Filter-based multiscale entropy analysis of complex physiological time series

Yuesheng Xu[1,2,*] and Liang Zhao[1]

[1]*Department of Mathematics, Syracuse University, Syracuse, New York 13244, USA*

[2]*Guangdong Province Key Lab of Computational Science, Sun Yat-sen University, Guangzhou 510275, China*

(Received 13 November 2012; revised manuscript received 1 August 2013; published 28 August 2013)

Multiscale entropy (MSE) has been widely and successfully used in analyzing the complexity of physiological time series. We reinterpret the averaging process in MSE as filtering a time series by a filter of a piecewise constant type. From this viewpoint, we introduce filter-based multiscale entropy (FME), which filters a time series to generate multiple frequency components, and then we compute the *blockwise* entropy of the resulting components. By choosing filters adapted to the feature of a given time series, FME is able to better capture its multiscale information and to provide more flexibility for studying its complexity. Motivated by the heart rate turbulence theory, which suggests that the human heartbeat interval time series can be described in piecewise linear patterns, we propose piecewise linear filter multiscale entropy (PLFME) for the complexity analysis of the time series. Numerical results from PLFME are more robust to data of various lengths than those from MSE. The numerical performance of the adaptive piecewise constant filter multiscale entropy without prior information is comparable to that of PLFME, whose design takes prior information into account.

DOI: 10.1103/PhysRevE.88.022716 PACS number(s): 87.18.−h

I. INTRODUCTION

The complex fluctuations exhibited by a signal generated from a physiological system contain information of underlying interacting mechanisms which regulate the system. Quantifying the "complexity" of physiological signals has drawn considerable attention [1–6]. In general, there is no precise mathematical definition for "physiological complexity," and no single statistical measure can be used to assess the complexity of physiological systems [6]. Intuitively, complexity is referred to as "meaningful structural richness." Several entropy metrics, such as approximation entropy [7] and sample entropy [8], were proposed to measure the regularity of a system by quantifying the degree of predictability of a series of data points generated from the system. However, irregularity is not the same as complexity. For example, the entropy measures mentioned above assign the highest values to uncorrelated random signals (Gaussian white noise), which are highly unpredictable but not structurally "complex." Moreover, when they are applied to the human heartbeat interval time series (HHITS), certain pathologies including cardiac arrhythmias such as atrial fibrillation are assigned a higher entropy value than healthy dynamics, which represent more physiologically complex, adaptive states. The reason these entropy matrices do not work on physiological systems is that they only measure complexity at the single scale. While biological systems operate across multiple spatial and temporal scales, their complexity should also be measured multiscaled. Therefore, these entropy matrices are not direct indices of physiological complexity. In this paper, as in [3,5], we take the point of view that for a physiological system, complexity should be measured across multiple scales using entropy matrices, and the higher the multiscale entropy (MSE) value is, the more complex the system is.

The remarkable MSE [3,5] takes into account that biological systems operate across multiple spatial and temporal scales when measuring the complexity of the physiological time series, and it examines the physiological dynamics over multiple scales. When applied to HHITS, MSE not only provides a meaningful measure for the complexity of the physiological time series, but it also shows good results in distinguishing different patterns from subjects with different ages and heart diseases. The hierarchical entropy [9] considers both the low- and high-frequency components of a time series and uses a binary structure in analyzing the signal.

A crucial step in MSE is the coarse-graining procedure, which assesses the entropy rate. It is achieved by an averaging process, extracting low-frequency components of the time series, at different scales. This procedure can be reinterpreted from a filter viewpoint as applying a piecewise constant low-pass filter which has a matrix representation to the time series. We shall take this point of view in studying filter-based multiscale entropy analysis.

The purpose of this paper is to introduce filter-based multiscale entropy analysis. We shall provide a general framework of filter-based multiscale entropy (FME) and theoretical results of FME for Guassian white noise as well as $1/f$ noise. The application of FME to HHITS will also be thoroughly studied.

Specifically, with FME the time series is passed through desired fine-to-coarse filter matrices at different scales, and a blockwise sample entropy value is calculated at each scale. On the one hand, this general setting will give us an insightful understanding of MSE; on the other hand, it will allow us to choose a filter that better fits the given data when certain prior information of the data is available to improve the entropy result. When prior information of the time series is not available, we can develop adaptive filters which extract the main feature of the time series.

We consider in this paper HHITS as our main study case. Heart rate turbulence (HRT), the technique of acceleration-deceleration oscillation analysis proposed in [10], suggests that HHITS can be described in the piecewise linear pattern.

*Author to whom all correspondence should be addressed: yxu06@syr.edu

The time series generated by different heart conditions show distinguished differences in this pattern. Therefore, using piecewise linear filters for capturing this pattern is highly desirable. We apply the piecewise linear filter to HHITS before measuring their complexity and find that aging may reduce the complexity of the cardiac system more than congestive heart failure. Numerical results from piecewise linear filter multiscale entropy (PLFME) are more robust to data of various lengths than those from MSE. We furthermore design an adaptive filter for HHITS (without prior information of HHITS) and use it to develop adaptive piecewise constant filter multiscale entropy (APCFME). In the study of HHITS, the numerical performance of APCFME is comparable to that of PLFME.

We organized this paper into six sections and two Appendixes. In Sec. II, we describe the coarse-graining processing using filters and a blockwise sample entropy for computing the resulting filter-based multiscale entropy. Then in Sec. III, we study FME for the Gaussian noise and the $1/f$ noise. For these cases of study, we provide theoretical results, with their detailed proofs reported in the Appendixes, and as well as numerical results. Section IV is devoted to application of PLFME to HHITS. In Sec. V, we design an adaptive piecewise constant filter and use it in developing APCFME. Numerical results of APCFME applied to Gaussian noise, $1/f$ noise, and HHITS are also presented in this section. We draw conclusions in Sec. VI.

II. FILTER-BASED MULTISCALE ENTROPY

We motivate FME from a filter viewpoint which reinterprets the averaging process in MSE as filtering a time series through a lower pass filter of a *piecewise constant* type in generating its multiple frequency components. Piecewise constant filters may be suitable for signals which can be described in piecewise constant patterns but may not be suitable for others. To make MSE more robust to signals with a different nature, we introduce FME which considers the meaningful structural complexity of a physiologic system over multiple spatial and temporal scales resulting from filters appropriate for the specific physiological system. Specifically, at each scale, from finer to coarser, the time series is passed through a desired filter to capture its characteristic pattern. For example, a piecewise polynomial filter of order k can be used to approximate a time series which can be intrinsically represented by such a function. When prior information of the signal is available, one can use it in the filter design, and when it is not available, one may construct filters adaptively from the given signal.

A filter may be described in terms of a matrix. For example, the Haar filter is the 1×2 matrix $[\frac{1}{2}, \frac{1}{2}]$. The piecewise polynomial filter may be derived from the wavelets on invariant sets [11], and a general construction of filters of this type was discussed in [12]. From [13], the piecewise linear filter is the 2×4 matrix given by

$$A := \frac{1}{2}\begin{bmatrix} 1 & 0 & 1 & 0 \\ -\frac{\sqrt{3}}{2} & \frac{1}{2} & \frac{\sqrt{3}}{2} & \frac{1}{2} \end{bmatrix} \tag{1}$$

and the piecewise quadratic filter is the 3×6 matrix given by

$$B := \frac{1}{2}\begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 \\ -\frac{\sqrt{3}}{2} & \frac{1}{2} & 0 & \frac{\sqrt{3}}{2} & \frac{1}{2} & 0 \\ 0 & -\frac{\sqrt{15}}{4} & \frac{1}{4} & 0 & \frac{\sqrt{15}}{4} & \frac{1}{4} \end{bmatrix}. \tag{2}$$

A coarse-graining process of a time series $\mathbf{x} := [x_0, \ldots, x_{N-1}]$ of real numbers can be viewed as a matrix multiplication of the vector $\mathbf{x}$ (we use the same notation $\mathbf{x}$ for the time series and the vector). Specifically, at each scale $\tau = 2,3, \ldots,$ a matrix $A^{(\tau)} \in \mathbb{R}^{p_\tau \times q_\tau}$ is chosen as a filter for $\mathbf{x}$. For matrices $P := [p_{jk}]$ and Q we define the Kronecker product $P \otimes Q := [p_{jk}Q]$. By $\lfloor \cdot \rfloor$ we denote the floor function. At scale τ, the coarse-grained time series is constructed by $A^{(\tau)}$ as $\mathbf{y}^\tau := (I_n \otimes A^{(\tau)})\mathbf{x}$, where $n := \lfloor \frac{N}{q_\tau} \rfloor$ and I_n is the $n \times n$ identity matrix. If $N = nq_\tau + k$, for some integers n and k with $1 \leqslant k < q_\tau$, we shall drop the last k components of $\mathbf{x}$ when constructing the coarse-grained time series since such insignificant loss of a few components will barely affect the complexity of the whole system. Thus, in each coarse-grained procedure, the time series is partitioned into n blocks with each having q_τ components and being transformed by $A^{(\tau)}$ to another block of p_τ components.

The coarse-graining process can also be viewed as the application of the same filter matrix recursively to $\mathbf{x}$. For a filter matrix $A \in \mathbb{R}^{p \times q}$ and $\tau = 2,3, \ldots,$ the coarse-grained time series at the scale τ is obtained recursively by $\mathbf{y}^\tau = (I_{n_{\tau-1}} \otimes A)\mathbf{y}^{\tau-1}$ with $\mathbf{y}^1 := \mathbf{x}$ and $n_{\tau-1} := \lfloor \frac{N_{\tau-1}}{q} \rfloor$, where $N_{\tau-1}$ is the length of the time series $\mathbf{y}^{\tau-1}$. For example, for PLFME, in the above formula A is the piecewise linear filter defined in (1) and $q = 4$.

We recall the notion of sample entropy [8]. Let $\mathbb{Z}_m := \{0,1, \ldots, m-1\}$ for a positive integer m. We denote by $\mathbf{x}(i)$ the ith component of a time series $\mathbf{x}$. For a given $\mathbf{x}$, we construct a sequence $\mathbf{u}_m := \{u_m(j) : j \in \mathbb{Z}_{N-m}\}$, where $u_m(j) := [x(j+k) : k \in \mathbb{Z}_m]$ are vectors of m data points, with m being the length of the pattern templates. The distance between $u_m(\ell)$ and $u_m(j)$ is defined as $d[u_m(\ell), u_m(j)] := \max\{|x(\ell+k) - x(j+k)| : k \in \mathbb{Z}_m\}$. For a given tolerance $r > 0$ and a fixed integer $\ell \in \mathbb{Z}_{N-m}$, we let B_ℓ^m denote the number of vectors $u_m(j)$ with $j > \ell$ which satisfy $d[u_m(\ell), u_m(j)] \leqslant r$. The number r serves as the tolerance for accepting matches, and $u_m(\ell)$ is called the template. Then the probability of vectors $u_m(j) \in \mathbf{u}_m$ that are near the template $u_m(\ell)$ within tolerance r is given by $C_\ell^m(\mathbf{x}, r) := B_\ell^m/(N-m+1)$. Let $C^m(\mathbf{x}, r) := \sum_{\ell=0}^{N-m} C_\ell^m(\mathbf{x}, r)$. The sample entropy of $\mathbf{x}$ is defined by

$$S_m(\mathbf{x}, r) := -\ln\left[\frac{C^{m+1}(\mathbf{x}, r)}{C^m(\mathbf{x}, r)}\right].$$

We propose a blockwise sample entropy (BSE) for the filtered time series $\mathbf{y}^\tau$ at each scale. The use of BSE (instead of the standard sample entropy) is to adjust to the size of the filters used in FME which result in output signals having blocks consisting of more than one component. We now introduce BSE for the coarse-grained time series $\mathbf{y}^\tau$ for $\tau \geqslant 2$. It follows the same idea as sample entropy and is designed to suit the structure of $\mathbf{y}^\tau$. Note that $\mathbf{y}^\tau$ consists of n blocks, each of which has p_τ components and is obtained from a block of $\mathbf{x}$ transformed by the same matrix $A^{(\tau)}$.

