring filter inside S' must be removed from the overall identified system in order to identify the original system S.

As mentioned, all the basic components of such an extension are already present in his 1962 report so that the method is fully developed in a period of about twenty years. At the beginning the main task was the inversion of the Volterra system and the extraction of the $2N$th-order white Gaussian components from a general signal, an effort that culminated in his 1976 contribution [53]. In fact, in [52], [54] we find the formula for $2N$th-order moments of a zero-mean Gaussian process in terms of a sum (over all pairs of variables) of products of the autocorrelation function; we also find the sentences "*if $h_1(t)$ is not a minimum-phase network, then an inverse to $h_1(t)$ does not exist, but can be approximated arbitrarily closely, although only at the expense of increasingly large time delay*" and "*any Gaussian process can be considered to be the white Gaussian noise response of a linear filter*". These points were much advanced for the year 1961; for example, the last sentence, which is not proved by Schetzen, still merits a theorem in Yasui's paper [51] about twenty

years later (see Theorem 6 in [51]: "The response of $\mathbf{S}$ to a stationary Gaussian input is Gaussian if and *only if* $\mathbf{S}$ is linear").

The second important contribution came in the middle of this two-decade period (Sixties and Seventies) in which Schetzen developed his method. In fact, Brillinger [55] showed that

$$H_N(f_1,\ldots,f_n) = \frac{P_{x\ldots xy}(-f_1,\ldots,-f_n)_{N+1}}{N!\,P_x(f_1)\ldots P_x(f_N)} \tag{15}$$

where $P_x(f)$ is the power spectral density of the zero-mean, Gaussian and stationary input[6] signal $x(t)$, and $P_{x\ldots xy}(\cdot,\ldots,\cdot)_{N+1}$ denotes the $(N+1)$th order joint input-output polyspectra. Brillinger and Rosenblatt are known for their introduction of higher-order cumulants, which is strictly linked to their reading of the contributions introduced by Shiryaev [56] in higher-order statistics. The relation reported here extends the famous Wiener input-output relation for linear systems (i.e., $S_{xy}(f) = S_x(f)H(-f)$) to the polyspectra and uses it to identify the highest-order Volterra kernel.

In the literature, the contribution of Brillinger is sometimes erroneously reported, e.g., the description of Brillinger's contribution given in [57] only refers of the identification of a homogeneous Volterra system while it is clearly stated in section 10 of [55] that the identification formula holds for the highest-order of a general Volterra system; moreover, in [57] it is also first noted that, analogously to the relation existing between (9) and (7), the

[6] To be precise, the mathematical details are slightly different. Brillinger follows the Wiener statistic notation and assumes that the input signal admits a Cramer representation, i.e., it can be written as $x(t) = \int e^{i\lambda t} dZ_x(\lambda)$ where $Z_x(\lambda)$ is a process with orthogonal increments.

general identification formula holds also for order $N-1$ when the Gaussian input has a zero-mean value. However, such a result had already been exploited in [58] to provide a full identification formula for the case of quadratic systems ($N=2$), in the first contribution along the approach that we are describing, about a decade before the Brillinger contribution.

The Brillinger identification formula is rediscovered by the Billings' group in [59]; the author, Fakhouri, does not mention Brillinger or other authors of higher-order statistics and seems not aware of having used the concept of cumulant, though he provides the exact expression of the cross-cumulant $C_{x\ldots xy}$ in terms of the moments and relates it to the multiple convolution in the time-domain. Moreover, he does pass to the frequency domain to reduce the computational burden and carries out the kernel identification on the basis of an ARMA (autoregressive moving average) parametric expression of the Nth-order Volterra transfer functions whose coefficients are determined as the solution to a linear system of equations, which can viewed as a method for parametric polyspectra estimation; such a method is not only advanced for that time and surely competitive with the higher-order periodogram defined by Brillinger in [60] and capable to quieting the "note of pessimism" with which Brillinger concludes his seminal paper [61], but merits being compared with the most advanced methods for polyspectra estimation reported in the second chapter of [62] and their further developments. Moreover, unlike Brillinger, the author of [59] also specializes the procedure to the case for which the input signal is white, thereby expressing with a different notation a method for solving the problem of the diagonal kernel values; in other

works, though Fakhouri gives no mention to the concept of cumulant, he provides a formula that, when read with knowledge of higher-order statistics, states that the long-searched-for solution to the problem of estimating the diagonal terms of the Volterra kernels for a white Gaussian input is given by the input-output cross-cumulant, as can be easily checked by expressing the counterpart of (15) in the time-domain for a white input signal.

The fact that, about ten years later than the Brillinger's paper, the Billings' group did not see that they were rewriting with a different notation his method (defining a cumulant without recognizing it) illustrates well the long-needed advances in Volterra system identification. The same group, however, is aware of the need for a synthesizing review of the methods for Volterra system identification and produced one of the first review paper [63] and Billings' book [64] still constitutes an important reference source for Volterra system identification.

It is surprising that the availability of (15) has not given an idea for solving the problem of the diagonal values in (12) by specializing it to the case of strict-sense white input and noticing that the cross-cumulant was the problem solution at least for the case of white Gaussian signals. Still the Billings' book [64] attributes to [57] the introduction of cumulants for estimating the Volterra transfer functions for Gaussian inputs, probably following the statement in [57] (according to which Brillinger limited himself to homogeneous Volterra systems). The neglectfulness of Brillinger's contribution is similar to that affecting other important utilizations of higher-order statistical signal processing to Volterra system identification discussed in the sequel. Furthermore, the following content of Billings' book

[64] in the passage in which that the author of one of the most effective methods for polyspectrum estimation (unaware of the possible interpretation of his contribution in those terms) declines the method (attributed to [57] rather than to [55]) on the basis of the difficulty to carry out polyspectra estimation. A careful reading of the entire paragraph, however, could be interpreted to be a disapproval of a nonparametric approach to polyspectra estimation in favor of a parametric one for any realistic case and not only for cases with data of limited sample size.

Rather than including the details regarding the correlation-based identification formulas in the case of colored Gaussian input signals, we include a revealing detail that is particularly clear from Schetzen's approach. The case of colored input signal has interest mainly when the signal is not under the control of the experimentalist; in fact, when the experimentalist has control of the input signal, it is difficult to imagine choosing a colored Gaussian signal instead of a white signal. In Schetzen's approach this is clear because his approach is to solve for the filter that produces the actual colored noise from white noise and then use knowledge of the inverse of this coloring filter to remove its effect from the identified modified system that includes the coloring filter.

However, the problem is common in both approaches since often an important motivation for considering the colored signal is missing; moreover, experiments showing the advantages of the proposed method in comparison with the existing ones for the case of white input signals are missing. The contribution in [59] is an exception because it mentions a need

to simplify the procedure associated with the choice of colored signal and carry out a simulation experiment that shows the efficiency of the method but no comparison with an alternative method is considered nor is a general discussion about the advantage for which an experimentalist should prefer the colored signal present. However, apart from the general principle that imposes an identification formula suited to the existing signal for the case where the input is not under the control of the experimentalist, we should not forget that the long-lasting difficulties in solving the problem of the diagonal kernel values had made the case of colored signal more attractive, since, at least for such a case, exact identification formulas (by Brillinger and Fakhouri, with the discussed differences between them) for the highest-order kernels existed.

