Energy Propagation in Deep Convolutional Neural Networks

Thomas Wiatowski, Philipp Grohs, and Helmut Bölcskei, Fellow, IEEE

Abstract—Many practical machine learning tasks employ very deep convolutional neural networks. Such large depths pose formidable computational challenges in training and operating the network. It is therefore important to understand how fast the energy contained in the propagated signals (a.k.a. feature maps) decays across layers. In addition, it is desirable that the feature extractor generated by the network be informative in the sense of the only signal mapping to the all-zeros feature vector being the zero input signal. This “trivial null-space” property can be accomplished by asking for “energy conservation” in the sense of the energy in the feature vector being proportional to that of the corresponding input signal. This paper establishes conditions for energy conservation (and thus for a trivial null-space) for a wide class of deep convolutional neural network-based feature extractors and characterizes corresponding feature map energy decay rates. Specifically, we consider general scattering networks employing the modulus non-linearity and we find that under mild analyticity and high-pass conditions on the filters (which encompass, inter alia, various constructions of Weyl-Heisenberg filters, wavelets, ridgelets, ($\alpha$)-curvelets, and shearlets) the feature map energy decays at least polynomially fast. For broad families of wavelets and Weyl-Heisenberg filters, the guaranteed decay rate is shown to be exponential. Moreover, we provide handy estimates of the number of layers needed to have at least $((1 - \varepsilon) \cdot 100)\%$ of the input signal energy be contained in the feature vector.

Index Terms—Machine learning, deep convolutional neural networks, scattering networks, energy decay and conservation, frame theory.

I. INTRODUCTION

FEATURE extraction based on deep convolutional neural networks (DCNNs) has been applied with significant success in a wide range of practical machine learning tasks [1]–[6]. Many of these applications, such as, e.g., the classification of images in the ImageNet data set, employ very deep networks with potentially hundreds of layers [7]. Such network depths entail formidable computational challenges in the training phase due to the large number of parameters to be learned, and in operating the network due to the large number of convolutions that need to be carried out. It is therefore paramount to understand how fast the energy contained in the signals generated in the individual network layers, a.k.a. feature maps, decays across layers. In addition, it is important that the feature vector—obtained by aggregating filtered versions of the feature maps—be informative in the sense of the only signal mapping to the all-zeros feature vector being the zero input signal. This “trivial null-space” property for the feature extractor can be obtained by asking for the energy in the feature vector being proportional to that of the corresponding input signal, a property we shall refer to as “energy conservation”.

Scattering networks as introduced in [8] and extended in [9] constitute an important class of feature extractors based on nodes that implement convolutional transforms with pre-specified or learned filters in each network layer (e.g., wavelets [8], [10], uniform covering filters [11], or general filters [9]), followed by a non-linearity (e.g., the modulus [8], [10], [11], or a general Lipschitz non-linearity [9]), and a pooling operation (e.g., sub-sampling or average-pooling [9]). Scattering network-based feature extractors were shown to yield classification performance competitive with the state-of-the-art on various data sets [12]–[17]. Moreover, a mathematical theory exists, which allows to establish formally that such feature extractors are—under certain technical conditions—horizontally [8] or vertically [9] translation-invariant and deformation-stable in the sense of [8], or exhibit limited sensitivity to deformations in the sense of [9] on input signal classes such as band-limited functions [9], [18], cartoon functions [19], and Lipschitz functions [19].

It was shown recently that the energy in the feature maps generated by scattering networks employing, in every network layer, the same set of (certain) Parseval wavelets [10] Section 5] or “uniform covering” [11] filters (both satisfying analyticity and vanishing moments conditions), the modulus non-linearity, and no pooling, decays at least exponentially fast and “strict” energy conservation (which, in turn, implies a trivial null-space) for the infinite-depth feature vector holds. Specifically, the feature map energy decay was shown to be at least of order $O(a^{-N})$, for some unspecified $a > 1$, where $N$ denotes the network depth. We note that $d$-dimensional uniform covering filters as introduced in [11] are functions whose Fourier transforms’ support sets can be covered by a union of finitely many balls. This covering condition is satisfied by, e.g., Weyl-Heisenberg filters [21] with a band-limited prototype function, but fails to hold for multi-scale filters such as wavelets [22], [23], ($\alpha$)-curvelets [24], [26], shearlets [27], [28], or ridgelets [29]–[31], see [11] Remark 2.2 (b).