Components in each block represent different information of $\mathbf{x}$, captured by different rows of the filter $A^{(\tau)}$. For example, different rows of the linear filter transform $\mathbf{x}$ to different components in the piecewise linear function basis. For this reason, in calculating BSE, we consider each block of $\mathbf{y}^\tau$ as a single unit and apply the sample entropy calculation process to $\mathbf{y}^\tau$. To this end, we write $\mathbf{y}^\tau = [\mathbf{y}^\tau_0, \ldots, \mathbf{y}^\tau_{n-1}]$, where each $\mathbf{y}^\tau_j$ is a vector with p_τ components and $[\mathbf{y}^\tau_j]_s := \mathbf{y}^\tau(jp_\tau + s)$ for each $s \in \mathbb{Z}_{p_\tau}$ with $[\mathbf{d}]_s$ denoting the sth component of a vector $\mathbf{d}$. Then the sequence corresponding to $\mathbf{u}_m$ in sample entropy is constructed as $\mathbf{u}^\tau_m := \{u^\tau_m(j) : j \in \mathbb{Z}_{n-m}\}$, where $u^\tau_m(j) := [\mathbf{y}^\tau_{j+k} : k \in \mathbb{Z}_m]$ consists of m vectors. Let $\sigma_s(u^\tau_m(\ell), u^\tau_m(j)) := \max\{|[\mathbf{y}^\tau_{\ell+k}]_s - [\mathbf{y}^\tau_{j+k}]_s| : s \in \mathbb{Z}_m\}$. The distance between $u^\tau_m(\ell)$ and $u^\tau_m(j)$ is then defined as a vector of p_τ components by $\mathbf{d}[u^\tau_m(\ell), u^\tau_m(j)] := [\sigma_s(u^\tau_m(\ell), u^\tau_m(j)) : s \in \mathbb{Z}_{p_\tau}]$. For a given $u^\tau_m(\ell)$, we denote by $B^{m,\tau}_\ell$ the number of vectors $u^\tau_m(j)$ with $j > \ell$ which satisfy $\sigma_s(u^\tau_m(\ell), u^\tau_m(j)) \leqslant r^\tau_s$ for each $s \in \mathbb{Z}_{p_\tau}$, where $r^\tau_s := \sum_{t=1}^{q} |A^{(\tau)}_{st}| r$. Here r is the tolerance used in calculating the sample entropy of $\mathbf{x}$, and $A^{(\tau)}_{st}$ is the (s,t) entry of $A^{(\tau)}$. Let $r^\tau := [r^\tau_0, r^\tau_1, \ldots, r^\tau_{p^\tau - 1}]$ and $\tilde{C}^m_\ell(\mathbf{y}^\tau, r^\tau) := B^{m,\tau}_\ell/(n-m+1)$, $\ell \in \mathbb{Z}_{n-m+1}$. BSE $\tilde{S}_m(\mathbf{y}^\tau, r^\tau)$ of $\mathbf{y}^\tau$ is defined in the same way as the sample entropy with $C^m_\ell(\mathbf{x}, r)$ replaced by $\tilde{C}^m_\ell(\mathbf{y}^\tau, r^\tau)$. That is,

$$\tilde{S}_m(\mathbf{y}^\tau, r^\tau) := -\ln\left[\frac{\tilde{C}^{m+1}_\ell(\mathbf{y}^\tau, r^\tau)}{\tilde{C}^m_\ell(\mathbf{y}^\tau, r^\tau)}\right]. \tag{3}$$

The choice of parameter r^τ in BSE is crucial. In MSE, the same value r was used for different scales. Adjusted to the decreasing variance of the filtered time series in scales, the parameter r in MSE was adjusted in [14] as a certain percentage of the standard deviation of the filtered time series at each scale. In BSE, the parameter r^τ is calculated from a different viewpoint. Since the filtering process in FME is a transformation of the time series by a matrix, measuring the similarity of the components in the filtered time series should be related to the filter matrix. Thus the parameter r^τ is a vector whose components are transformed by the corresponding rows of the filter matrix. Each component r^τ_s, $s \in \mathbb{Z}_{p_\tau}$, is to measure the similarity of the corresponding components among different blocks in the filtered time series.

We now elaborate the relation of MSE and FME. In MSE, the consecutive coarse-grained time series $\{\mathbf{y}^\tau\}$ is constructed according to the equation $\mathbf{y}^\tau(j) = 1/\tau \sum_{k=j\tau}^{(j+1)\tau-1} \mathbf{x}(k)$, $j \in \mathbb{Z}_{\lfloor N/\tau \rfloor}$. The time series is actually filtered by the $1 \times \tau$ matrix $C^{(\tau)} := [\frac{1}{\tau}, \frac{1}{\tau}, \ldots, \frac{1}{\tau}]$ at scale τ. That is, $\mathbf{y}^\tau = (I_{\lfloor \frac{N}{\tau} \rfloor} \otimes C^{(\tau)})\mathbf{x}$. Thus $\mathbf{y}^\tau$ has $\lfloor N/\tau \rfloor$ blocks with each having one component. In this case, BSE degenerates to the standard sample entropy. Therefore, MSE is actually a special case of FME with the piecewise constant filter $C^{(\tau)}$ at scale τ. We also remark that when $\tau = 2^k$, for $k = 1, 2, \ldots$, filtering a time series by $C^{(\tau)}$ is equivalent to filtering it by the Haar filter k times recursively.

III. FME FOR GAUSSIAN AND $1/f$ NOISE

In this section, we discuss the behavior of FME in simulating white noise, a completely irregular signal, and $1/f$ noise, a correlated signal.

A. FME for Gaussian noise

We first apply FME to Gaussian white noise. Intuitively, complexity is associated with "meaningful structural richness" [15]. There is no straightforward correspondence between regularity and complexity. For example, uncorrelated random signals, such as Gaussian white noise, are highly unpredictable but not structurally complex. We present both theoretical and numerical results of Gaussian white noise from FME. We shall see that the entropy measure of the Gaussian white noise decreases as the scale increases. This indicates the lack of the complexity of the Gaussian white noise.

For a positive integer N, let $\mathbf{x} := [x_j : j \in \mathbb{Z}_N]$ denote a random vector taking values in $\mathbb{R}^N$. When the components $x_j, j \in \mathbb{Z}_N$, are independent and have the same Gaussian distribution in the sense that they have the same mean and standard deviation, we call $\mathbf{x}$ a real Gaussian random vector, we call a component of $\mathbf{x}$ a Gaussian random variable, and we call an instance of $\mathbf{x}$ the Gaussian white noise. In this subsection, we use $\mathbf{g} := [g_j : j \in \mathbb{Z}_N]$ for the real Gaussian random vector, with the mean of g_j being 0 and the standard deviation of g_j being δ.

Given a filter $A^{(\tau)} \in \mathbb{R}^{p_\tau \times q_\tau}$, we consider BSE of the filtered Gaussian random vector,

$$\mathbf{g}^\tau := (I_n \otimes A^{(\tau)})\mathbf{g}, \tag{4}$$

where $n = \lfloor \frac{N}{q_\tau} \rfloor$. Recall that BSE estimates the negative natural logarithm of the conditional probability that the distance between two blocks in $\mathbf{g}^\tau$ is small (measured by r_s in the description of BSE) given that the distance between the two preceding blocks is also small. This conditional probability can be analytically expressed by the probability density function of all of the blocks in $\mathbf{g}^\tau$ (Lemma 2 in Appendix A) and the independence of blocks in $\mathbf{g}^\tau$ (Lemma 3 in Appendix A). For notational convenience, in the remaining part of this section and Appendix A we use the same notation $\tilde{S}_m(\mathbf{g}^\tau, r)$ for the theoretical value of BSE of $\mathbf{g}^\tau$. For a positive integer N, let $\mathbb{Z}^+_N := \{1, 2, \ldots, N\}$. For a given $r > 0$, we let

$$r^\tau_s := \sum_{t=1}^{q_\tau} |A^{(\tau)}_{st}| r, \quad s \in \mathbb{Z}^+_{p_\tau}.$$

For a given vector $\mathbf{y} \in \mathbb{R}^{p_\tau}$, we define

$$\Omega_\mathbf{y} := \left([\mathbf{y}]_1 - r^\tau_1, [\mathbf{y}]_1 + r^\tau_1\right) \times \cdots \times \left([\mathbf{y}]_{p_\tau} - r^\tau_{p_\tau}, [\mathbf{y}]_{p_\tau} + r^\tau_{p_\tau}\right).$$

Clearly, $\Omega_\mathbf{y} \subset \mathbb{R}^{p_\tau}$. For the standard deviation δ of g_j, we define the matrix $\Sigma \in \mathbb{R}^{p_\tau \times p_\tau}$ by

$$\Sigma_{st} := \sum_{j=1}^{p_\tau} A^{(\tau)}_{sj} A^{(\tau)}_{tj} \delta^2, \quad 1 \leqslant s, \quad t \leqslant p_\tau.$$

We assume that the matrix Σ is invertible. We shall show in Lemma 2 that all of the blocks in $\mathbf{g}^\tau$ have the same probability density function with the covariance matrix Σ. We next present $\tilde{S}_m(\mathbf{g}^\tau, r)$ in terms of the matrix Σ. To this end, we let

$$\mathbf{I}(\Omega_\mathbf{y}) := \int_{\Omega_\mathbf{y}} \lambda_\Sigma \exp\left(-\frac{1}{2}\mathbf{x}^T \Sigma^{-1}\mathbf{x}\right) d\mathbf{x}, \tag{5}$$

where $\lambda_\Sigma := \frac{1}{(2\pi)^{p_\tau/2}|\Sigma|^{1/2}}$.

Proposition 1. If $A^{(\tau)} \in \mathbb{R}^{p_\tau \times q_\tau}$ and $\mathbf{g}^\tau$ is defined as in (4), then for any $r > 0$

$$\tilde{S}_m(\mathbf{g}^\tau, r^\tau) = -\ln\left\{\int_{\mathbb{R}^{p_\tau}} \mathbf{I}(\Omega_\mathbf{y})\lambda_\Sigma \exp\left(-\frac{1}{2}\mathbf{y}^T\Sigma^{-1}\mathbf{y}\right) d\mathbf{y}\right\}. \tag{6}$$

Details of the proof of Proposition 1 can be found in Appendix A.

Formula (6) holds for a general filter matrix $A^{(\tau)}$. In particular, when we choose $A^{(\tau)} = C^{(\tau)}$, it recovers the theoretical results originally proved in [5] of MSE for Gaussian white noise.

We next present a special result when the rows of the filter matrix A are orthogonal. Examples of such filter matrices include the piecewise polynomial filter of order k. In this case, in addition to the independence of blocks of the filtered Gaussian white noise, elements within each block are also independent (Lemma 4 in Appendix A). We use erf to denote the error function defined by

$$\text{erf}(x) := \frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2} dx.$$

For given $r > 0$ and matrix $A \in \mathbb{R}^{p\times q}$, we let $\delta(A,j) := \sqrt{\sum_{k=1}^p A_{jk}^2}\delta$ and

$$r_s := \sum_{t=1}^q |A_{st}| r, s \in \mathbb{Z}_p^+. \tag{7}$$

Then $\delta(A,j)$ is the standard deviation of the jth element in each block of the filtered Gaussian random vector (see Lemma 4). For real numbers a and b, we define

$$E(a,b) := \text{erf}\left(\frac{a+b}{\delta(A,j)}\right) - \text{erf}\left(\frac{a-b}{\delta(A,j)}\right).$$

Proposition 2. If matrix $A \in \mathbb{R}^{p\times q}$ has orthogonal rows, then for $\tilde{\mathbf{g}} := (I_{\lfloor \frac{N}{q} \rfloor} \otimes A)\mathbf{g}$ and for any $r > 0$,

$$\tilde{S}_m(\tilde{\mathbf{g}}, r^\tau) = -\ln\left(\prod_{j=1}^p \frac{1}{\sqrt{2\pi}\delta(A,j)}\int_{\mathbb{R}} E(x_j, r_j) \times \exp\frac{-x_j^2}{2[\delta(A,j)]^2} dx_j\right). \tag{8}$$

The proof of Proposition 2 is presented in Appendix A. Since the Gaussian noise is a completely irregular signal, one expects that its complexity, measured by FME, decreases as the scale increases. This is the case when the filter matrix A used in FME has orthogonal rows and satisfies the condition

$$A^T A = \rho^2 I_p \tag{9}$$

for some constant $\rho \in (0,1)$. This is stated in the next theorem. For this purpose, we let

$$\mathbf{g}^\tau := \left(I_{\lfloor \frac{N_{\tau-1}}{q} \rfloor} \otimes A\right)\left(I_{\lfloor \frac{N_{\tau-2}}{q} \rfloor} \otimes A\right)\cdots\left(I_{\lfloor \frac{N_1}{q} \rfloor} \otimes A\right)\mathbf{g}, \tag{10}$$

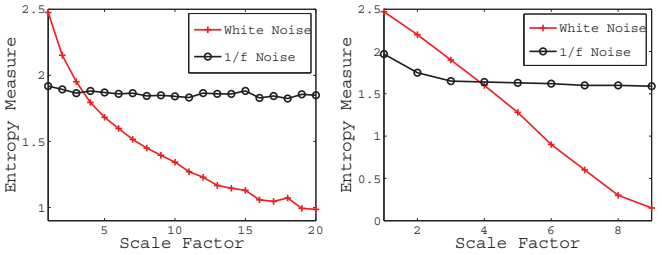
where N_j is the length of the time series $\mathbf{g}^j, j = 2,3,\ldots,\tau-1$, and N_1 is the length of $\mathbf{g}$.