The important contribution in [59] consisting of making clear that, when a multidimensional ARMA model is reasonable for the higher-order Volterra kernel, then a great computational advance can be gained by assuming it in the identification procedure in comparison with an estimation procedure that assumes simply a non-recursive expression for the kernel coefficients, which significantly increases the number of parameters to be estimated and the variances of their estimators. The simulation experiments carried out in [59] are executed with kernels that satisfy an ARMA model and their parameters are essentially perfectly estimated for a third order model. Also on the basis of the problem of estimating the diagonal values of the Volterra kernels in correlation-based methods for white input signals, this is one of the first contributions in the literature associated with an

experimental section describing a fully working estimation method! Such a sentence should be seen in connection with the quality of the method for polyspectrum estimation implicitly derived (together with the same concept of cumulant) in [59].

The contribution by Brillinger is much more profound than those reviewed here with reference to the identification formula (15), and is discussed in the sequel.

C. The general input signal

Following the general reasoning at the end of the previous subsection and in some sense confirming it, the non-existence of identification formulas for the general case of non-white non-Gaussian input has often limited the approach considered in such a general case. More specifically, the limitation consists in the exclusive reference to a scenario where the input signal is not under the control of the experimentalist. Not only has an identification formula not been proposed for this case but also no one seems to have argued that the ability to treat this case would be an important step ahead for system identification.

The motivation for the research however includes the important case for which the input signal is present and can only be measured by the experimentalist. All the research in this line has used the important property of the Volterra representation of being linear in the parameters so that the problem of their estimation can be carried out as a linear estimation problem.

The approaches can be roughly classified into two cases:

- the first approach defines a cost function in terms of the time average of the squared

error signal; the time average that defines the cost function is theoretically conceived over an infinite time interval and in practice is performed over finite time intervals determined by the available data. The method in this approach tries to minimize the cost functions and expresses the values of the Volterra kernels at various orders in terms of the averages of different kinds of processed data, mainly input and joint input/output higher-order moments of the input and output signals. In general, this approach leads to linear systems of equations whose solutions provide the kernel values at various orders.

- the second approach does not directly define a cost function in terms of time averages but its definition is such that a cost function based on time averages could well approximate it. The problem is generally defined in mathematical terms; a cost function is defined and the optimum value of the kernel values is determined by a suitable procedure. Different procedures using the linearity of the Volterra representation with respect to its free parameters (the kernels) can be used. Here the approach to Volterra system estimation is not much different from any approach based on generalized linear models and admits the same method for solutions. In particular, the method for minimizing the cost function can be based on the gradient technique or the stochastic gradient technique (LMS Volterra method) or the least squares technique or a recursive version (RLS Volterra method). Moreover, a method for selecting a subset of significant coefficients (nonnull values of the Volterra kernels) in the generalized linear model on the basis of the data can be selected according to a general procedure for sparse solution of a linear

system.

IV. Advances derived from using fraction-of-time approach to the identification of time-invariant Volterra systems

The contributions of Archer and Gardner in [68], [69] represent important advances to the problem of time-invariant Volterra system identification. The first contribution lies in a systematic return to Wiener's idea of identification by means of input-output higher-order cross-correlation. On this point the authors introduce a powerful extension of the Klein-Yasui approach [51] for avoiding the problem (never really solved by the M.I.T. group) of the diagonal kernel values with white stationary input signal. This extension is mainly due to the use of the non-probabilistic framework that puts on a solid ground an approach that had become weak. Norbert Wiener had also provided a solid basis to his algorithms but at the expense of dedicating much effort to relating statistical means and time averages by means of the Birkhoff ergodic theorem. However, in the following decades, the ergodic theorem became the ergodic hypothesis (assumption) and such a firm correspondence between the two averages was missing from the work on many algorithms proposed in the literature. By constructing a theory around the fraction-of-time natural density [15], [16] inherent to any measured time series, i.e., the density according to which the time averages are performed, research is freed from the problem of establishing which is the density according to which orthogonalization procedure has to be carried out. Also the approach followed by Klein and Yasui [51] rests on the ergodic theorem to construct the orthogonalization functionals: according to this approach, one should first solve the fundamental problem of statistics

to find the density of the stochastic process modeling the actual data before constructing the functionals whose orthogonalization is based on this density. Using the fraction-of-time natural density greatly simplifies the procedure for finding an appropriate orthogonalization since the construction is based on the use of the time-average of the product of two time series as inner product on which the Gram-Schmidt procedure of orthonormalization can be based, at least for the case of temporally independent input sequences, which is the case considered in [68].

The best previous contribution on this issue is given by Barrett; in fact, in [70] it is clearly stated how to avoid the limit of the diagonal elements by using the Hermite functionals. However, such a solution rests on the assumption of Gaussian statistics. Klein and Yasui had extended such approach to the case of a general statistics of the white input signal but continue to assume that this statistics is known. The approach proposed in [68], [69], instead, is defined in terms of temporal averages of the input and output signal, and is independent of the statistics of the input signal. This advantage is important when the stationary input signal is not under control of the experimentalist but can be well measured.

If we read what is written about Yasui's contribution in [71], the great importance of the contribution in [69], [68] becomes evident since in [71] the fact that [72], [51] do not contain identification methods is clearly surfaced. In particular, the first Yasui's contribution [72] is not mentioned in the list of methods for Volterra systems but it is mentioned with reference to the Wiener systems identification (it is recognized that it eliminates the problem of the

diagonal kernels but it is stated about it that the “theoretical presentation is mainly focused upon a clarification and unification of the Wiener, Volterra, and several other methods based on the series approach to a nonlinear system”) while the more important contribution [51], which is used in [68] to construct the orthogonalization for the case of stationary input signal, is simply not mentioned.

The contribution in [68], [69] constitutes an important advance since, thanks to the use of the fraction-of-time approach, an immediate identification procedure for the case where stationary input sequences are employed is much more directly available than it was before (i.e., the procedure is now defined in terms of temporal averages of signals depending on the input and output ones). For example, a relatively recent contribution [47] in this line of researches, being unaware of [69], [68], has proposed to use the interpolation of the values obtained by (7) as an alternative to other (already discussed and not efficient) methods to estimate the kernels on the diagonal points.

V. Applying the theory of cyclostationarity to the problem of Volterra system identification

A general method for evaluating the Volterra kernels by cross-correlation between the output signal and a function of the input signal is proposed in [69], [68] by adopting a non-probabilistic approach, i.e., by modeling the system output as a time series rather than a realization of a random process. It is assumed that the input signal is cyclostationary rather than stationary; with reference to discrete time, two cases are considered: (1) the case in

which the cyclostationary input signal has the form

$$x(k) = z(k)e^{j\omega k} \tag{16}$$

where $z(k)$ is a stationary time-series, possibly complex-valued, and (2) the case where the input signal is given by

$$x(k) = \Re\{z(k)e^{j\omega k}\} \tag{17}$$

where $\Re\{\cdot\}$ denotes the real-part extraction operation.

In the first case the input signal is complex-valued and therefore can be applied to the system only if a mathematical description of the system[7] is available.