Contributions. The first main contribution of this paper is a characterization of the feature map energy decay rate in
DCNNs employing the modulus non-linearity, no pooling, and general filters that constitute a frame\cite{22,32,34}, but not necessarily a Parseval frame, are allowed to be different in different network layers. We find that, under mild analyticity and high-pass conditions on the filters, the energy decay rate is at least polynomial in the network depth, i.e., the decay is at least of order $O(N^{-\alpha})$, and we explicitly specify the decay exponent $\alpha > 0$. This result encompasses, inter alia, various constructions of Weyl-Heisenberg filters, wavelets, ridgelets, $(\alpha)$-curvettes, shearlets, and learned filters (of course as long as the learning algorithm imposes the analyticity and high-pass conditions we require). For broad families of wavelets and Weyl-Heisenberg filters, the guaranteed energy decay rate is shown to be exponential in the network depth, i.e., the decay is at least of order $O(a^{-N})$ with the decay factor given as $a = \frac{\pi}{2}$ in the wavelet case and $a = \frac{\pi}{2}$ in the Weyl-Heisenberg case. We hasten to add that our results constitute guaranteed decay rates and do not preclude the energy from decaying faster in practice.

Our second main contribution shows that the energy decay results above are compatible with a trivial null-space for finite- and infinite-depth networks. Specifically, this is accomplished by establishing energy proportionality between the feature vector and the underlying input signal with the proportionality constant lower- and upper-bounded by the frame bounds of the filters employed in the different layers. We show that this energy conservation result is a consequence of a demodulation effect induced by the modulus non-linearity in combination with the analyticity and high-pass properties of the filters. Specifically, in every network layer, the modulus non-linearity moves the spectral content of each individual feature map to base-band (i.e., to low frequencies), where it is subsequently extended (i.e., fed into the feature vector) by a low-pass output-generating filter.

Finally, for input signals that belong to the class of Sobolev functions\cite{1}, our energy decay and conservation results are shown to yield handy estimates of the number of layers needed to have at least $(1 - \varepsilon) \cdot 100\%$ of the input signal energy be contained in the feature vector. For example, in the case of exponential energy decay with $a = \frac{\pi}{2}$ and for band-limited input signals, only 8 layers are needed to absorb 95\% of the input signal’s energy.

We emphasize that throughout energy decay results pertain to the feature maps, whereas energy conservation statements apply to the feature vector, obtained by aggregating filtered versions of the feature maps.

Notation. The complex conjugate of $z \in \mathbb{C}$ is denoted by $\bar{z}$. We write $\text{Re}(z)$ for the real, and $\text{Im}(z)$ for the imaginary part of $z \in \mathbb{C}$. The Euclidean inner product of $x, y \in \mathbb{C}$ is $\langle x, y \rangle := \sum_{i=1}^{d} x_i y_i$, with associated norm $\|x\| := \sqrt{\langle x, x \rangle}$. For $x \in \mathbb{R}$, $\langle x \rangle := \max\{0, x\}$ and $\langle x \rangle := (1 + |x|^2)^{1/2}$. We denote the open ball of radius $r > 0$ centered at $x \in \mathbb{R}^d$ by $B_r(x) \subseteq \mathbb{R}^d$. The first canonical orthonth is $H := \{ x \in \mathbb{R}^d \mid x_k \geq 0, k = 1, \ldots, d \}$, and we define the rotated orthonth $H_A := \{ Ax \mid x \in H \}$ for $A \in O(d)$, where $O(d)$ stands for the orthogonal group of dimension $d \in \mathbb{N}$. The Minkowski sum of sets $A, B \subseteq \mathbb{R}^d$ is $A + B := \{ a + b \mid a \in A, b \in B \}$, and $A \Delta B := (A \setminus B) \cup (B \setminus A)$ denotes their symmetric difference. A Lipschitz domain $D$ is a set $D \subseteq \mathbb{R}^d$ whose boundary $\partial D$ is “sufficiently regular” to be thought of locally as the graph of a Lipschitz-continuous function; for a formal definition we refer to \cite{39} Definition 1.40. A multi-index $\alpha = (\alpha_1, \ldots, \alpha_d) \in \mathbb{N}_0^d$ is an ordered $d$-tuple of non-negative integers $\alpha_i \in \mathbb{N}_0$.