Theorem 1. If $\mathbf{g} := [g_j : j \in \mathbb{Z}_N]$ is a real Gaussian random vector with mean 0 and standard deviation δ, $A \in \mathbb{R}^{p\times q}$ is a filter with orthogonal rows and satisfies condition (9), $m \geqslant 1$, $r > 0$, and $\mathbf{g}^\tau$ is defined as in (10), then for $\tau_1 > \tau_2$,

$$\tilde{S}_m(\mathbf{g}^{\tau_1}, r^{\tau_1}) < \tilde{S}_m(\mathbf{g}^{\tau_2}, r^{\tau_2}).$$

The proof of Theorem 1 follows from Proposition 2 and the fact that the error function is strictly increasing. Details of the proof are presented in Appendix A.

It can be verified that the piecewise polynomial filters of order k whose construction was described in [12] have orthogonal rows and satisfy the condition (9). Hence, the hypotheses of Theorem 1 are satisfied for this class of filters, and as a result, the corresponding filter-based multiscale entropy of the Gaussian noise decreases as the scale increases. This fact is further confirmed by the numerical example. Numerical results from PLFME and also MSE for Gaussian white noise are presented in Fig. 1. Unless stated otherwise, all entropy values presented in this paper are computed by choosing $m := 2$ and r being 15% of the time-series standard deviation. A similar pattern in which the entropy value decreases as the scale increases is shown in both of the methods.



(a) MSE (b) PLFME

FIG. 1. (Color online) MSE and PLFME for Gaussian white noise (mean 0, variance 1) and $1/f$ noise, $N = 8 \times 10^4$.

B. FME for $1/f$ noise

Now we apply FME to $1/f$ noise. Note that $1/f$ noise can be observed in various physical, chemical, and biological systems [16]. It is the signal whose power spectral density is proportional to the reciprocal of its frequency. To describe $1/f$ noise, we recall complex Gaussian variables and the discrete Fourier transform. As usual, we let $i = \sqrt{-1}$ be the imaginary unit and denote the complex plane by $\mathbb{C}$. A complex variable $z := x + iy$ is called a complex Gaussian random variable if both x and y are real independent Gaussian variables with the mean 0 and the same standard deviation δ. The corresponding probability density function for the complex Gaussian random variable z is given by

$$\rho(z) := \frac{1}{\pi\delta_z^2} e^{|z|^2/\delta_z^2}, \quad z \in \mathbb{C},$$

where $\delta_z := \sqrt{2}\delta$. Given $n \in \mathbb{N}$, we let $\theta_n := \frac{2\pi}{2^n}$ and define the discrete Fourier transform F_n by a $2^n \times 2^n$ matrix,

$$F_n := \frac{1}{2^n}[e^{-i\theta_n k\ell} : k \in \mathbb{Z}_{2^n}, \ell \in \mathbb{Z}_{2^n}]. \tag{11}$$

For a random vector x taking values in $\mathbb{R}^{2^N}$, we use $\hat{\mathbf{x}}$ to denote the discrete Fourier transform of $\mathbf{x}$, that is, $\hat{\mathbf{x}} := F_N\mathbf{x}$. We write $\hat{\mathbf{x}} := [z_k : k \in \mathbb{Z}_{2^N}]^T$. It is well known that the discrete Fourier

transform has the symmetric property

$$z_{2^{N-1}+k} = \bar{z}_{2^{N-1}-k}, \quad k \in \mathbb{Z}^+_{2^{N-1}}. \tag{12}$$

We need only to obtain the first $2^{N-1}+1$ components of the vector $\hat{\mathbf{x}}$ since the remaining components may be obtained from the symmetry property.

We describe the $1/f$ noise following [17]. If z_k, $k \in \mathbb{Z}^+_{2^{N-1}-1}$, are independent complex Gaussian random variables with mean 0, z_0 and z_{2^N-1} are real Gaussian random variables with mean 0, and there is a positive constant c such that for all $k \in \mathbb{Z}_{2^{N-1}+1}$ the standard deviation δ_k of z_k satisfies $\delta_k \leqslant \frac{c}{k+1}$, then we call $\mathbf{x}$ a $1/f$ random vector and we call an instance of $\mathbf{x}$ a $1/f$ noise. In this subsection, we use $\mathbf{f} := [f_k : k \in \mathbb{Z}_{2^N}]^T$ to denote a $1/f$ random vector and $\hat{\mathbf{f}} := [z_k : k \in \mathbb{Z}_{2^N}]^T$ to denote the discrete Fourier transform of $\mathbf{f}$.

It is known that $1/f$ noise contains complex structures across multiple time scales [18,19]. We shall show that the filtered $1/f$ noise is again $1/f$ noise if the filters satisfy certain conditions. This indicates that the filtered $1/f$ signal is as complex as the original $1/f$ signal.

We start with a simple filter A which has the form

$$A := [\alpha, \beta],$$

where $\alpha, \beta \in \mathbb{R}$. The next proposition shows that the filtered $1/f$ signal, $(I_{2^{N-1}} \otimes A)\mathbf{f}$, is again a $1/f$ signal for most of the filters of this form. The proof will be given in Appendix B by using Lemma 6 to Lemma 12. To present this result, we write $A\mathbf{f} := (I_{2^{N-1}} \otimes A)\mathbf{f}$ and let $\delta_k, \delta_{A,k}$ denote the standard derivations of the $(k+1)$th random variable of $\hat{\mathbf{f}}$ and $\hat{A\mathbf{f}}$, respectively.

Proposition 3. If $\mathbf{f}$ is a $1/f$ random vector and (α, β) satisfies the condition

$$\alpha + \beta \neq 0, \tag{13}$$

then $A\mathbf{f}$ is also a $1/f$ random vector. Moreover, if there exists a positive constant c such that for all $k \in \mathbb{Z}_{2^{N-1}+1}$, $\delta_k^2 \leqslant \frac{c}{1+k}$, then for all $k \in \mathbb{Z}_{2^{N-2}+1}$, $\delta_{A,k}^2 \leqslant \frac{c'}{k+1}$, where $c' = 2(\alpha^2 + \beta^2)c$.

Note that the special case of Proposition 3 with $[\alpha, \beta] := [\frac{1}{2}, \frac{1}{2}]$ was proved in [9]. If (α, β) does not satisfy condition (13), then $\alpha = -\beta$. In this case, $(I_{2^{N-1}} \otimes A)\mathbf{f}$ may not be $1/f$ noise. For example, when A is the high pass Haar filter, that is, $A := [\frac{1}{2}, -\frac{1}{2}]$, it was verified in [9] by a numerical experiment that $(I_{2^{N-1}} \otimes A)\mathbf{f}$ is not $1/f$ noise and the entropy value of the filtered $1/f$ signal will decrease as the scale increases. In addition, if we take $A := [1,0]$ ($A := [0,1]$) in Proposition 3, we can see that the random vector consisting of the odd components (the even components) of a $1/f$ random vector is still a $1/f$ random vector. This result is summarized in the next corollary.

Corollary 1. If $\mathbf{f} := [f_0, f_1, \ldots, f_{N-1}]$ is a $1/f$ random vector, then $\mathbf{u} := [u_m : m \in \mathbb{Z}_{N/2+1}]$ with $u_m = f_{2m}$ and $\mathbf{v} := [v_m : m \in \mathbb{Z}_{N/2+1}]$ with $v_m = f_{2m+1}$ are also $1/f$ random vectors.

Though the result in Proposition 3 is only for filters of a simple form, it can be utilized (together with Corollary 1) to establish below that a filtered $1/f$ noise is still $1/f$ noise. The proof of this result will be provided in Appendix B.

Proposition 4. If $\mathbf{f}$ is $1/f$ random vector and the matrix A is defined as in (1), then $(I_{2^{N-2}} \otimes A)\mathbf{f}$ is also a $1/f$ random vector.

Numerical results from MSE and PLFME for $1/f$ noise are presented in Fig. 1. Results from both of these methods are consistent with the fact that $1/f$ noise contains complex structures across multiple scales [18,19].

IV. APPLICATION TO HUMAN HEARTBEAT INTERVAL TIME SERIES

We apply FME to HHITS to study the loss of complexity, a generic feature of pathologic dynamics. Specifically, we apply the piecewise linear filter recursively to HHITS of healthy young subjects (YOUNG), healthy old subjects (OLD), subjects with cardiac arrhythmia, atrial fibrillation (AF), and subjects with severe congestive heart failure (CHF), and we compute BSE of the resulting signals of multiple scales. We test the hypothesis that healthy heart interbeat interval dynamics are more complex than those with pathology. Our numerical results also suggest that aging may reduce the complexity of the heart interbeat interval more than CHF. This finding is robust to data of different lengths.

In this consideration, the use of the piecewise linear filter in FME is motivated by a study of the biological mechanism of the cardiac system described by HRT [10]. HRT describes short-term fluctuations in the sinus cycle length that follow spontaneous ventricular premature complexes (VPCs). The physiological pattern described in HRT consists of brief heart rate acceleration, which is followed by more gradual heart deceleration before the rate returns to a pre-ectopic level. Following singular VPCs, the HRT pattern is frequently masked by a heartbeat interval time series. Consequently, HRT is usually assessed from Holter recordings as an average response to VPCs over longer periods (e.g., 24 h). From such recordings, the VPC tachogram is constructed by aligning and averaging sequences of heartbeat interval time series surrounding isolated VPCs. Figure 2 from [20] is VPCs tachograms showing normal (left) and abnormal (right) HRT. From Fig. 2, we observe that different heart conditions show distinguished differences in HRT pattern.

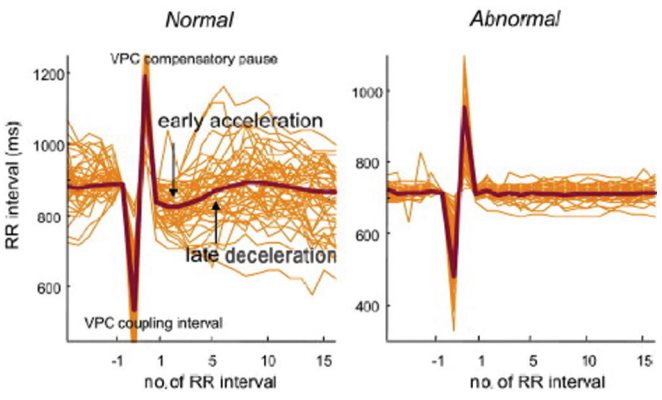


FIG. 2. (Color online) VPC tachograms showing normal (left) and abnormal (right) HRT. Thin orange curves show single VPC tachograms. Bold brown curves show the averaged VPC tachogram over 24 h.