In the case of complex-valued signal (16) the general system

$$y(k) = \sum_{n=1}^{\infty} \sum_{\boldsymbol{j}_n} h_n(\boldsymbol{j}_n)\lambda_n(k, \boldsymbol{j}_n, x(\cdot)\ ) \tag{18}$$

is considered where

$$\lambda_n(k, \boldsymbol{j}_n, x(\cdot)\ ) = \prod_{r=1}^{n} x(k - j_{n_r}) \tag{19}$$

and $\boldsymbol{j}_n = \{j_{n_1}, \ldots, j_{n_n}\ \}$. It is shown that, provided that $\dfrac{\omega}{2\pi}$ is irrational:

$$h_n(\boldsymbol{k}_n) = \frac{1}{P(\boldsymbol{k}_n)} \left\langle \quad y(k) \cdot \gamma_n^*(k, \boldsymbol{k}_n, z(\cdot)) \right\rangle \tag{20}$$

[7] Already in [73] there was proposed a method for identifying the Volterra transfer functions by employing a complex-valued input signal $x(t) = e^{j\omega_1 t} + \ldots e^{j\omega_n t}$, with the values ω_i irrational among themselves; the component with frequency $\omega_1 + \ldots + \omega_n$ is multiplied by the coefficient $n!\ H_n(\omega_1, \ldots, \omega_n)$; in [73] there is an attempt to use the equations that implicitly define the system to explicitly express the input-output relation with a Volterra series.

where $< \ldots >$ is the temporal average, $\boldsymbol{k}_n = \{\underbrace{k_{m_1}, \ldots, k_{m_1}}_{n_1}, \ldots, \underbrace{k_{m_p}, \ldots, k_{m_p}}_{n_p}\}$ in which km_q is repeated nq times (and, obviously, $n_1 + \ldots + n_p = n$) and

$$P(\boldsymbol{k}_n) = \frac{n!}{n_1! \ \ldots n_p!} \tag{21}$$

$$\gamma_n(k, \boldsymbol{k}_n, z(\cdot)) = e^{j\omega nk} \ e^{-j\omega \sum_{q=1}^{p} n_q k_{m_q}} \ \phi_n(k, \boldsymbol{k}_n, z(\cdot)) \tag{22}$$

and ϕ_n is such that

$$\left\langle \quad \lambda_n(k, \boldsymbol{j}_n, z(\cdot)\ ) \cdot \phi_n^*(k, \boldsymbol{k}_n, z(\cdot)) \right\rangle \ = \delta_{\tilde{\boldsymbol{j}}_n \boldsymbol{k}_n} \tag{23}$$

in which $\tilde{\boldsymbol{j}}_n$ is a general permutation of $\boldsymbol{j}_n$.

In the case of real system (17), an identification formula for the highest-order kernel is proposed, which solves the problem of estimating the diagonal values of the highest-order kernel

$$y(k) = \sum_{n=1}^{N} \sum_{\boldsymbol{j}_n} h_n(\boldsymbol{j}_n) \lambda_n(k, \boldsymbol{j}_n, x(\cdot)\ ) \tag{24}$$

and it is also shown that, in such a case, the relation (20) still holds for any ω provided that $n = N$. It is therefore possible to identify the N-order Volterra kernel and to subtract the N-order output to the overall system output and so to identify the lower-order Volterra kernels.

For complete specification of the algorithm, it is necessary to be able to calculate the time-series $\phi_n(k, \boldsymbol{k}_n, z(\cdot))$ that satisfy the condition in (23). It is shown how to determine the functions ϕ_n in the case where $z(k)$ is a sequence of random variables temporally independent

and the powers $z^m(k)$ are linearly independent among themselves for the various values of m. In such a case,

$$\phi_n(k, \boldsymbol{k}_n, z(\cdot)) = \prod_{q=1}^{p} \psi_{n_q}(z(k - k_{m_q})) \tag{25}$$

where the functions $\psi_n(\cdot)$ are such that

$$\left\langle \; z^m(k - j_0) \cdot \psi_n(z(k - k_0)) \right\rangle \; = \delta_{mn}\delta_{j_0 k_0} \qquad m \leq n \tag{26}$$

The $\psi_n(\cdot)$ can be determined starting from $\{z^m(k)\}_{m=0,\ldots,n}$ by means of the Gram-Schmidt procedure: $\psi_n(z(k))$ is the (properly normalized) component of $z^n(k)$, orthogonal to the subspace identified by the vectors $\{z^m(k)\}_{m=0,\ldots,n-1}$. With this choice the sequences ϕ_n satisfy a condition stronger than (23):

$$\left\langle \; \lambda_m(k, \boldsymbol{j}_m, z(\cdot)\,) \cdot \phi_n^*(k, \boldsymbol{k}_n, z(\cdot)) \; \right\rangle \; = \delta_{mn}\delta_{\boldsymbol{j}_n \boldsymbol{k}_n} \qquad ; \;\; m \leq n \tag{27}$$

on which the validity of the algorithm for every value of the variable ω is based; the validity of the algorithm for the case $\omega = 0$ allows to use the considered method also for stationary input sequences.

The results are also expressed in the frequency domain to be able to utilize the computational advantages of the FFT:

$$H_n(\boldsymbol{f}_n) \cong \left[\frac{1}{T} w_T^*(-\boldsymbol{f}_n) Y_T(\mathbf{1}_n^T \boldsymbol{f}_n)\right] \otimes A_K(f_1) \otimes \ldots \otimes A_K(f_n) \tag{28}$$

in which

$$A_K(f) = \frac{\sin[\pi f(K+1)]}{\sin(\pi f)} \tag{29}$$

$$Y_T(f) = \sum_{|k| \leq \frac{T}{2}} y(k) e^{-j2\pi fk} \tag{30}$$

$$w_T(\boldsymbol{f}_n) = \sum_{|k| \leq \frac{T}{2}} w(\boldsymbol{k}_n) e^{-j2\pi \boldsymbol{f}_n^T \boldsymbol{k}_n} \tag{31}$$

$$w(\boldsymbol{k}_n) = \frac{1}{P(\boldsymbol{k}_n)} \gamma_n(0, \boldsymbol{k}_n, z(\cdot)) \tag{32}$$

and K must be larger than the system memory L.

The approximation error in (28) tends to zero when $T \to \infty$. The computational efficiency increases significantly if a rectangular window of length $\frac{1}{K}$ is used to replace the function $A_K(f)$.

The proposed method has been verified in both the case in which the mathematical description is available and therefore the complex-valued input sequences can be used and the case in which the use of the real-valued input sequence is mandatory. The input sequence is the PM sequence based on the definition

$$z(k) = \sigma \; e^{j\theta(k)} \tag{33}$$

with $\theta(k)$ a sequence of IID variables that assume with the same probability M discrete values in $[-\pi, \pi]$. This test case is further extended to the case in which $z(k)$ is a purely stationary fraction-of-time-probability independent Gaussian sequence with known standard deviation.

The capability to solve the problem of the diagonal kernel values is elegant and arrives at perfection. It is likely that this also corresponds to reduced variance of the estimator

but unfortunately this issue has not been verified by simulation experiments reported in the literature.

Therefore, we can conclude that the contribution in [68] represents an advance in comparison with the methods operating with stationary inputs also if the method proposed in it is used in the setting $\omega = 0$ (stationary input); moreover, the possibility to choose a value $\omega \neq 0$ enables direct identification of any value of the Volterra kernel at any order provided that a mathematical description of the system is available. Moreover, also when the mathematical description is not available, it opens the possibility of extending all the identification methods already proposed in the literature to the case of cyclostationary input sequences. Such an extension is likely associated with performance improvements of the identification methods; however, no successive literature contribution has further investigated this issue. This may be due to the assumption that past failures of efforts to solve long-standing problems with earlier methods indicated there was no promise for success.

A. Input signal affected by measurement noise

The method developed in [68] considers both the case where the input signal is under control of the experimentalist and the case where it is only measured by the experimentalist, though the first case is given much more attention. In [74], [75] the second case is given much more attention by assuming that both input and output signals have to be measured by the experimentalist and may be affected by measurement noise, a case usually neglected in the literature. The presence of an additive, zero-mean noise term, independent of the input

signal, has been considered only for the output signal, which is still a necessary assumption even when the input signal is under control of the experimentalist. Moreover, the majority of earlier methods proposed in the literature are not able to correctly operate when the input signal is affected by noise.