For functions $W : \mathbb{N} \rightarrow \mathbb{R}$ and $G : \mathbb{N} \rightarrow \mathbb{R}$, we say that $W(N) = O(G(N))$ if there exist $C > 0$ and $N_0 \in \mathbb{N}$ such that $W(N) \leq CG(N)$, for all $N \geq N_0$. The support supp$(f)$ of a function $f : \mathbb{R}^d \rightarrow \mathbb{C}$ is the closure of the set $\{ x \in \mathbb{R}^d \mid f(x) \neq 0 \}$ in the topology induced by the Euclidean norm $|\cdot|$. For a Lebesgue-measurable function $f : \mathbb{R}^d \rightarrow \mathbb{C}$, we write $\int f(x) dx$ for its integral w.r.t. Lebesgue measure. The indicator function of a set $B \subseteq \mathbb{R}^d$ is defined as $1_B(x) = 1$, for $x \in B$, and $1_B(x) = 0$, for $x \in \mathbb{R}^d \setminus B$. For a measurable set $B \subseteq \mathbb{R}^d$, we let vol$^d(B) := \int_{\mathbb{R}^d} 1_B(x) dx = \int_B 1 dx$, and we write $\mathcal{B}$ for its boundary. $L^p(\mathbb{R}^d)$, with $p \in [1, \infty)$, stands for the space of Lebesgue-measurable functions $f : \mathbb{R}^d \rightarrow \mathbb{C}$ satisfying $\|f\|_p := (\int_{\mathbb{R}^d} |f(x)|^p dx)^{1/p} < \infty$. $L^\infty(\mathbb{R}^d)$ denotes the space of Lebesgue-measurable functions $f : \mathbb{R}^d \rightarrow \mathbb{C}$ such that $\|f\|_\infty := \inf \{ \alpha > 0 \mid |f(x)| \leq \alpha \} \text{ for a.e. } x \in \mathbb{R}^d < \infty$. For a countable set $Q$, $(L^2(\mathbb{R}^d))^Q$ stands for the space of sets $S := \{ f_q \}_{q \in Q}$, with $f_q \in L^2(\mathbb{R}^d)$ for all $q \in Q$, satisfying $\|S\| := (\sum_{q \in Q} \|f_q\|_2^2)^{1/2} \leq \infty$. We denote the Fourier transform of $f \in L^1(\mathbb{R}^d)$ by $f(\omega) := \int_{\mathbb{R}^d} f(x) e^{-2\pi i x \cdot \omega} dx$ and extend it in the usual way to $L^2(\mathbb{R}^d)$\cite{40} Theorem 7.9).