Several parameters characterizing the HRT pattern were proposed [10,21–23]. Among these parameters, turbulence onset (TO) and turbulence slope (TS) [10] quantify two phases of HRT, namely early acceleration and late deceleration. These two parameters are meaningful in the clinical sense, especially in risk prediction and monitoring of disease progression in several pathologies (see [20] and references therein). TO and TS represent the regression slope of the corresponding sequences of HHITS (acceleration and deceleration) surrounding isolated VPCs. The acceleration phase surrounding an isolated VPC is characterized by a negative value of TO, and the deceleration phase surrounding an isolated VPC is characterized by a positive value of TS. Both TO and TS are constants for each sequence of HHITS surrounding VPCs and representing the acceleration and deceleration phase, respectively. This implies that regression slopes of the acceleration and deceleration sequences of HHITS surrounding VPCs have a piecewise constant pattern. Therefore, sequences of HHITS surrounding VPC may have a piecewise linear pattern. This inspires us to use a piecewise linear filter for the cardiac signal to capture its piecewise linear pattern. Though HRT only describes sequences of HHITS surrounding VPCs, we use a piecewise linear filter for the entire HHITS assuming that the piecewise linear pattern represents the entire HHITS better than the piecewise constant pattern, which is captured by the piecewise constant filter used in MSE. In Fig. 3, we compare the HHITS with their piecewise linear representations, where the signals in column (b) are obtained from the original HHITS in column (a) by the linear filter $\frac{1}{4}[1,2,1]$. This shows that HHITS may be well represented by a piecewise linear signal.

According to the discussion above, we use PLFME for the cardiac signal. Specifically, in PLFME, the original time series is filtered by the piecewise linear filter A defined in (1) recursively at multiple scales. That is, we define

$$\mathbf{y}^1 := \mathbf{x},\ \mathbf{y}^\tau := \left(I_{n_{\tau-1}} \otimes A\right)\mathbf{y}^{\tau-1},$$

where $n_{\tau-1} := \lfloor \frac{N_{\tau-1}}{4} \rfloor$ and $N_{\tau-1}$ is the length of the time series $\mathbf{y}^{\tau-1}$.

We next present numerical results of PLFME. We compare MSE and PLFME of the time series of consecutive heart beat intervals derived from 20 YOUNG, 20 OLD, 7 AF, and 20 CHF subjects of data lengths $N = t \times 10^4$, where $t = 3,5,8$. The entropy value of each group shown in Fig. 4 is the mean entropy value of the group. MSE values are computed using the software provided in [24]. We see that PLFME improves the robustness of MSE to the data of different lengths.

The most significant difference between the results of MSE and PLFME occurs in the OLD and CHF groups.

We first look at the results of MSE [Fig. 4(a)]. When $N = 3 \times 10^4$, the same data length as used in [5,25], we obtain the same results as those in [5,25]. As N increases, the curve representing entropy values at different scales of the OLD group (black dot) gradually drops below that representing entropy values of the CHF group (blue circle). In particular, the black dot curve is above the blue circle curve at scales 4 to 12 when $N = 5 \times 10^4$. The blue circle curve exceeds the black dot curve at all scales when $N = 8 \times 10^4$, suggesting the cardiac system from the CHF group is more complex than that from the OLD group. It is known that both disease and aging will reduce biological complexity [6]. However, which condition reduces the complexity more remains inconclusive from MSE, which provides different results from data of different lengths.

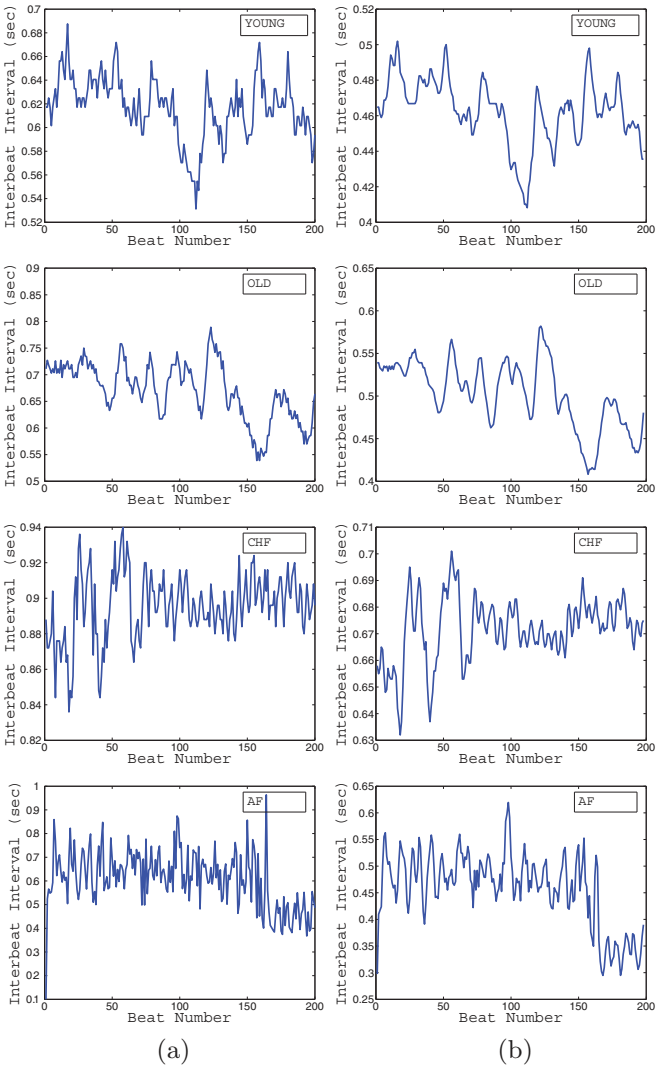
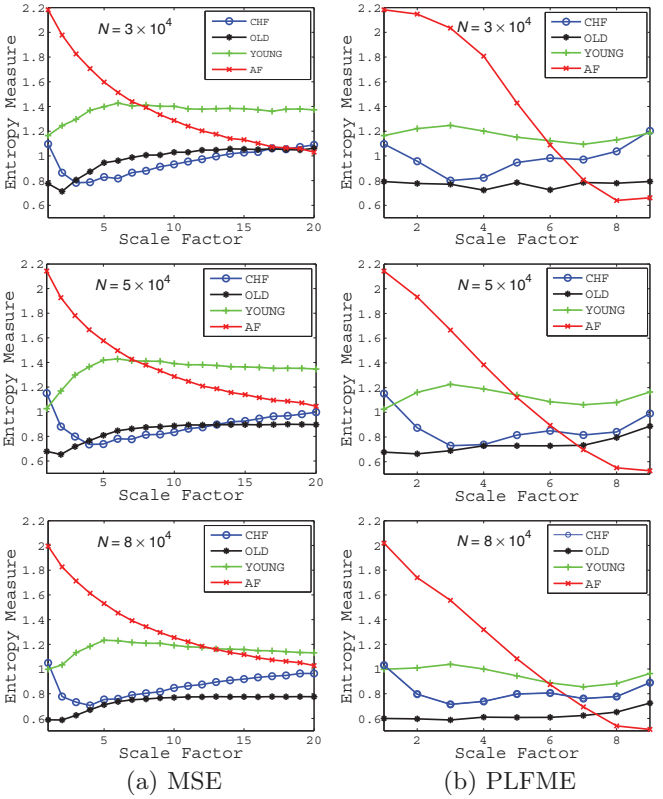


FIG. 3. (Color online) A comparison of the original signal [column (a)] and their piecewise linear representations [column (b)].

Next we look at the results of PLFME. Notice that for data of the OLD and CHF groups of different lengths, the results of FLFME are more robust than those of MSE. For all of the values of N reported in Fig. 4(b), the blue circle curve is above the black dot curve at all scales. This suggests that the cardiac interbeat intervals may lose more complexity from aging than from the CHF group. This result is consistent with that of MSE when $N = 8 \times 10^4$. Moreover, the curves representing entropy values at different scales of the OLD and CHF groups are better separated by PLFME than by MSE. This is further confirmed by the classification result that we present below.

We develop the MSE classifier and PLFME classifier by using entropy values at all scales of MSE and PLFME, respectively, as features to classify the CHF and OLD groups via the support vector machine classifier [26]. In this experiment, we use the data of length $N = 8 \times 10^4$. The training set consists of 22 CHF and 23 OLD subjects and the test set consists of 21



(a) MSE (b) PLFME

FIG. 4. (Color online) MSE and PLFME for longer HHITS data.

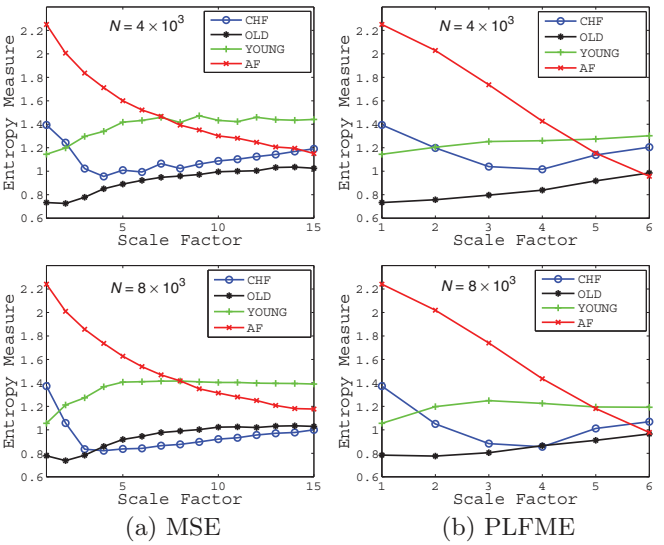
CHF and 23 OLD subjects. The correct classification rates of both methods for the OLD and the CHF subjects are presented in Table I. Clearly, the correct classification rate of PLFME is significantly higher than that of MSE. This may be due to the fact that the piecewise linear filter captures the multiscale information of HHITS more accurately than the piecewise constant filter.

Obtaining long time series may be difficult and expensive in practical applications. Data analysis with short time series is highly desirable. Applications of MSE to short-term physiological recordings were recently studied in [27,28]. We compare the performance of MSE and PLFME for shorter HHITS with data lengths $N = 4,8 \times 10^3$. The numerical results shown in Fig. 5 from the shorter data are consistent with those shown in Fig. 4 from the longer data.

To close this section, we compare PLFME with two refinements of MSE proposed by [29,30]. In [29], the averaging process of MSE was interpreted as the finite-impulse filter (FIR) and a refined MSE (RMSE) was proposed based on the replacement of the FIR filter with a low-pass Butterworth filter, which aims to reduces aliasing when the filtered series are down-sampled. In [30], the adaptive MSE (AMSE) method was proposed by using empirical mode decomposition to

TABLE I. Correct classification rates of the MSE classifier and the PLFME classifier for the OLD subject and the CHF subject.

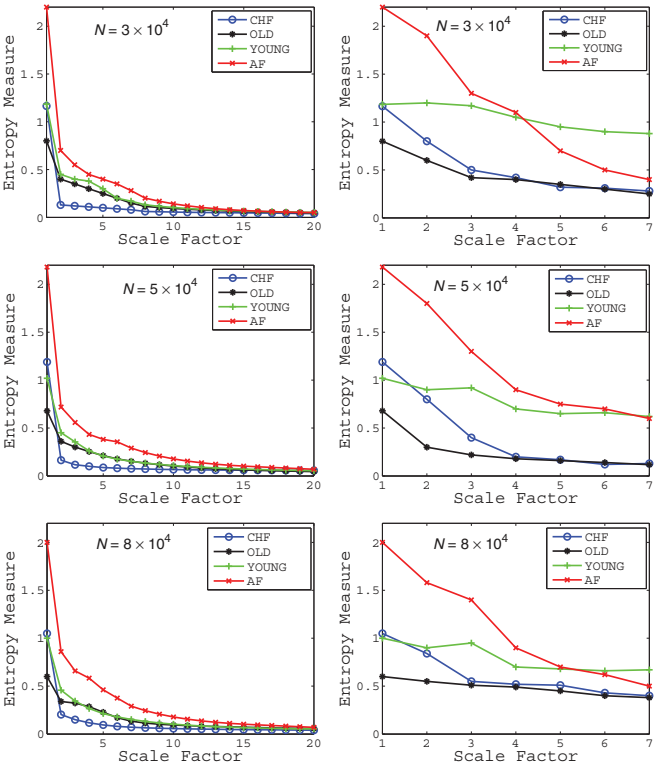
Classifiers	OLD	CHF
MSE	78.3%	76.2%
PLFME	87.0%	81.0%



(a) MSE (b) PLFME

FIG. 5. (Color online) MSE and PLFME for shorter HHITS data.

extract the lower-frequency components of the time series at different scales. We performed both RMSE and AMSE methods on our data (Fig. 6). Both of these methods do not provide satisfactory numerical results. Numerical results of RMSE [Fig. 6(a)] do not give any evidence that cardiac systems from healthy young subjects are more complex than those from pathologic subjects. Numerical results of AMSE [Fig. 6(b)] do not discriminate OLD and CHF groups as well as those from MSE and PLFME, especially when the length of the time series is small. The reason that PLFME and MSE give a better



(a) RMSE (b) AMSE

FIG. 6. (Color online) RMSE and AMSE for HHITS data.

description of the complexity of heartbeat interval time series than RMSE and AMSE may be due to the fact that PLFME and MSE filter the time series more "locally" than RMSE and AMSE.