A deep analysis of the existing methods could show that only the method of Gardner and Archer in [68] is suited for adaptation to a noisy-input scenario: for an input signal of the form $z(k)\cos(2\pi\nu_0 k)$, with $z(k)$ a sequence of temporally independent random variables with identical distributions, and such that the powers $\{z^m(k)\}$ are linearly independent of each other, the maximum order Volterra kernel can be identified. However, the method proposed in [68] assumes that $z(k)$ is available to the experimentalist, but, when only noisy measurements of the system input and output are available, it is not possible to assume such knowledge. Two cases in which this problem is addressed are proposed in [74]. Before considering the relative details, the needed signal notation is introduced: signal $v(k) = x(k) + \eta(k)$ denotes the noisy measurement of the input signal $x(k)$ and the signal $w(k) = y(k) + \zeta(k)$ denotes the noisy measurement of the output signal $y(k)$; therefore, we denote with $\eta(k)$ and $\zeta(k)$ the disturbance signals affecting the measurements of the input signal $x(k)$ and output signal $y(k)$, respectively.

It is also noted that in [68] the input signal is perfectly known while in the two cases described in the next two subsections only a noisy version $v(k)$ of the useful input signal $x(k)$ is available. However, it is assumed that the useful signal $x(k)$ is a general IID sequence

modulated by an AM sequence, whose frequency parameter ν_0 is assumed known. The method in [68] cannot be straightforwardly applied in this case because it assumes that the signal $z(k)$ is available while in the considered scenario only the noisy version $v(k)$ of the input signal $x(k)$ is available. The first case considered in [74] and described in Subsection V-B, is an extension of the method already proposed in [68]; the extension lies in the fact that its application only resorts to a noisy version of the input signal $x(k)$ and does not require knowledge of the intermediate signal $z(k)$, which is needed in the method described in [68]. Moreover, the second case, considered in [74] and described in Subsection V-C, includes at same time the extension from the stationary to the cyclostationary input of the Brillinger's method relative to the case of colored Gaussian input and the extension of this method to the case in which the input signal measurements are affected by noise.

B. First case: IID modulating sequence

To extend the method in [68], which assumes an AM cyclostationary signal, to the case where only the signal $v(k)$ is available, it is necessary to modify the orthogonalization method proposed in [68], which uses the signal $z(k)$, by resorting to the sinusoidally-weighted time-averages of signals depending only on the noisy version of the input signal $x(k)$. In particular, it is necessary to resort to the sinusoidally-weighted time-average $\langle x(k) \rangle_{\nu_0}$ of a discrete-time signal $x(k)$) as follows:

$$\langle x(k) \rangle_{\nu_0} = \left\langle x(k) e^{-j2\pi\nu_0 k} \right\rangle \tag{34}$$

and, using it, to define the following functions:

$$\psi_0[x(k)] \stackrel{\triangle}{=} 1 \tag{35}$$

$$\psi_1[x(k)] \stackrel{\triangle}{=} x(k) - 2\left\langle x(k)\right\rangle_{\nu_0}\cos(2\pi\nu_0 k) \tag{36}$$

$$\begin{aligned}\psi_2[x(k)] \stackrel{\triangle}{=}\ & x^2(k) - 2\left\langle x^2(k)\right\rangle_{2\nu_0}\cos[2\pi(2\nu_0)k] \\ -\ & 2\frac{\langle\, x^2(k)\psi_1[x(k)]\,\rangle_{3\nu_0}}{\langle\, \psi_1[x(k)]\psi_1[x(k)]\,\rangle_{2\nu_0}}\psi_1[x(k)]\cos[2\pi\nu_0 k]\end{aligned} \tag{37}$$

and, for general n,

$$\psi_n[x(k)] \stackrel{\triangle}{=} x^n(k) - \sum_{i=0}^{n-1} 2\frac{\langle\, x^n(k)\cdot\psi_i[x(k)]\,\rangle_{(n+i)\nu_0}}{\langle\, \psi_i[x(k)]\cdot\psi_i[x(k)]\,\rangle_{2i\nu_0}}\psi_i[x(k)]\cdot\cos[2\pi(n-i)\nu_0 k] \tag{38}$$

Then it is shown in [74] that

$$\left\langle \psi_n[x(k)]\cdot\psi_i[x(k)]\right\rangle_{(n+i)\nu_0} = 0 \quad \forall i<n,\ n\le N,\ 4N\nu_o<1 \tag{39}$$

The assumption $4N\nu_o < 1$ can be replaced also by the condition that ν_0 is irrational. From (39) it can be shown that the following relation holds:

$$\left\langle \psi_n[x(k)]\cdot x^i(k)\right\rangle_{(n+i)\nu_0} = 0 \quad \forall i<n,\ n\le N,\ 4N\nu_o<1 \tag{40}$$

and, therefore, defined

$$\mathbf{k_N} = \{\underbrace{k_{m_1},\ldots,k_{m_1}}_{n_1},\ldots,\underbrace{k_{m_p},\ldots,k_{m_p}}_{n_p}\} \tag{41}$$

with $n_1 + \ldots + n_p = N$, for a finite-order system, we can write

$$h_N(\mathbf{K_N}) = \frac{\left\langle y(k)\psi_{n_1}[\, x(k-k_{m_1})] \ldots \psi_{n_p}[\, x(k-k_{m_p})] \right\rangle_{2N\nu_0} \cdot \frac{n_1! \ldots n_p!}{N!}}{\left\langle x^{n_1}(k-k_{m_p})\psi_{n_1}[x(k-k_{m_p})] \right\rangle_{2n_1\nu_0} \ldots \left\langle x^{n_p}(k-k_{m_p})\psi_{n_p}[x(k-k_{m_p})] \right\rangle_{2n_p\nu_0}} \tag{42}$$

Such expression can be viewed as a re-writing of the method (20) proposed in [68]. The main contribution in [74] lies in the proof that the identification formula (42) still holds when the input signal $x(k)$ is replaced by its noisy measurement $v(k)$ and the output signal $y(k)$ is replaced by its noisy measurement $w(k)$ provided that the input noise $\eta(k)$ and output noise $\zeta(k)$ are independent of the input signal and do not exhibit cyclostationarity properties analogous to those of the useful signals, specifically described by the following conditions:

$$R_{\eta}^{l\nu_0}(\cdot)_n \equiv 0 \qquad \forall n \leq N \qquad l \in \{1, 2, \ldots, 4N\} \tag{43}$$

and

$$R_{\eta \ldots \eta\zeta}^{l\nu_0}(\cdot)_n \equiv 0 \qquad \forall n \leq N+1 \qquad l \in \{N, N+1, \ldots, 3N\} \tag{44}$$

provided that $l\nu_0$ is not an integer. In (43) $R_{\eta}^{l\nu_0}(\cdot)_n$ denotes the nth-order reduced-dimension cyclic temporal moment function [76] of the disturbance signal $\eta(\cdot)$ at the cycle frequency $l\nu_0$ while in (44) $R_{\eta \ldots \eta\zeta}^{l\nu_0}(\cdot)_n$ denotes the nth-order reduced-dimension cyclic temporal cross-moment function of the disturbance signals $\eta(\cdot)$ and $\zeta(\cdot)$ at the cycle frequency $l\nu_0$.