Id : $L^p(\mathbb{R}^d) \rightarrow L^p(\mathbb{R}^d)$ stands for the identity operator on $L^p(\mathbb{R}^d)$. The convolution of $f \in L^2(\mathbb{R}^d)$ and $g \in L^1(\mathbb{R}^d)$ is $(f * g)(y) := \int_{\mathbb{R}^d} f(x) g(y - x) dx$. We write $(T_\ell)(f)(x) := f(x - \ell)$, for the translation operator, and $(M_\omega)(f)(x) := e^{2\pi i x \cdot \omega} f(x)$, $\omega \in \mathbb{R}^d$, for the modulation operator. We set $(f, g) := \int_{\mathbb{R}^d} f(x) g(x) dx$, for $f, g \in L^2(\mathbb{R}^d)$. $H^\alpha(\mathbb{R}^d)$, with $s \geq 0$, stands for the Sobolev space of functions $f \in L^2(\mathbb{R}^d)$ satisfying $\int_{\mathbb{R}^d} (1 + |x|^2)^s |f(x)|^2 dx < \infty$, see \cite{41} Section 6.2.1. Here, the index $s$ reflects the degree of smoothness of $f \in H^s(\mathbb{R}^d)$, i.e., larger $s$ entails smoother $f$. For a multi-index $\alpha \in \mathbb{N}_0^d$, $D^\alpha$ denotes the differential operator $D^\alpha := (\partial / \partial x_1)^{\alpha_1} \ldots (\partial / \partial x_d)^{\alpha_d}$, with order $|\alpha| := \sum_{i=1}^{d} \alpha_i$. The space of functions $f : \mathbb{R}^d \rightarrow \mathbb{C}$ whose derivatives $D^\alpha f$ of order at most $k \in \mathbb{N}_0$ are continuous is designated by $C_k(\mathbb{R}^d, \mathbb{C})$. Moreover, we denote the gradient of a function $f : \mathbb{R}^d \rightarrow \mathbb{C}$ as $\nabla f$.

II. DCNN-BASED FEATURE EXTRACTORS

Throughout the paper we employ the terminology of \cite{9}, consider (unless explicitly stated otherwise) input signals $f \in L^2(\mathbb{R}^d)$, and employ the module-sequence

$$\Omega := ((\Psi_n \cdot | \cdot |, \text{Id}))_{n \in \mathbb{N}^+} \tag{1}$$

i.e., each network layer is associated with (i) a collection of filters $\Psi_n := \{ \chi_n \} \cup \{ q_{\lambda_n} \}_{\lambda_n \in \Lambda_n} \subseteq L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$, where

\footnote{A wide range of practically relevant signal classes are Sobolev functions, for example, square-integrable functions, band-limited functions, and—as established in the present paper—cartoon functions\cite{35}.\footnote{Throughout the paper “a.e.” is w.r.t. Lebesgue measure.}\cite{19,29,56,57} as a model for natural images such as e.g., images of handwritten digits\cite{38}.}
\[ \chi_n, \text{ referred to as output-generating filter, and the } g_{\lambda_n}, \text{ indexed by a countable set } \Lambda_n, \text{ satisfy the frame condition } \begin{aligned} A_n \| f \|^2 &\leq \| f \ast \chi_n \|^2 + \sum_{\lambda_n \in \Lambda_n} \| f \ast g_{\lambda_n} \|^2 \leq B_n \| f \|^2, \tag{2} \end{aligned} \]

for all \( f \in L^2(\mathbb{R}^d) \), for some \( A_n, B_n > 0 \), (ii) the modulus non-linearity \( | \cdot | : L^2(\mathbb{R}^d) \to L^2(\mathbb{R}^d), \ | f(x) : = | f(x) |, \) and (iii) no pooling, which, in the terminology of [9], corresponds to pooling through the identity operator with pooling factor equal to one. Associated with the module \( (\Psi_n, | \cdot | _{\text{Id}}) \), the operator \( U_n[\lambda_n] \) defined in [9, Eq. 12] particularizes to

\[ U_n[\lambda_n] f = f \ast g_{\lambda_n}, \tag{3} \]

We extend (3) to paths on index sets

\[ q = (\lambda_1, \lambda_2, \ldots, \lambda_n) \in \Lambda_1 \times \Lambda_2 \times \cdots \times \Lambda_n =: \Lambda^n, \quad n \in \mathbb{N}, \]

according to

\[ U[q] f = U[(\lambda_1, \lambda_2, \ldots, \lambda_n)] f := U_n[\lambda_1] \cdots U_2[\lambda_2] U_1[\lambda_1] f, \tag{4} \]

where, for the empty path \( e := \emptyset \), we set \( \Lambda^0 := \{ e \} \) and \( U[e] f := f \). The signals \( U[q] f, q \in \Lambda^n \), associated with the \( n \)-th network layer, are often referred to as feature maps in the deep learning literature. The feature vector \( \Phi_\Omega(f) \) is obtained by aggregating filtered versions of the feature maps. More formally, \( \Phi_\Omega(f) \) is defined as [9, Def. 3]