V. ADAPTIVE FILTERS

When prior knowledge of the system that generates a time series is not available, we propose to use adaptive filters constructed from the time series to compute its FME. Since traditional entropy methods quantify the degree of regularity of a time series by evaluating the appearance of its repetitive patterns, and since consecutive components which are close to each other (measured by r in sample entropy) are considered to be a repetitive pattern, one may consider these repetitive patterns as single units which will present the regularity of the whole system. We further consider the multiscale structure of these patterns to measure the complexity of the physiological system by using adaptive filters. To illustrate the process of constructing an adaptive filter, we present below APCFME as an example. The idea is applicable to constructing adaptive piecewise polynomial filters of order k.

Given a time series, we group it according to its repetitive patterns. Specifically, for a time series $\mathbf{x}$ of length N, we write $\mathbf{x} = \{\mathbf{x}_0, \ldots, \mathbf{x}_{n-1}\}$ with disjoint $\mathbf{x}_j$, being a repetitive pattern of $\mathbf{x}$. Each $\mathbf{x}_j$ sequence consisting of several consecutive components of $\mathbf{x}$ and the distance between any two components in $\mathbf{x}_j$ is not bigger than a preselected tolerance r. For each $\mathbf{x}_j$, we use $|\mathbf{x}_j|$ to denote the number of its elements. A new coarse-grained time series $\mathbf{y}$ of length n is generated by

$$\mathbf{y}(k) := \frac{1}{|\mathbf{x}_k|} \sum_{x(j)\in\mathbf{x}_k} x(j), \quad k \in \mathbb{Z}_n. \tag{14}$$

In other words, $\mathbf{x}$ is filtered by an $n \times n$ block diagonal matrix, whose jth diagonal block is the $1 \times |\mathbf{x}_j|$ matrix,

$$\left[\frac{1}{|\mathbf{x}_j|}, \frac{1}{|\mathbf{x}_j|}, \ldots, \frac{1}{|\mathbf{x}_j|}\right].$$

This *adaptive piecewise constant filter* (APCF) is different from the piecewise constant filter used in MSE. We choose $m = 1$ and an increasing sequence $\{r_0, r_1, \ldots,\}$ at different scales to compute sample entropy. APCFME is then computed by the following procedure:

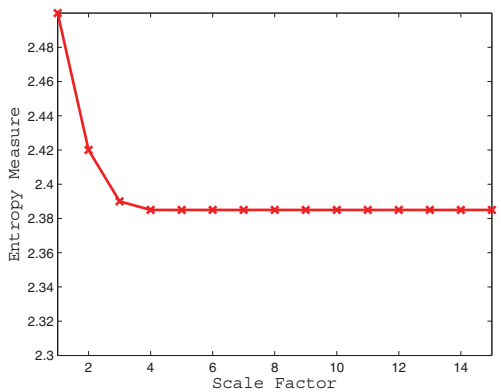
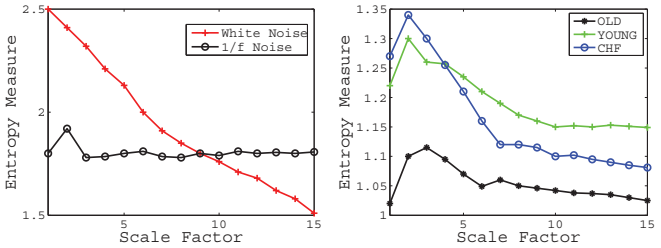


FIG. 7. (Color online) APCFME results for simulated Gaussian white noise when the parameters r_j are constant.



(a) (b)

FIG. 8. (Color online) APCFME results for simulated Gaussian white noise, $1/f$ noise, and HHITS.

(i) Compute $S_1(\mathbf{x}, r_0)$.

(ii) At scale $\tau \geqslant 1$, a new time series $\mathbf{y}^\tau$ is generated from $\mathbf{y}^{\tau-1}$ with $\mathbf{y}^0 := \mathbf{x}$ using APCF. The parameter r used to construct APCF is chosen as $r_{\tau-1}$.

(iii) Compute $S_1(\mathbf{y}^\tau, r_\tau)$.

We now discuss the choice of the parameters r_j used in APCFME. Constructing coarse-grained time series as described in (14) is equivalent to averaging several consecutive components of the original time series among which any two components are close to each other (measured by the tolerance r). Thus the distance between any two components in the coarse-grained time series is potentially bigger than that in the original time series. To further construct the coarse time series, a bigger tolerance in the next coarse-grained procedure is desired. We present the numerical results of APCFME applied to Gaussian white noise when the parameters r_j are chosen as a constant ($r_j = 0.15$) at all scales in Fig. 7. It shows that the entropy value of the coarse-grained time series remains a constant after scale 3, which indicates that the construction of the coarse-grained time series fails when the scale is bigger than 3.

According to the discussion above, we choose parameters $r_0 = 0.15$, $r_{j+1} = 1.1 \times r_j$ for $0 \leqslant j \leqslant 5$ and $r_{j+1} = 1.05 \times r_j$ for $j > 5$ in the numerical results shown in the remaining part of this section. We present in Fig. 8(a) results of APCFME applied to Gaussian white noise and $1/f$ noise. The results are similar to those in Fig. 1. In Fig. 8(b), we present numerical results of APCFME applied to HHITS of length $N = 3 \times 10^4$ for the CHF, YOUNG, and OLD groups. We find that YOUNG is most complex when the scale is bigger than 3. Moreover, CHF is more complex than OLD, which is consistent with the results of PLFME [Fig. 4(b)]. This example shows that APCFME without using any prior information is comparable to PLFME, whose construction uses prior information of HHITS.

VI. CONCLUSION

FME improves MSE in analyzing the complexity of physical and biological systems. It gives a wider range of applicability due to the flexibility of choosing different filters according to data patterns. For HHITS, unlike MSE, which uses piecewise constant filters, PLFME uses the piecewise linear filter (motivated from HRT) and APCFME uses APCF. The results of PLFME are more robust to the data length than those of MSE, and the results of APCFME are consistent with those of PLFME. From both PLFME and APCFME, we find that

aging may reduce the complexity of the cardiac system more than CHF. This finding is consistent for different data lengths.

ACKNOWLEDGMENTS

This work is supported in part by the NSF under Grant No. DMS-1115523, by Guangdong Provincial Government of China through the "Computational Science Innovative Research Team" program, and by the NSFC under Grants No. 11071336 and No. 91130009.

APPENDIX A: PROOFS OF FME RESULTS FOR GAUSSIAN NOISE

In this Appendix, we provide detailed analytical derivations of FME for Gaussian white noise. The analysis mainly relies on the statistical properties of a linear combination of Gaussian random variables. Theorem 4.2.14 in [31] states that a linear combination of two independent real Gaussian random variables is also a real Gaussian random variable. This result can be easily generalized to the following lemma. The proof of the lemma is straightforward and will be omitted.

Lemma 1. If X_j are n independent real Gaussian random variables with mean 0 and standard deviation δ_j, then for $\alpha_j \in \mathbb{R}, j \in \mathbb{Z}_n$, $\sum_{j=0}^{n-1} \alpha_j X_j$ is a real Gaussian random variable with mean 0 and standard deviation $\tilde{\delta} := (\sum_{j=0}^{n-1} \alpha_j^2 \delta_j^2)^{1/2}$.

As usual, the expectation of a random variable X taking values in $\mathbb{R}$ is defined by $E(X) := \int_{\mathbb{R}} tp(t)dt$, where p is the probability density function for the random variable X. Two random variables X and Y are independent if and only if $E(XY) = 0$. We let $\mathrm{cov}(X,Y)$ denote the covariance of X and Y. It is known from Theorem 4.5.3 in [31] that

$$\mathrm{cov}(X,Y) = E(XY) - E(X)E(Y).$$

We let $\mathbf{g} := [g_j, j \in \mathbb{Z}_N]$ denote the real Gaussian random vector. At scale τ, we let $A^\tau \in \mathbb{R}^{p_\tau \times q_\tau}$ be a filter and $n = \lfloor \frac{N}{q_\tau} \rfloor$. We write $\mathbf{g}^\tau := (I_n \otimes A^{(\tau)})\mathbf{g}$ in block form,

$$\mathbf{g}^\tau := \left[\mathbf{g}_0^\tau, \ldots, \mathbf{g}_{n-1}^\tau\right], \tag{A1}$$

where $\mathbf{g}_j^\tau, j \in \mathbb{Z}_n$, is a vector with p_τ components, and for each $k \in \mathbb{Z}_{p_\tau}$ let $[\mathbf{g}_j^\tau]_k := \mathbf{g}^\tau(jp_\tau + k)$. The proof of Proposition 1 is based on the probability density function of $\mathbf{g}_j^\tau$ and the independence of blocks in $\mathbf{g}^\tau$. We present the probability density function of $\mathbf{g}_j^\tau, j \in \mathbb{Z}_n$, in the following lemma.

Lemma 2. If $\mathbf{g}_j^\tau$ is defined in (A1), $j \in \mathbb{Z}_n$, then $\mathbf{g}_j^\tau$ is a p_τ-variate normally distributed random vector with its mean being the zero vector and and its covariance matrix being Σ with

$$\Sigma_{st} := \sum_{k=1}^{p_\tau} A_{sk}^{(\tau)} A_{tk}^{(\tau)} \delta^2, 1 \leqslant s,t \leqslant p_\tau, \tag{A2}$$

where δ is the standard deviation of g_j.

Proof. From the construction of $\mathbf{g}_j^\tau$, for each $j \in \mathbb{Z}_n$, we have that $\mathbf{g}_j^\tau = A^{(\tau)}\mathbf{g}_j$, where $\mathbf{g}_j \in \mathbb{R}^{q_\tau}$ with $[\mathbf{g}_j]_k = \mathbf{g}(q_\tau + k)$ for each $k \in \mathbb{Z}_{q_\tau}$. For each $j \in \mathbb{Z}_n$, $\mathbf{g}_j$ consists of q_τ independent real Gaussian random variables with mean 0 and standard deviation δ, and $[\mathbf{g}_j^\tau]_k$, $k \in \mathbb{Z}_{p_\tau}$, is a linear combination of these random variables. Hence, it follows from Lemma 1 that for each $k \in \mathbb{Z}_{p_\tau}$, $[\mathbf{g}_j^\tau]_k$ is a Gaussian random variable with mean 0 and standard deviation $\sqrt{\sum_{\ell=1}^{p_\tau} (A_{k\ell}^{(\tau)})^2}\delta$. Hence, for any $1 \leqslant s,t \leqslant p_\tau$,

$$\mathrm{cov}\left(\left[\mathbf{g}_j^\tau\right]_s, \left[\mathbf{g}_j^\tau\right]_t\right) = E\left(A_s^{(\tau)}\mathbf{g}_j \cdot A_t^{(\tau)}\mathbf{g}_j\right) = \sum_{k=1}^{p_\tau} A_{sk}^{(\tau)} A_{tk}^{(\tau)} \delta^2,$$

where $A_k^{(\tau)}$ is the kth row of $A^{(\tau)}$ for $1 \leqslant k \leqslant p_\tau$. The desired result follows directly from the definition of the multivariate normal distribution.

Noting that formula (A2) is independent of j, all blocks in $\mathbf{g}^\tau$ have the same probability density function. In the next lemma, we show the independence of blocks in $\mathbf{g}^\tau$. We say that two random variable vectors $\mathbf{X}$ and $\mathbf{Y}$ are independent if the elements of $\mathbf{X}$ (as a collection of random variables) are independent of the elements of $\mathbf{Y}$. Elements within $\mathbf{X}$ or $\mathbf{Y}$ need not be independent when $\mathbf{X}$ and $\mathbf{Y}$ are independent.