C. Second case: Gaussian modulating sequence

For the case for which the input signal is $x(k) = z(k)\cos(2\pi\nu_0 k)$, with $z(k)$ Gaussian, zero-mean and stationary, the following identification formula for the highest-order kernel is derived in [74]:

$$H_N(\nu_1 + 2\nu_0, \ldots, \nu_N + 2\nu_0) = \frac{P^{2N\nu_0}_{x\ldots xy}(-\nu_1, \ldots, -\nu_N)_{N+1}}{N!\, P^{2\nu_0}_x(\nu_1 + 2\nu_0)_2 \ldots P^{2\nu_0}_x(\nu_N + 2\nu_0)_2} \tag{45}$$

where $P^{2N\nu_0}_{x\ldots xy}(-\nu_1, \ldots, -\nu_N)_{N+1}$ denotes the $(N+1)$th-order cyclic cross-polyspectrum [76] of the signals $x(k)$ and $y(k)$ and $P^{2\nu_0}_x(\nu_1 + 2\nu_0)_2$ denotes the second-order cyclic polyspectrum of the signal $x(k)$. This formula can be seen as an extension to a general ν_0 of the Brillinger's formula (15) proposed in [55] for the case $\nu_0 = 0$ and continues to be strictly linked to the Shiryaev [56] contribution regarding the cumulants of a Gaussian signal.

The main contribution of [74] is the proof that the identification method is not affected by the presence of measurement noise, i.e., (45) still holds when the input signal $x(k)$ is replaced by its noisy measurement $v(k)$ and the output signal $y(k)$ is replaced by its noisy measurement $w(k)$ provided that the additive input noise $\eta(k)$ and the additive output noise $\zeta(k)$ are independent of the input signal and do not exhibit cyclostationarity properties analogous to those of the useful signals, specifically described by the following conditions:

$$C^{2n\nu_0}_{\eta\ldots\eta\zeta}(\cdot)_{n+1} \equiv C^{2\nu_0}_{\eta}(\cdot)_2 \equiv 0 \quad \forall n \leq N \tag{46}$$

In (46) $C^{2n\nu_0}_{\eta\ldots\eta\zeta}(\cdot)_{n+1}$ denotes the $(n+1)$th-order reduced-dimension cyclic temporal cross-cumulant function of the signals $x(k)$ and $y(k)$ at the cycle frequency $2n\nu_0$ and $\overline{C}^{2\nu_0}_{\eta}(\cdot)_2$

denotes the second-order cyclic covariance (also, named, second-order reduced-dimension cyclic temporal cumulant function) of the signal $\eta(k)$ at the cycle frequency $2\nu_0$.

Moreover, in [74] it is also shown for both methods how to derive the lower-order kernels of the considered system in the presence of noise affecting the measurements of the input signal. Furthermore, it is noted that both methods are robust with respect to an overestimation of the maximum order N of the Volterra system, an issue rarely considered in the literature.

Finally, let us note that the methods in [68] have been extended to the case of polyperiodic Volterra systems in [77]; the polyperiodic Volterra systems are Volterra systems where the kernels are not constant but are polyperiodic functions of time and therefore can be specified by a set of cyclic Volterra kernels (Fourier coefficients of the Volterra kernels). Moreover, the methods in [74], suited to the case where the input/output system measurements are affected by additive noise, have been extended to the case of polyperiodic Volterra systems in [78].

D. Further contributions using the processing of cyclostationary signals

A method regarding the blind estimation of a nonlinear system that admits a homogeneous quadratic Volterra description is introduced in [79], [80]. The contribution refers to the case where the purely quadratic nonlinear system describes a communication channel and utilizes the cyclic properties of typical signals at the output of communication systems. The proposed methods use the cyclic autocorrelation or cyclic cepstrum of the output signal and assume the presence of simple system input (that is typical of those used in communication systems)

that induces the typical property of the channel output.

Moreover, very recently a new contribution on the use of a cyclostationary input signal to identify a Volterra system model has been proposed in [81]. It assumes that the nonlinear system is an industrial device that admits an LN block description (as specified in Subsection II-C) and that the input signal is under the control of the experimentalist. At the system input, the authors choose to employ a cyclostationary sequence whose fraction-of-time mean and variance are periodic functions of the discrete time. This kind of cyclostationary signal is well suited to the scenario where the experimentalist has full choice of the possible input signal.

VI. Input-output relations

To complement the material above, which is specific to nonlinear system identification, this section reviews some useful mathematical expressions for input/output relations given two Volterra systems S_1 and S_2 for which $y(t)$ and $z(t)$, respectively, denote the output signals corresponding to the same Gaussian input. The mean, crosscorrelation $E\{y(t+\tau)z(t)\}$, and power spectral density of $y(t)$ are evaluated in [73] and are given by:

$$\langle y(t)\rangle = \sum_{n=1}^{\infty} \frac{1}{(2n)!\ n!\ 2^n} \int_{\mathcal{R}^n} S_x(f_1)\dots S_x(f_n) H_{2n}(f_1,-f_1,\dots,f_n,-f_n) df_1\dots df_n \quad (47)$$

$$\langle y(t+\tau)z(t)\rangle =$$

$$= a_0 b_0 + \sum_{k=1}^{\infty} \int_{\mathcal{R}^k} \frac{e^{i2\pi\tau(f_1+\dots+f_k)}}{k!} S_x(f_1)\dots S_x(f_k) a_k(f_1,\dots,f_k) b_k(f_1,\dots,f_k) df_1\dots df_k \quad (48)$$

$$S_y(f) = \mid a_0 \mid^2 \delta(f) + \sum_{k=1}^{\infty} \int_{\mathcal{R}^k} \frac{S_x(f_1)\dots S_x(f_k)}{k!} \mid a_k(f_1,\dots,f_k) \mid^2 \ \delta(f-f_1-\dots-f_k) df_1\dots df_k \quad (49)$$

where $a_0 = \langle y(t) \rangle$, $b_0 = \langle z(t) \rangle$ and for $k > 0$:

$$a_k(f_1, \ldots, f_k) = \sum_{n=0}^{\infty} \int_{\mathcal{R}^n} \frac{S_x(f_1^{'}) \ldots S_x(f_n^{'})}{(2n+k)!\ n!\ 2^n} H_{2n+k}(f_1, f_2, \ldots, f_k, f_1^{'}, -f_1^{'}, \ldots, f_n^{'}, -f_n^{'}) df_1^{'} \ldots df_n^{'} \tag{50}$$

and b_k is obtained from (50) by replacing the transfer function of S_2 evaluated in the points $-f_1, \ldots, -f_k$ with the transfer function of S_1 $H_k(f_1, \ldots, f_k)$. The relation (49) can be extended to the case where a deterministic sine signal is added to the Gaussian signal.

The first contribution of this type is given in [22] where the output autocorrelation is evaluated independently of the assumption of Gaussian white input.

In [82] the assumption of Gaussian input is removed. The joint input-output correlation and the output autocorrelation (and their transforms) are expressed in terms of the higher-order moments of the input signal (and of their transforms):

$$R_{YX}(\tau) = \langle y(t+\tau) x^*(t) \rangle = \sum_{n=1}^{\infty} \int_{\mathcal{R}^n} h_n(u_1, \ldots, u_n) R_{XX(n,1)}(\tau - u_1, \ldots, \tau - u_n) du_1 \ldots du_n \tag{51}$$

$$\begin{aligned} R_{YY}(\tau) = \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} \int_{\mathcal{R}^{n+m}} h_n(u_1, \ldots, u_n) h_m^*(u_{n+1}, \ldots, u_{n+m}) \cdot \\ \cdot R_{XX(n,m)}(\tau - u_1 + u_{n+m}, \ldots, \tau - u_n + u_{n+m}, u_{n+m} - u_{n+1}, \ldots, u_{n+m} - u_{n+m-1}) \\ du_1 \ldots du_{n+m} \end{aligned} \tag{52}$$

$$S_{YX}(f) = \sum_{n=1}^{\infty} \int_{\mathcal{R}^{n+1}} H_n(f_1, \ldots, f_n) \hat{S}_{XX(n,1)}(f_1, \ldots, f_{n+1}) \delta(f - f_1 \ldots - f_n) df_1 \ldots df_{n+1} \tag{53}$$

$$S_{YY}(f) = \sum_{n=1}^{\infty}\sum_{m=1}^{\infty}\int_{\mathcal{R}^{n+m}} H_n(f_1,\ldots,f_n)H_m^*(-f_{n+1},\ldots,-f_{n+m-1},f_1+\ldots f_{n+m-1})\cdot$$
$$\cdot\hat{S}_{XX(n,m)}(f_1,\ldots,f_{n+m})\delta(f-f_1\ldots-f_n)df_1\ldots df_{n+m} \quad (54)$$

where

$$R_{XX(n,m)}(\tau_1,\ldots,\tau_{n+m-1}) = \langle x(t+\tau_1)\ldots x(t+\tau_n)x^*(t+\tau_{n+1})\ldots x^*(t+\tau_{n+m-1})x^*(t)\rangle \quad (55)$$

and

$$\hat{S}_{XX(n,m)}(f_1,\ldots,f_{n+m}) = \mathcal{F}\left[R_{XX(n,m)}(\tau_1,\ldots,\tau_{n+m-1})\right]\delta(f_1+\ldots+f_{n+m}) \quad (56)$$

In [83] the same authors have derived expressions much similar to those reported in [73] by specializing the results obtained in [82] to the case of Gaussian input.

The mean and the autocorrelation function of the output signal of a discrete-time nonlinear Volterra system whose input consists in a sequence of independent random variables:

$$\langle y(n)\rangle = \sum_{k=1}^{\infty}\sum_{p=1}^{k}\sum_{\boldsymbol{\nu}\in\mathcal{P}(k,p)}\Lambda(\boldsymbol{\nu})\sum_{\boldsymbol{v}\in Z^p}S(\boldsymbol{\nu},\boldsymbol{v})h_k \quad (57)$$

$$R_{yy}(m) = \sum_{k=1}^{\infty}\sum_{l=1}^{\infty}\sum_{p=1}^{k+l}\sum_{\boldsymbol{\nu}\in\mathcal{P}(k,p)}\Lambda(\boldsymbol{\nu})\sum_{\boldsymbol{v}\in Z^p}S(\boldsymbol{\nu},\boldsymbol{v})h_{kl}^m \quad (58)$$

where $\mathcal{P}(k,p)$ denotes the set of partitions $\boldsymbol{\nu} = \{\nu_1,\ldots,\nu_p\}$ of a positive integer k in p parts, i.e.,

$$1 \le \nu_1 \le \ldots \le \nu_p \quad \nu_1+\nu_2+\ldots+\nu_p = k \quad (59)$$

and, moreover,

$$\Lambda(\boldsymbol{\nu}) = \prod_{i=1}^{p} C_{\nu_i} \quad (60)$$

in which C_i is the cumulant of order i of $x(n)$ and

$$S(\boldsymbol{\nu}, \boldsymbol{v})h_k = \sum_{\pi(\boldsymbol{\nu})} h_k \left(\sum_{i=1}^{p} v_i I_i(1),\ \ldots\ , \sum_{i=1}^{p} v_i I_i(k) \right) \tag{61}$$

where $\pi(\boldsymbol{\nu})$ is the set of partitions of the set $\{1, 2, \ldots, k\}$ in p subsets such that each subset $\mathcal{K}_i$ contains ν_i elements and $I_i(j)$ is the function that denotes the belonging of j to the set $\mathcal{K}_i$ and

$$h_{kl}^{m}(u_1, \ldots, u_{k+l}) = h_k(u_1, \ldots, u_k) h_l(u_{k+1} + m, \ldots, u_{k+l} + m) \tag{62}$$

In [84] for a nonlinear Volterra system with multiple inputs and outputs, almost periodically time-variant, the general input-output relations in terms of the higher-order cyclostationary statistics that, when specialized to the case of N nonlinear time-invariant systems with a single input and output, become:

$$\overline{S}^{\alpha}_{y_1,\ldots,y_N}(\boldsymbol{f'})_N = \sum_{n_1,\ldots,n_N\,=\,1}^{+\infty} \int_{\mathcal{R}^{n_1+\ldots+n_N-N}} \overline{S}^{\alpha}_{\boldsymbol{x}^{(n_1)},\ldots,\boldsymbol{x}^{(n_N)}} \left([\boldsymbol{\lambda}^{(n_1)'\mathbf{T}},\right.$$

$$\left. f_1 - \boldsymbol{\lambda}^{(n_1)'\mathbf{T}}\mathbf{1}, \ldots, \boldsymbol{\lambda}^{(n_N)'\mathbf{T}}]^{\mathbf{T}} \right)_{n_1+\ldots+n_N} \cdot H^{(n_N)} \left([\boldsymbol{\lambda}^{(n_N)'\mathbf{T}}, \alpha - \boldsymbol{f'}^{\mathbf{T}}\mathbf{1} - \boldsymbol{\lambda}^{(n_N)'\mathbf{T}}\mathbf{1}]^{\mathbf{T}} \right)$$

$$\prod_{m\,=\,1}^{N-1} H^{(n_m)} \left([\boldsymbol{\lambda}^{(n_m)'\mathbf{T}}, f_m - \boldsymbol{\lambda}^{(n_m)'\mathbf{T}}\mathbf{1})]^{\mathbf{T}} \right) d\boldsymbol{\lambda}^{(n_1)'} \cdots d\boldsymbol{\lambda}^{(n_N)'} \tag{63}$$

where $H^{(n_m)}(\boldsymbol{f}^{(n_m)})$ is the n_mth-order transfer function, $x_m(t)$ is the input and $y_m(t)$ is the output of the mth system; $\boldsymbol{x}^{(n_m)}$ is the n_m-dimensional vector $[x_m(t), ..., x_m(t)]^{\mathbf{T}}$, and $\boldsymbol{\lambda}^{(n_m)'}$ and $\boldsymbol{f'}$ are vectors whose dimensions are $n_m - 1$ and $N - 1$, respectively.

VII. Conclusions

The research on Volterra system identification has received much more attention for the case of stationary input signals than cyclostationary input signal in the last thirty years since the seminal contribution in [68]. The dispersion of identification methods among quite different journals complicates the integration of the different methods already proposed and recent contributions ignore the most important results achieved by the use of cyclostationary input signals. As a result, further investigation that has been needed in order to provide experimental evidence of the advantages associated with the use of the cyclic properties of the input signals has not occurred. It is the intention of this review article to bring attention to the substantial merit of cyclostationarity exploitation for purposes of nonlinear system identification.

VIII. Acknowledgements

The author expresses his gratitude to Profs. William A. Gardner and Antonio Napolitano for their careful reading of the original manuscript, which led to significant improvements in the presentation.

References

[1] V. Volterra, *Sopra le funzioni che dipendono da altre funzioni*. Tipografia della R. Accademia dei Lincei, 1887.

[2] ——, *Sopra le funzioni dipendenti da linee*. Tip. della R. Accademia dei LIncei, 1887.