Lemma 3. If $\mathbf{g}_j^\tau$ is defined as in (A1), $j \in \mathbb{Z}_n$, then $\mathbf{g}_j^\tau$ and $\mathbf{g}_k^\tau$, $j,k \in \mathbb{Z}_n$, are independent when $j \neq k$.

Proof. It suffices to prove that $[\mathbf{g}_j^\tau]_s$ and $[\mathbf{g}_k^\tau]_t$ are independent for any $s,t \in \mathbb{Z}_{p_\tau}$ if $j \neq k, j,k \in \mathbb{Z}_n$. Since $\mathbf{g}(jp_\tau + s)$ and $\mathbf{g}(kp_\tau + t)$ are independent, we have that $E[\mathbf{g}(jp_\tau + s)\mathbf{g}(kp_\tau + t)] = 0$. It follows that

$$\begin{aligned} &E\left(\left[\mathbf{g}_j^\tau\right]_s\left[\mathbf{g}_k^\tau\right]_t\right) \\ &\quad = E\left[\sum_{\ell=1}^{p_\tau} A_{s\ell}^{(\tau)}\mathbf{g}(jq_\tau + \ell - 1)\sum_{\ell=1}^{p_\tau} A_{t\ell}^{(\tau)}\mathbf{g}(kq_\tau + \ell - 1)\right] \\ &\quad = \sum_{\ell=1}^{p_\tau}\sum_{u=1}^{p_\tau} A_{s\ell}^{(\tau)} A_{tu}^{(\tau)} E[\mathbf{g}(jq_\tau + \ell - 1)\mathbf{g}(kq_\tau + u - 1)] = 0 \end{aligned} \tag{A3}$$

if $j \neq k$. Thus, $[\mathbf{g}_j^\tau]_s$ and $[\mathbf{g}_k^\tau]_t$, $s,t \in \mathbb{Z}_{p_\tau}$, are independent. We then conclude that $\mathbf{g}_j^\tau$ and $\mathbf{g}_k^\tau$ are independent for any $j,k \in \mathbb{Z}_n$ with $j \neq k$.

We next prove Proposition 1. We say that the distance between two blocks $\mathbf{g}_j^\tau$ and $\mathbf{g}_k^\tau$, $j,k \in \mathbb{Z}_n$, is less than r^τ if $[|\mathbf{g}_j^\tau - \mathbf{g}_k^\tau|]_s < r_s^\tau$ for all $s \in \mathbb{Z}_{p_\tau}$ and we write it as $|\mathbf{g}_j^\tau - \mathbf{g}_k^\tau| < r^\tau$.

Proof of Proposition 1. BSE $\tilde{S}_m(\mathbf{g}^\tau, r^\tau)$ is the negative natural logarithm of the conditional probability that the distance between two blocks is less than r provided that the distance between the two preceding blocks is also less than r. We write $\mathbf{g}^\tau$ in block form as in (A1), and from Lemma 3 we know that $\mathbf{g}_j^\tau$ and $\mathbf{g}_k^\tau$ are independent if $j \neq k$. Thus, when $m = 1$ the conditional probability is

$$\begin{aligned} &P\left(\left|\mathbf{g}_j^\tau - \mathbf{g}_k^\tau\right| < r^\tau \middle| \left|\mathbf{g}_{j-1}^\tau - \mathbf{g}_{k-1}^\tau\right| < r^\tau\right) \\ &\quad = \frac{P\left(\left|\mathbf{g}_j^\tau - \mathbf{g}_k^\tau\right| < r^\tau\right) \times P\left(\left|\mathbf{g}_{j-1}^\tau - \mathbf{g}_{k-1}^\tau\right| < r\right)}{P\left(\left|\mathbf{g}_{j-1}^\tau - \mathbf{g}_{k-1}^\tau\right| < r^\tau\right)} \\ &\quad = P\left(\left|\mathbf{g}_j^\tau - \mathbf{g}_k^\tau\right| < r^\tau\right). \end{aligned} \tag{A4}$$

Using this approach recursively, it can be proved that this result is valid for any value of m.

From Lemma 2 and the definition of multivariate normal distribution, we know that the probability density functions of $\mathbf{g}_j^\tau$, $j \in \mathbb{Z}_n$, are all equal to

$$f(\mathbf{x}) := \frac{1}{(2\pi)^{p_\tau/2}|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}\mathbf{x}^T \Sigma^{-1}\mathbf{x}\right\} \tag{A5}$$

for $\mathbf{x} \in \mathbb{R}^{p_\tau}$, where Σ is defined as in (A2). Then the desired formula follows from the definition of BSE, (A4) and (A5). ■

We next present the proof of Proposition 2. Let $A \in \mathbb{R}^{p\times q}$ be a matrix with orthogonal rows and $\tilde{\mathbf{g}} := (I_{\lfloor \frac{N}{q} \rfloor} \otimes A)\mathbf{g}$. We also write $\tilde{\mathbf{g}}$ in block form, $\tilde{\mathbf{g}} := [\tilde{\mathbf{g}}_0, \ldots, \tilde{\mathbf{g}}_{n-1}]$. In this case, a set of independent variables are transformed by A to another set of independent variables due to the orthogonality of the rows of A. In particular, given a real Gaussian random vector $\mathbf{g}$, we have the following result for each block $\tilde{\mathbf{g}}_j$, $j \in \mathbb{Z}_n$.

Lemma 4. If $A \in \mathbb{R}^{p\times q}$ is a matrix with orthogonal rows and $\tilde{\mathbf{g}} := [\tilde{\mathbf{g}}_0, \ldots, \tilde{\mathbf{g}}_{n-1}]$ is defined via a real Gaussian random vector $\mathbf{g}$, then for each $j \in \mathbb{Z}_n$, $\tilde{\mathbf{g}}_j$ consists of p independent Gaussian random variables, with the mean and the standard deviation of $[\tilde{\mathbf{g}}_j]_k$ being 0 and $\sqrt{\sum_{\ell=1}^{p}[A_{k\ell}]^2}\delta$, respectively, for $k \in \mathbb{Z}_p$, where δ is the standard deviation of g_j.

Proof. For each $j \in \mathbb{Z}_n$ and $k \in \mathbb{Z}_p$, we know from Lemma 1 that $[\tilde{\mathbf{g}}_j]_k$ is a real Gaussian variable since it is a linear transformation of real Gaussian variables. The mean and standard deviation of $[\tilde{\mathbf{g}}_j]_k$ are easily obtained from Lemma 1.

It remains to prove that $[\tilde{\mathbf{g}}_j]_s$ and $[\tilde{\mathbf{g}}_j]_t$, $s,t \in \mathbb{Z}_p$, are independent when $s \neq t$ for each $j \in \mathbb{Z}_n$. This follows from the orthogonality of the rows of A. In fact,

$$\begin{aligned} &E([\tilde{\mathbf{g}}_j]_s[\tilde{\mathbf{g}}_j]_t) \\ &= E\left[\sum_{k=1}^{p} A_{sk}\mathbf{g}[jq+(k-1)] \cdot \sum_{k=1}^{p} A_{tk}\mathbf{g}[jq+(k-1)]\right] \\ &= \sum_{k=1}^{p} A_{sk}A_{tk}E\{\mathbf{g}^2[jq+(k-1)]\} = \sum_{k=1}^{p} A_{sk}A_{tk}\delta = 0, \end{aligned}$$

completing the proof of this lemma.

We are now ready to prove Proposition 2.

Proof of Proposition 2. From Lemma 4, we know that the probability density function of $\mathbf{g}_j^\tau$, for each $j \in \mathbb{Z}_n$, is given by

$$f(\mathbf{x}) = \prod_{\ell=1}^{p_\tau} \frac{1}{\sqrt{2\pi\sum_{k=1}^{p} A_{\ell k}^2}\delta} \exp\left(\frac{-x_\ell^2}{2\sum_{k=1}^{p} A_{\ell k}^2\delta^2}\right), \quad \text{(A6)}$$

where $\mathbf{x}$ is a p_τ random vector. By applying (A4) to (A6) and the definition of BSE, we obtain the desired result. ■

Since $\sqrt{\sum_{k=1}^{p} A_{\ell k}^2}\delta$ is actually the standard deviation of $[\tilde{\mathbf{g}}_j]_\ell$ for each block $\tilde{\mathbf{g}}_j$ in the filtered time series, BSE of $\tilde{\mathbf{g}}$ is determined by the standard deviation of $[\tilde{\mathbf{g}}_j]_\ell$ by Proposition 2. This leads us to investigate the standard deviation of the filtered time series at different scales if we apply the filter matrix A recursively to $\mathbf{g}$. At scale τ, we define $\mathbf{g}^\tau$ as in (10) and write it in block form as in (A1). By Proposition 1, at a fixed scale, all blocks have the same probability density function. Let $\delta_\ell^{(\tau)}$ denote the standard deviation of the ℓth element in a block at scale τ. If the matrix A has orthogonal rows and satisfies condition (9), we have the following result for $\delta_\ell^{(\tau)}$:

Lemma 5. If $A \in \mathbb{R}^{p\times q}$ has orthogonal rows and satisfies condition (9), then for $\tau = 2,3,\ldots$, $\delta_k^{(\tau)} = \rho^{\tau-1}\delta$ for $k \in \mathbb{Z}_p$, where ρ is the constant that appears in (9) and δ is the standard deviation of g_j.

Proof. We prove this result by induction on τ. When $\tau = 2$, it follows from Lemma 4 and the fact that A satisfies condition (9) that $\delta_k^{(2)} = \rho\delta$. By the induction hypothesis, we have that $\delta_k^{(\tau-1)} = \rho^{\tau-2}\delta$ for some $\tau \geqslant 3$. Following the computation similar to that used in the proof of Lemma 4 for computing the standard deviation, we obtain that for any $k \in \mathbb{Z}_p$,

$$\delta_k^{(\tau)} = \sqrt{\sum_{t=1}^{q} A_{kt}^2\delta_k^{(\tau-1)}} = \rho\rho^{\tau-2}\delta = \rho^{\tau-1}\delta.$$

This completes the induction and thus the proof.

We next present the proof for Theorem 1.

Proof of Theorem 1. We first reexpress $\tilde{S}_m(\mathbf{g}^\tau, r^\tau)$ at scale τ. For each $s \in \mathbb{Z}_p$, let r_s be defined in (7), x_s be a fixed number, and

$$E(x_s, r_s, \tau) := \operatorname{erf}\left(\frac{x_s + r_s}{\delta_s^{(\tau)}}\right) - \operatorname{erf}\left(\frac{x_s - r_s}{\delta_s^{(\tau)}}\right),$$

where $\delta_s^{(\tau)}$ is the standard deviation of the sth element in a block of $\mathbf{g}^\tau$ at scale τ. In Proposition 2, replacing $\delta(A,s)$ by $\delta_s^{(\tau)}$, we have that

$$\tilde{S}_m(\mathbf{g}^\tau, r^\tau) = -\ln\prod_{s=1}^{p} I_s(\tau),$$

where

$$I_s(\tau) := \frac{1}{\sqrt{2\pi}\delta_s^{(\tau)}} \int_{\mathbb{R}} E(x_s, r_s, \tau) \exp\frac{-x_s^2}{2\left(\delta_s^{(\tau)}\right)^2} dx_s.$$

It remains to prove that for each $s \in \mathbb{Z}_p$, I_s is strictly increasing. By employing Lemma 5 with a change of variable, $y_s = \frac{x_s}{\rho^{\tau-1}}$, we observe that

$$I_s(\tau) = \frac{1}{\sqrt{2\pi}\delta} \int_{\mathbb{R}} \tilde{E}(y_s, r_s, \tau) \exp\frac{-y_s^2}{2\delta^2} dy_s,$$

where

$$\tilde{E}(y_s, r_s, \tau) := \operatorname{erf}\left(\frac{y_s + r_s/\rho^{\tau-1}}{\delta}\right) - \operatorname{erf}\left(\frac{y_s - r_s/\rho^{\tau-1}}{\delta}\right).$$

Since for $0 < \rho < 1$, erf is strictly increasing, we have that $\tilde{E}(y_s, r_s, \tau_1) > \tilde{E}(y_s, r_s, \tau_2)$ when $\tau_1 > \tau_2$. Thus, I_s is strictly increasing. ■

APPENDIX B: PROOFS OF FME RESULTS FOR $1/f$ NOISE

In this Appendix, we provide detailed derivations for Propositions 3 and 4 regarding $1/f$ noise. We let $\mathbf{f}$ denote the $1/f$ noise and assume that the length of $\mathbf{f}$ is 2^N for a positive integer N.