[3] ——, *Le**c**cons sur les équations intégrales et les équations intégro-différentielles: Le**c**cons professées à la Faculté des sciences de Rome en 1910*. Gauthier-Villars, 1913.

[4] ——, *Teoria de las funcionales y de las ecuaciones integrales e integro-differenciales. Conferencias explicadas en la Facultad de la Ciencas de la Universitad, 1925, redactas por L. Fantappié, Madrid*. Madrid: L. Fantappié, 1927.

[5] ——, *Theory of Functionals and of Integral and Integro-differential Equations*. London and Glasgow: Blackie & Son, Ltd., 1930, edited by Luigi Fantappiè. Authorized Translation by Miss M. Long.

[6] ——, *Theory of Functionals and of Integral and Integro-differential Equations*. New York: Dover Publications, 1948.

[7] A. Rosenblueth, N. Wiener, and J. Bigelow, "Behavior, purpose and teleology," *Philosophy of science*, vol. 10, no. 1, pp. 18–24, 1943.

[8] N. Wiener, "Response of a nonlinear device to noise," Radiation Lab Rept, Massachusetts Inst. Technology, Tech. Rep. 129, April 1942.

[9] ——, *Norbert Wiener# A Life in Cybernetics: Ex-Prodigy: My Childhood and Youth and I Am a Mathematician: The Later Life of a Prodigy*. MIT Press, 1956.

[10] ——, *Cybernetics or Control and Communication in the Animal and the Machine*. MIT press, 1948.

[11] H. E. Singleton, "A digital electronic correlator," Massachusetts Institute of Technology, Research Laboratory of Electronics, Tech. Rep. 152, February 1950.

[12] ——, "A digital electronic correlator," *Proceedings of the IRE*, vol. 38, no. 12, pp. 1422–1428, 1950.

[13] D. Brennan, "Probability theory in communication system engineering," in *Lectures on communication system theory*. McGraw-Hill, 1961.

[14] E. M. Hofstetter, "Random processes," in *The Mathematics of Physics and Chemistry*, vol. 2. D. Van Nostrand Co., 1964.

[15] W. A. Gardner, *Statistical Spectral Analysis: a Nonprobabilistic Theory*. Prentice-Hall, 1987.

[16] W. A. Gardner and W. A. Brown, "Fraction-of-time probability for time-series that exhibit cyclostationarity," *Signal Processing*, vol. 23, no. 3, pp. 273–292, 1991.

[17] S. Ikehara, "A method of Wiener in a nonlinear circuit," Research Laboratory of Electronics, Massachusetts Institute of Technology, Tech. Rep. 217, December 1951.

[18] A. G. Bose, "A theory of nonlinear systems," Research Laboratory of Electronics, Massachusetts Institute of Technology, Tech. Rep. 309, may 1956.

[19] A. Bose, "Nonlinear system characterization and optimization," *IRE Transactions on Information Theory*, vol. 5, no. 5, pp. 30–40, 1959.

[20] N. Wiener, *Nonlinear problems in random theory*. The Technology Press of MIT (later named The MIT Press), 1958.

[21] M. B. Brilliant, "Theory of the analysis of nonlinear systems," Massachusetts Inst of Tech, Cambridge, Research Lab of Electronics, Tech. Rep. 345, march 1958.

[22] D. A. George, "Continuous nonlinear systems," Massachusetts Institute of Technology. Research Laboratory of Electronics, Tech. Rep. 355, July 1959.

[23] G. Zames, "Nonlinear operators of system analysis," Massachusetts Institute of Technology, Research Laboratory of Electronics, Tech. Rep. 370, august 1960.

[24] Y. Ku and A. A. Wolf, "Volterra-Wiener functionals for the analysis of nonlinear systems," *Journal of the Franklin Institute*, vol. 281, no. 1, pp. 9–26, 1966.

[25] M. Schetzen, *The Volterra and Wiener Theories of Nonlinear Systems*. Wiley, 1980.

[26] W. J. Rugh, *Nonlinear System Theory: The Volterra/Wiener Approach*. Baltimore: The John Hopkins University Press, 1981.

[27] J. Barrett, "The use of functionals in the analysis of non-linear physical systems," *Ministry of Supply Great Britain, Statist, Advisory Unit Rept*, vol. 1, p. 57, 1957.

[28] ——, "The use of functionals in the analysis of non-linear physical systems," *International Journal of Electronics*, vol. 15, no. 6, pp. 567–615, 1963.

[29] ——, *Bibliography on Volterra series Hermite functional expansions and related subjects*. University of Technology, 1977.

[30] ——, "Problems arising from the analysis of randomly disturbed non-linear control systems," Ph.D. dissertation, University of Cambridge, 1960.

[31] H. Smets, "Analysis and synthesis of nonlinear systems," *IRE Transactions on Circuit Theory*, vol. 7, no. 4, pp. 459–469, 1960.

[32] L. A. Zadeh, "From circuit theory to system theory," *Proceedings of the IRE*, vol. 50, no. 5, pp. 856–865, 1962.

[33] G. Palm, "On representation and approximation of nonlinear systems," *Biological Cybernetics*, vol. 34, no. 1, pp. 49–52, 1979.

[34] S. Boyd and L. Chua, "Fading memory and the problem of approximating nonlinear operators with Volterra series," *IEEE Transactions on circuits and systems*, vol. 32, no. 11, pp. 1150–1161, 1985.

[35] P. D'Alessandro, A. Isidori, and A. Ruberti, "Realization and structure theory of bilinear dynamical systems," *SIAM Journal on Control*, vol. 12, no. 3, pp. 517–535, 1974.

[36] C. Bruni, G. Di Pillo, and G. Koch, "On the mathematical models of bilinear systems," *Ricerche di Automatica*,

vol. 2, no. 1, pp. 11–26, 1971.

[37] A. J. Krener, "Linearization and bilinearization of control systems," in *Proceedings of the 1974 Allerton Conference on Circuit and Systems Theory, Urbana III*, 1974.

[38] R. W. Brockett, "Volterra series and geometric control theory," *Automatica*, vol. 12, no. 2, pp. 167–176, 1976.

[39] R. H. Cameron and W. T. Martin, "The orthogonal development of non-linear functionals in series of Fourier-Hermite functionals," *Annals of Mathematics*, pp. 385–392, 1947.

[40] Y. Ku, "Taylor-Cauchy transforms for analysis of varying-parameter systems," *Proceedings of the IRE*, vol. 49, no. 6, pp. 1096–1097, 1961.

[41] M. Schetzen, "Nonlinear system modeling based on the Wiener theory," *Proceedings of the IEEE*, vol. 69, no. 12, pp. 1557–1573, 1981.

[42] J. Barrett, *Nonlinear System Analysis by Volterra and Hermite Functional Expansions*. University of Southampton, 2012.

[43] P. Gallman, "An iterative method for the identification of nonlinear systems using a Uryson model," *IEEE Transactions on Automatic Control*, vol. 20, no. 6, pp. 771–775, 1975.

[44] G. Palm and T. Poggio, "The Volterra representation and the Wiener expansion: validity and pitfalls," *SIAM Journal on Applied Mathematics*, vol. 33, no. 2, pp. 195–216, 1977.

[45] P. Z. Marmarelis, "Nonlinear dynamic transfer functions for certain retinal neuronal systems," Ph.D. dissertation, California Institute of Technology, 1972.

[46] D. G. Elson, "Prediction and simulation of stochastic time series using a non-linear learning filter," Ph.D. dissertation, Electrical Engineering Department, Imperial College of Science and Technology, South Kensington, London, august 1963.

[47] S. Orcioni, M. Pirani, and C. Turchetti, "Advances in Lee–Schetzen method for Volterra filter identification," *Multidimensional Systems and Signal Processing*, vol. 16, no. 3, pp. 265–284, 2005.

[48] H. Ogura, "Orthogonal functionals of the Poisson process," *IEEE Transactions on Information Theory*, vol. 18, no. 4, pp. 473–481, 1972.

[49] A. Segall and T. Kailath, "Orthogonal functionals of independent-increment processes," *IEEE Transactions on Information Theory*, vol. 22, no. 3, pp. 287–298, 1976.

[50] D. R. Brillinger, "The identification of point process systems," *The Annals of Probability*, pp. 909–924, 1975.

[51] S. Klein and S. Yasui, "Nonlinear systems analysis with non-Gaussian white stimuli; General basis functionals and

kernels (corresp.)," *IEEE Transactions on Information Theory*, vol. 25, no. 4, pp. 495–500, 1979.

[52] M. Schetzen, "Some problems in nonlinear theory," Ph.D. dissertation, Department of Electrical Engineering, M.I.T., Cambridge (MA), may 1961.

[53] ——, "Theory of pth-order inverses of nonlinear systems," *IEEE Transactions on Circuits and Systems*, vol. 23, no. 5, pp. 285–291, 1976.

[54] ——, "Some problems in nonlinear theory," Massachusetts Institute of Technology, Research Laboratory of Electronics, Tech. Rep. 390, July 1962.

[55] D. Brillinger, "The identification of polynomial systems by means of higher order spectra," *Journal of Sound and Vibration*, vol. 12, no. 3, pp. 301–313, 1970.

[56] A. N. Shiryaev, "Some problems in the spectral theory of higher-order moments. i," *Theory of Probability & Its Applications*, vol. 5, no. 3, pp. 265–284, 1960.

[57] P. Koukoulas and N. Kalouptsidis, "Nonlinear system identification using Gaussian inputs," *IEEE Transactions on signal Processing*, vol. 43, no. 8, pp. 1831–1841, 1995.

[58] L. J. Tick, "The estimation of "transfer functions" of quadratic systems," *Technometrics*, vol. 3, no. 4, pp. 563–567, 1961.

[59] S. Fakhouri, "Identification of the Volterra kernels of nonlinear systems," in *IEE Proceedings D-Control Theory and Applications*, vol. 6, no. 127, 1980, pp. 296–304.

[60] D. R. Brillinger and M. Rosenblatt, "Computation and interpretation of k-th order spectra," *Spectral Analysis of Time Series*, vol. 57, no. 2, pp. 189–232, 1967.

[61] D. R. Brillinger, "An introduction to polyspectra," *The Annals of Mathematical Statistics*, vol. 36, no. 5, pp. 1351–1374, 1965.

[62] W. A. Gardner, Ed., *Cyclostationarity in communications and signal processing*. IEEE, 1994, book edited by Gardner, William A.

[63] S. Billings, "Identification of nonlinear system - a survey," *IEE Proceedings D (Control Theory and Applications)*, vol. 127, pp. 272–285(13), November 1980.

[64] S. A. Billings, *Nonlinear system identification: NARMAX methods in the time, frequency, and spatio-temporal domains*. John Wiley & Sons, 2013.

[65] A. Dandawate and G. Giannakis, "Nonparametric polyspectral estimators for kth-order (almost) cyclostationary processes," *IEEE Transactions on Information Theory*, vol. 40, no. 1, pp. 67–84, 1994.

[66] C. Spooner and W. Gardner, "The cumulant theory of cyclostationary time-series. II. Development and applications," *IEEE Transactions on Signal Processing*, vol. 42, no. 12, pp. 3409–3429, 1994.

[67] A. Napolitano and C. Spooner, "Median-based cyclic polyspectrum estimation," *IEEE Transactions on Signal Processing*, vol. 48, no. 5, pp. 1462–1466, 2000.

[68] W. A. Gardner and T. L. Archer, "Exploitation of cyclostationarity for identifying the Volterra kernels of nonlinear systems," *IEEE transactions on Information Theory*, vol. 39, no. 2, pp. 535–542, 1993.

[69] T. Archer and W. Gardner, "New methods for identifying the Volterra kernels of a nonlinear system," in *1990 Conference Record Twenty-Fourth Asilomar Conference on Signals, Systems and Computers, 1990.*, vol. 2. IEEE, 1990, p. 592.

[70] J. Barrett, "Functional series representation of nonlinear systems–some theoretical comments," *IFAC Proceedings Volumes*, vol. 15, no. 4, pp. 313–318, 1982.

[71] J. Bartos and H. H. Sun, "Characterization of abrupt nonlinearity by the Volterra-Fourier method," *Journal of the Franklin Institute*, vol. 325, no. 4, pp. 465–484, 1988.

[72] S. Yasui, "Stochastic functional Fourier series, Volterra series, and nonlinear systems analysis," *IEEE Transactions on Automatic Control*, vol. 24, no. 2, pp. 230–242, 1979.

[73] E. Bedrosian and S. O. Rice, "The output properties of Volterra systems (nonlinear systems with memory) driven by harmonic and Gaussian inputs," *Proceedings of the IEEE*, vol. 59, no. 12, pp. 1688–1707, 1971.

[74] D. Mattera and L. Paura, "Higher-order cyclostationarity-based methods for identifying Volterra systems by input-output noisy measurements," *Signal Processing*, vol. 67, no. 1, pp. 77–98, 1998.

[75] ——, "Exploitation of cyclostationarity for identifying nonlinear Volterra systems by input-output noisy measurements," in *1996 Conference Record Thirtieth Asilomar Conference on Signals, Systems and Computers, 1996.*, vol. 1. IEEE, 1996, pp. 166–170.

[76] W. Gardner and C. Spooner, "The cumulant theory of cyclostationary time-series. I. Foundation," *IEEE Transactions on Signal Processing*, vol. 42, no. 12, pp. 3387–3408, 1994.

[77] W. Gardner and L. Paura, "Identification of polyperiodic nonlinear systems," *Signal Processing*, vol. 46, no. 1, pp. 75–83, 1995.

[78] D. Mattera, "Identification of polyperiodic Volterra systems by means of input-output noisy measurements," *Signal Processing*, vol. 75, no. 1, pp. 41–50, 1999.

[79] S. Prakriya and D. Hatzinakos, "Identification of symmetric/asymmetric quadratic nonlinear channels with cyclo-

stationary inputs," in *Proceedings of MILCOM '93 - IEEE Military Communications Conference*, vol. 2, 1993, pp. 653–657 vol.2.

[80] ——, "Efficient nonlinear channel identification using cyclostationary signal analysis," in *1993 IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 4, 1993, pp. 264–267 vol.4.

[81] G. Mzyk and G. Maik, "Nonparametric identification of Wiener system with a subclass of wide-sense cyclostationary excitations," *International Journal of Adaptive Control and Signal Processing*, vol. 38, no. 1, pp. 323–341, 2024.

[82] M. Rudko and D. Weiner, "Volterra systems with random inputs: A formalized approach," *IEEE Transactions on Communications*, vol. 26, no. 2, pp. 217–227, 1978.

[83] ——, "Addendum to "Volterra systems with random inputs: A formalized approach"," *IEEE Transactions on Communications*, vol. 27, no. 3, pp. 636–638, 1979.

[84] L. Izzo, A. Napolitano, and L. Paura, "MIMO Volterra system input/output relations for cyclic higher-order statistics," in *Proceedings of VII European Signal Processing Conference (EUSIPCO'94), Edinburgh, UK*, 1994.