We first prove Proposition 3. Let $A := [\alpha, \beta]$ for $\alpha, \beta \in \mathbb{R}$ and $\mathbf{f}_A := (I_{2^{N-1}} \otimes A)\mathbf{f}$. From the definition of $1/f$ noise, we need to consider the Fourier transform of the filtered signal $\hat{\mathbf{f}_A} := F_{N-1}(I_{2^{N-1}} \otimes A)\mathbf{f}$. Since $\mathbf{f} = F_N^{-1}F_N\mathbf{f}$, $\hat{\mathbf{f}_A}$ may be rewritten as

$$\hat{\mathbf{f}_A} = F_{N-1}(I_{2^{N-1}} \otimes A)F_N^{-1}F_N\mathbf{f},$$

where F_N^{-1} is the inverse discrete Fourier transform which has the form $F_N^{-1} := [e^{i\theta_N k\ell} : k, \ell \in \mathbb{Z}_{2^N}]$. We shall express the Fourier transform of the filtered signal in terms of the Fourier transform of the original signal. To this end, we investigate the matrix

$$\tilde{A} := F_{N-1}(I_{2^{N-1}} \otimes A)F_N^{-1}. \quad \text{(B1)}$$

In the next lemma, we express $\tilde{A}$ in terms of two diagonal matrices

$$D^+ := \mathrm{diag}[d_k^+ : k \in \mathbb{Z}_{2^{N-1}}], \text{ where } d_k^+ := \alpha + \beta e^{i\theta_N k},$$

and D^- whose definition is obtained from that of D^+ by replacing the "+" sign by the "−" sign.

Lemma 6. For any positive integer N, there holds

$$\tilde{A} = [D^+, D^-]. \tag{B2}$$

Proof. From the definition of F_N^{-1} and A, for $\ell \in \mathbb{Z}_{2^{N-1}}$, $k \in \mathbb{Z}_{2^N}$ it is straightforward to compute that

$$\left[(I_{2^{N-1}} \otimes A)F_N^{-1}\right]_{\ell,k} = e^{i\theta_{N-1}\ell k}(\alpha + \beta e^{i\theta_N k}). \tag{B3}$$

Letting $D := \mathrm{diag}[D^+, D^-]$ and $\tilde{F}_N := [\alpha F_{N-1}^{-1}, \beta F_{N-1}^{-1}]$, it follows from (B3) that

$$(I_{2^{N-1}} \otimes A)F_N^{-1} = \tilde{F}_N D. \tag{B4}$$

From (B4) we have that

$$\tilde{A} = [I_{N-1}, I_{N-1}]D. \tag{B5}$$

Formula (B2) is then obtained by substituting the expression of D into (B5).

The discrete Fourier transform of a real vector is a complex vector. When analyzing $1/f$ noise, it is convenient to separate the real and imaginary parts of the discrete Fourier transform of a signal. Considering the symmetry property of the discrete Fourier transform, we define two operators T_1 and T_2. The operator T_1 projects a vector of length 2^N to the vector of length 2^{N-1} consisting of the first 2^{N-1} components of the original vector, and the operator T_2 projects a vector of length 2^N to the vector of length 2^{N-1} consisting of the last 2^{N-1} components of the original vector. For a real vector $\mathbf{x}$ of length 2^N, we write its discrete Fourier transform as $\hat{\mathbf{x}} := \tilde{\mathbf{x}} + i\tilde{\mathbf{y}}$, where $\tilde{\mathbf{x}}$ and $\tilde{\mathbf{y}}$ are two real vectors of length 2^N. We let

$$\tilde{\mathbf{x}}_1 := T_1\tilde{\mathbf{x}}, \quad \tilde{\mathbf{x}}_2 := T_2\tilde{\mathbf{x}}, \quad \tilde{\mathbf{y}}_1 := T_1\tilde{\mathbf{y}}, \quad \tilde{\mathbf{y}}_2 := T_2\tilde{\mathbf{y}}.$$

In the next lemma, we express the discrete Fourier transform $\hat{\mathbf{x}}_A$ of $\mathbf{x}_A := (I_{2^{N-1}} \otimes A)\mathbf{x}$ in terms of $\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2, \tilde{\mathbf{y}}_1$, and $\tilde{\mathbf{y}}_2$. To simplify the notation, we introduce three diagonal matrices:

$$A^+ := \mathrm{diag}[\alpha + \beta\cos(\theta_N k) : k \in \mathbb{Z}_{2^{N-1}}],$$

A^-, whose definition is obtained from that of A^+ by replacing the "+" sign by the "−" sign, and $B = \mathrm{diag}[\beta\sin(\theta_N k) : k \in \mathbb{Z}_{2^{N-1}}]$. Moreover, we denote by $\mathrm{Re}(\hat{\mathbf{x}}_A)$ and $\mathrm{Im}(\hat{\mathbf{x}}_A)$ the real part and the imaginary part of $\hat{\mathbf{x}}_A$, respectively.

Lemma 7. If $\mathbf{x}$ is a real vector of length 2^N, then

$$\mathrm{Re}(\hat{\mathbf{x}}_A) = A^+\tilde{\mathbf{x}}_1 + A^-\tilde{\mathbf{x}}_2 - B\tilde{\mathbf{y}}_1 + B\tilde{\mathbf{y}}_2$$

and

$$\mathrm{Im}(\hat{\mathbf{x}}_A) = A^+\tilde{\mathbf{y}}_1 + A^-\tilde{\mathbf{y}}_2 + B\tilde{\mathbf{x}}_1 - B\tilde{\mathbf{x}}_2.$$

Proof. According to the definition of $\hat{\mathbf{x}}$, $\hat{\mathbf{x}}_A$, and $\mathbf{x}_A$, we have that

$$\hat{\mathbf{x}}_A = F_{N-1}A\mathbf{x} = F_{N-1}AF_N^{-1}\hat{\mathbf{x}} = \tilde{A}\hat{\mathbf{x}}. \tag{B6}$$

Applying Lemma 6 to (B6) yields that $\hat{\mathbf{x}}_A = (D^+, D^-)\hat{\mathbf{x}}$. We partition the real and imaginary parts of the vector $\hat{\mathbf{x}}$ as

$$\hat{\mathbf{x}} = \begin{pmatrix} \tilde{\mathbf{x}}_1 \\ \tilde{\mathbf{x}}_2 \end{pmatrix} + i\begin{pmatrix} \tilde{\mathbf{y}}_1 \\ \tilde{\mathbf{y}}_2 \end{pmatrix}. \tag{B7}$$

Notice that the definition of D^+, D^-, A^+, A^-, and B gives us $D^+ = A^+ + iB$ and $D^- = A^- - iB$, which together with (B7) are substituted into the above formula of $\hat{\mathbf{x}}_A$ yielding the desired result.

The next lemma states that the independence of components of $\hat{\mathbf{x}}$ determines the independence of components of $\hat{\mathbf{x}}_A$.

Lemma 8. Suppose that $\mathbf{x}$ is a real random vector of length 2^N. If the first $2^{N-1}+1$ components of $\hat{\mathbf{x}}$ are independent, then the first $2^{N-2}+1$ components of $\hat{\mathbf{x}}_A$ are independent.

Lemma 8 is a straightforward extension of a known result, Lemma 4.3 of [9]. We thus omit the proof.

We next recall a known result that describes the statistical property of the linear combinations of two independent complex Gaussian random variables. Its proof can be found in [9].

Lemma 9. If z_1 and z_2 are independent complex Gaussian random variables with mean 0 and standard derivation δ_1 and δ_2, respectively, then for each pair of complex numbers $a := a_1 + ia_2, b := b_1 + ib_2$ with $a_1, a_2, b_1, b_2 \in \mathbb{R}$, $az_1 + bz_2$ is a complex Gaussian random variable with mean 0 and standard derivation

$$\delta := \left(a_1^2\delta_1^2 + a_2^2\delta_1^2 + b_1^2\delta_2^2 + b_2^2\delta_2^2\right)^{1/2}. \tag{B8}$$

In the next lemma, we describe $\hat{\mathbf{x}}_A$ in terms of $\hat{\mathbf{x}}$.

Lemma 10. If the first $2^{N-1}+1$ components of $\hat{\mathbf{x}}$ are independent complex Gaussian random variables, then the first $2^{N-2}+1$ components of $\hat{\mathbf{x}}_A$ are independent complex [except the first and the $(2^{N-2}+1)$th components, which are real] Gaussian random variables. Moreover, for each $k \in \mathbb{Z}_{2^{N-1}+1}$, if the mean of the $(k+1)$th component of $\hat{\mathbf{x}}$ is 0 and its standard deviation is δ_k, then the mean of each component of $\hat{\mathbf{x}}_A$ is 0 and the standard deviation of the $(k+1)$th component of $\hat{\mathbf{x}}_A$ is

$$\delta_{A,k} = \left[(\alpha^2 + \beta^2 + \gamma_k)\delta_k^2 + (\alpha^2 + \beta^2 - \gamma_k)\delta_{2^{N-1}-k}^2\right]^{1/2}, \tag{B9}$$

where $\gamma_k := 2\alpha\beta\cos(\theta_N k)$.

Proof. The description of the components of $\hat{\mathbf{x}}_A$ relies on Lemma 7. We write $\hat{\mathbf{x}} := [z_k : k \in \mathbb{Z}_{2^N}]$ and $\hat{\mathbf{x}}_A := [z_{A,k} : k \in \mathbb{Z}_{2^{N-1}}]$. We first prove that $z_{A,0}$ and $z_{A,2^{N-2}}$ are real Gaussian random variables. By Lemma 7, we have that

$$z_{A,0} = \alpha(z_0 + z_{2^{N-1}}). \tag{B10}$$

Since z_0 and $z_{2^{N-1}}$ are real Gaussian random variables, from (B10) and Lemma 1 we know that $z_{A,0}$ is a real Gaussian random variable. Noting $2^{N-1} - 2^{N-2} = 2^{N-2}$, it also follows from Lemma 7 that

$$z_{A,2^{N-2}} = \alpha(x_{2^{N-2}} - y_{2^{N-2}}), \tag{B11}$$

where $x_{2^{N-2}}$ and $y_{2^{N-2}}$ are, respectively, the real part and the imaginary part of $z_{2^{N-2}}$. Hence, we conclude that $z_{A,2^{N-2}}$ is also a real Gaussian random variable. By Lemma 7, we have that

$$z_{A,k} = (\alpha + \beta e^{i\theta_N k})z_k + (\alpha - \beta e^{i\theta_N k})z_{2^{N-1}+k}. \tag{B12}$$

Thus, by Lemma 9, $z_{A,k}$ is a complex Gaussian random variable. The independence of the random variables $z_{A,k}$,

$k \in \mathbb{Z}_{2^{N-2}+1}$, is ensured by Lemma 8 from the fact that the random variables z_k, $k \in \mathbb{Z}_{2^{N-1}+1}$, are independent. By applying Lemma 9 to Eq. (B12) with the symmetric property (12) and Lemma 1 to Eqs. (B10) and (B11), we obtain (B9).

Now we are ready to prove Proposition 3.

Proof of Proposition 3. We denote by $\hat{A}\mathbf{f}$ the discrete Fourier transform of $A\mathbf{f}$. To prove $\hat{A}\mathbf{f}$ is a $1/f$ random vector, we need to show that the first $2^{N-2}+1$ elements of $\hat{A}\mathbf{f}$ are independent Gaussian random variables with mean 0, and for all $k \in \mathbb{Z}_{2^{N-2}+1}$, $\delta_{A,k} \leqslant \frac{c'}{1+k}$ for some constant c'.

Since $\mathbf{f}$ is a $1/f$ random vector, from the definition of a $1/f$ random vector we know that the first $2^{N-1}+1$ elements of $\hat{\mathbf{f}}$ are independent Gaussian random variables with mean 0. Thus, from Lemma 10, we have that the first $2^{N-2}+1$ elements of $\hat{A}\mathbf{f}$ are independent Gaussian random variables with mean 0.

We next show that for all $k \in \mathbb{Z}_{2^{N-1}+1}$, $\delta_{A,k} \leqslant \frac{c'}{1+k}$ for some constant c'. Let $\gamma_k^+ := \alpha^2+\beta^2+2\alpha\beta\cos(\theta_N k)$ and define γ_k^- by replacing the second "+" in the definition of γ_k^+ with "−." When $\alpha+\beta \neq 0$, we have that $\gamma_k^+ \neq 0$ and $\gamma_k^- \neq 0$ for all $k \in \mathbb{Z}_{2N-1+1}$. Thus, from (B9), we have for all $k \in \mathbb{Z}_{2^{N-2}+1}$ that

$$\begin{aligned}\delta_{A,k}^2 &= \gamma_k^+\delta_k^2 + \gamma_k^-\delta_{2^{N-1}-k}^2 \\ &\leqslant \frac{c}{1+k}\,\frac{(1+2^{N-1}-k)\gamma_k^+ + (1+k)\gamma_k^-}{1+2^{N-1}-k}. \end{aligned} \tag{B13}$$

Note that $k \in \mathbb{Z}_{2^{N-2}+1}$ and thus $0 \leqslant \cos(\theta_N k) \leqslant 1$. Then the second fraction in the last term of formula (B13) is an increasing function of k, which has the maximum value $2(\alpha^2+\beta^2)$ when $k = 2^{N-2}$. This gives the desired estimate. ■

Finally, we prove Proposition 4. We first recall a known fact of the discrete Fourier transform, whose proof may be found in [32] (Lemma 2.37).

Lemma 11. If $M \in \mathbb{N}$, $N = 2M$, $\mathbf{z}$ is a vector of length N, and $\mathbf{u}$, $\mathbf{v}$ are vectors of length M defined by $u_k := z_{2k}$ and $v_k := z_{2k+1}$ for $k \in \mathbb{Z}_M$, then for $m \in \mathbb{Z}_M$,

$$\hat{z}_m = \hat{u}_m + e^{-2\pi\frac{m}{N}}\hat{v}_m,\ \ \hat{z}_{m+M} = \hat{u}_m - e^{-2\pi i\frac{m+M}{N}}\hat{v}_m.$$

From Lemma 11, we shall show in the next lemma that the vector obtained by interlacing two $1/f$ random vectors is again a $1/f$ random vector.

Lemma 12. If $\mathbf{g}$ and $\mathbf{h}$ are $1/f$ random vectors of length n, then the random vector $\mathbf{f}$ defined by

$$\mathbf{f} := [g_0, h_0, g_1, h_1, \ldots, g_{n-1}, h_{n-1}]$$

is also a $1/f$ random vector.

Proof. Since $\mathbf{g}$ and $\mathbf{h}$ are $1/f$ random vectors, by the definition of the $1/f$ random vector we know that $\hat{g}_k$ and $\hat{h}_k$, $k \in \mathbb{Z}_n$, are independent Gaussian random variables with mean 0 and there exist positive constants c, c' such that their standard deviations $\delta_{g,k}$ and $\delta_{h,k}$ satisfy

$$\delta_{g,k} \leqslant \frac{c}{1+k} \ \text{ and } \ \delta_{h,k} \leqslant \frac{c'}{1+k} \tag{B14}$$

for any $k \in \mathbb{Z}_n$.

It follows from Lemma 11 that for $m \in \mathbb{Z}_n$, $\hat{f}_m = \hat{g}_m + e^{-2\pi i m/N}\hat{h}_m$, and for $m = n, n+1, n+2, \ldots, 2n-1$, $\hat{f}_m = \hat{g}_{m-n} - e^{-2\pi i m/N}\hat{h}_{m-n}$. By applying Lemma 9 to the formulas for $\hat{f}_m$ and using conditions (B14), we know for $m \in \mathbb{Z}_{2n}$ that $\hat{f}_m$ is a Gaussian random vector with mean 0 and the standard deviation of $\hat{f}_m$, $\delta_{f,m}$, satisfies $\delta_{f,m} \leqslant \frac{c''}{1+m}$, where $c'' = \sqrt{(1+n)[2c^2+(c')^2]}$. The independence of $\hat{f}_m$, $m \in \mathbb{Z}_{2n}$, can be proved by using a similar computation used in the proof of Lemma 4 for proving the independence considering that $[1, e^{-2\pi i m/N}]$ and $[1, -e^{-2\pi i m/N}]$ are orthogonal. Therefore, $\mathbf{f}$ is a $1/f$ random vector. ■

We next recall a known fact that the sum of two $1/f$ random vectors is also a $1/f$ random vector. Its proof can be found in [33].

Lemma 13. If $\mathbf{f}_1$ and $\mathbf{f}_2$ are $1/f$ random vectors, then $\mathbf{f}_1 + \mathbf{f}_2$ is also a $1/f$ random vector.

We are now ready to prove Proposition 4. The filtered $1/f$ random vector through the piecewise linear filter A defined in (1) is again a $1/f$ random vector. Here, A is referred to as the linear filter defined in (1).

Proof of Proposition 4. Let A_1 and A_2 denote the matrices formed, respectively, by the first row and the second row of A. Since $(I_{2^{N-2}} \otimes A)\mathbf{f}$ may be obtained by interlacing $(I_{2^{N-2}} \otimes A_1)\mathbf{f}$ and $(I_{2^{N-2}} \otimes A_2)\mathbf{f}$, according to Lemma 12, it suffices to prove that $(I_{2^{N-2}} \otimes A_1)\mathbf{f}$ and $(I_{2^{N-2}} \otimes A_2)\mathbf{f}$ are $1/f$ random vectors.

Let $\mathbf{f}_1 := [f_0, f_2, \ldots, f_{N-2}]$, $\mathbf{f}_2 := [f_0, f_1, f_4, f_5, \ldots, f_{N-4}, f_{N-3}]$, and $\mathbf{f}_3 := [f_2, f_3, f_6, f_7, \ldots, f_{N-2}, f_{N-1}]$. Since $\mathbf{f}$ is a $1/f$ random vector, it follows from Lemma 12 and Corollary 1 that $\mathbf{f}_1$, $\mathbf{f}_2$, and $\mathbf{f}_3$ are $1/f$ random vectors. By Proposition 3, $(I_{2^{N-2}} \otimes [1/2, 1/2])\mathbf{f}_1$ is a $1/f$ random vector. Since $(I_{2^{N-2}} \otimes A_1)\mathbf{f} = (I_{2^{N-2}} \otimes [1/2, 1/2])\mathbf{f}_1$, we conclude that $(I_{2^{N-2}} \otimes A_1)\mathbf{f}$ is a $1/f$ random vector. Note that

$$\begin{aligned}(I_{2^{N-2}} \otimes A_2)\mathbf{f} &= (I_{2^{N-2}} \otimes [-\sqrt{3}/4, 1/4])\mathbf{f}_2 \\ &\quad + (I_{2^{N-2}} \otimes [\sqrt{3}/4, 1/4])\mathbf{f}_3. \end{aligned} \tag{B15}$$

By Proposition 3, both $(I_{2^{N-2}} \otimes [-\sqrt{3}/4, 1/4])\mathbf{f}_2$ and $(I_{2^{N-2}} \otimes [\sqrt{3}/4, 1/4])\mathbf{f}_3$ are $1/f$ random vectors. Hence, by Lemma 13 and (B15), $(I_{2^{N-2}} \otimes A_2)\mathbf{f}$ is a $1/f$ random vector. ■

[1] S. M. Pincus, Ann. N.Y. Acad. Sci. **954**, 245 (2001).

[2] A. Porta, S. Guzzetti, N. Montano, R. Furlan, M. Pagani, A. Malliani, and S. Cerutti, IEEE Trans. Biomed. Eng. **48**, 1282 (2001).

[3] M. Costa, A. L. Goldberger, and C.-K. Peng, Phys. Rev. Lett. **89**, 068102 (2002).

[4] M. Costa, C.-K. Peng, A. L. Goldberger, and J. M. Hausdorff, Physica A **330**, 53 (2003).

[5] M. Costa, A. L. Goldberger, and C.-K. Peng, Phys. Rev. E **71**, 021906 (2005).

[6] A. L. Goldberger, C.-K. Peng, L. A. Lipsitz, D. E. Vaillancourt, and K. M. Newell, Neurobiol. Aging **23**, 27 (2002).

[7] S. M. Pincus, Proc. Natl. Acad. Sci. (USA) **88**, 2297 (1991).
[8] J. S. Richman and J. R. Moorman, Am. J. Physiol. **278**, H2039 (2000).
[9] Y. Jiang, C.-K. Peng, and Y. Xu, J. Comput. Appl. Math. **236**, 728 (2011).
[10] G. Schmidt, M. Malik, P. Barthel, R. Schneider, K. Ulm, L. Rolnitzky, A. J. Camm, J. T. Bigger, Jr., and A. Schömig, The Lancet **353**, 1390 (1999).
[11] C. A. Micchelli and Y. Xu, Appl. Comput. Harmon. Anal. **1**, 391 (1994).
[12] C. A. Micchelli and Y. Xu, Multidimens. Systems Signal Process. **8**, 31 (1997).
[13] Q. Lian, L. Shen, Y. Xu, and L. Yang, Appl. Anal. **90**, 1299 (2011).
[14] V. V. Nikulin and T. Brismar, Phys. Rev. Lett. **92**, 089803 (2004).
[15] H. Atmanspacher and H. Scheingraber, *Information Dynamics* (Plenum, New York, 1991), Vol. 1990.
[16] J. Berntgen, K. Heime, W. Daumann, U. Auer, F.-J. Tegude, and A. Matulionis, IEEE Trans. Electron Devices **46**, 194 (1999).
[17] T. Antal, M. Droz, G. Györgyi, and Z. Rácz, Phys. Rev. E **65**, 046140 (2002).
[18] Y.-C. Zhang, J. Phys. I France **1**, 971 (1991).
[19] H. C. Fogedby, J. Stat. Phys. **69**, 411 (1992).
[20] A. Bauer, M. Malik, G. Schmidt, P. Barthel, H. Bonnemeier, I. Cygankiewicz, P. Guzik, F. Lombardi, A. Muller, A. Oto, R. Schneider, M. Watanabe, D. Wichterle, and W. Zareba, J. Am. Coll. Cardiol. **52**, 1353 (2008).
[21] M. A. Watanabe, J. E. Marine, R. Sheldon, and M. E. Josephson, Circulation **106**, 325 (2002).
[22] P. Guzik and G. Schmidt, Cardiac Electrophysiol. Rev. **6**, 256 (2002).
[23] A. Berkowitsch, N. Guettler, T. Neumann, D. Vukajlovic, B. Schulte, J. Neuzner, and H. Pitschner, Eur. Heart J. **22**(suppl), P2941 (2001).
[24] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, Ch. R. G. Mark, J. E. Mietus, G. B. Moody, C. K. Peng, and H. E. Stanley, Circulation **101**, e215 (2000).
[25] M. Costa, A. L. Goldberger, and C.-K. Peng, Phys. Rev. Lett. **92**, 089804 (2004).
[26] J. Suykens and J. Vandewalle, Neural Proc. Lett. **9**, 293 (1999).
[27] L. Angelini, R. Maestri, D. Marinazzo, L. Nitti, M. Pellicoro, G. Pinna, S. Stramaglia, and S. Tupputi, Artificial Intell. Med. **41**, 237 (2007).
[28] Z. Trunkvalterova, M. Javorka, I. Tonhajzerova, J. Javorkova, Z. Lazarova, K. Javorka, and M. Baumert, Physiol. Meas. **29**, 817 (2008).
[29] J. F. Valencia, A. Porta, M. Vallverdú, F. Claria, R. Baranowski, E. Orlowska-Baranowska, and P. Caminal, IEEE Trans. Biomed. Eng. **56**, 2202 (2009).
[30] M. Hu and H. Liang, IEEE Trans. Biomed. Eng. **59**, 12 (2012).
[31] G. Casella and R. Berger, *Statistical Inference* (Duxbury Resource Center, Belmont, CA, 2001).
[32] M. Frazier, *An Introduction to Wavelets through Linear Algebra* (Springer, New York, 1999).
[33] H. Takayasu, *Fractals in Physical Sciences* (Manchester University Press, Manchester, 1990).