\[ \Phi_\Omega(f) := \bigcup_{n=0}^{\infty} \Phi^n_\Omega(f), \tag{5} \]

where

\[ \Phi^n_\Omega(f) := \{(U[q] f) \ast \chi_{n+1}\}_{q \in \Lambda^n} \]

are the features generated in the \( n \)-th network layer, see Fig. [1]. Here, \( n = 0 \) corresponds to the root of the network. The function \( \chi_{n+1} \) is the output-generating filter of the \( n \)-th network layer. The feature extractor

\[ \Phi_\Omega : L^2(\mathbb{R}^d) \to \left( L^2(\mathbb{R}^d) \right)^{\bigcup_{n=0}^{\infty} \Lambda^n} \]

was shown in [9, Theorem 1] to be vertically translation-invariant, provided although that pooling is employed, with pooling factors \( S_n \geq 1, \ n \in \mathbb{N}, \) (see [9, Eq. 6] for the definition of the general pooling operator) such that \( \lim_{N \to \infty} \sum_{n=1}^{N} S_n = \infty \). Moreover, \( \Phi_\Omega \) exhibits limited sensitivity to certain non-linear deformations on (input) signal classes such as band-limited functions [9, Theorem 2], cartoon functions [19, Theorem 1], and Lipschitz functions [19, Corollary 1].

### III. ENERGY DECAY AND TRIVIAL NULL-SPACE

The first central goal of this paper is to understand how fast the energy contained in the feature maps decays across layers. Specifically, we shall study the decay of

\[ W_N(f) := \sum_{q \in \Lambda^N} \| U[q] f \|^2_2, \quad f \in L^2(\mathbb{R}^d), \tag{6} \]

as a function of network depth \( N \). Moreover, it is desirable that the infinite-depth feature vector \( \Phi_\Omega(f) \) be informative in the sense of the only signal mapping to the all-zeros feature vector being the zero input signal, i.e., \( \Phi_\Omega \) has a trivial null-space

\[ \mathcal{N}(\Phi_\Omega) := \{ f \in L^2(\mathbb{R}^d) \mid \Phi_\Omega(f) = 0 \} \ni \{ 0 \}. \tag{7} \]

Fig. [2] illustrates the practical ramifications of a non-trivial null-space in a binary classification task. \( \mathcal{N}(\Phi_\Omega) = \{ 0 \} \) can be guaranteed by asking for “energy conservation” in the sense of

\[ A_\Omega \| f \|^2_2 \leq \| \Phi_\Omega(f) \|^2 \leq B_\Omega \| f \|^2_2, \quad \forall f \in L^2(\mathbb{R}^d), \tag{8} \]

for some constants \( A_\Omega, B_\Omega > 0 \) (possibly depending on the module-sequence \( \Omega \)) and with the feature space norm \( \| \Phi_\Omega(f) \| := (\sum_{n=0}^{\infty} \| \Phi^n_\Omega(f) \|^2)^{1/2} \), where \( \| \Phi^n_\Omega(f) \| := \| U[n] f \|^2_2 \).
(\sum_{q \in \Lambda^n} \|(U[q]f) \ast x_n\|_2^2)^{1/2}. Indeed, \(2\) follows from \(8\) as the upper bound in \(8\) yields \(0 \subseteq N(\Phi_{\Omega})\), and the lower bound implies \(0 \supseteq N(\Phi_{\Omega})\). We emphasize that, as \(\Phi_{\Omega}\) is a non-linear operator (owing to the modulus non-linearities), characterizing its null-space is non-trivial in general. The upper bound in \(8\) was established in \([9]\) Appendix E]. While the existence of this upper bound is implied by the filters \(\Psi_a, n \in \mathbb{N}\), satisfying the frame property \(2\) \([9]\) Appendix E], perhaps surprisingly, this is not enough to guarantee \(A_\Omega > 0\) (see Appendix A) for an example). We refer the reader to Section V for results on the null-space of the finite-depth feature extractor \(J_{n=0}^{N} \Phi_{\Omega}^n\).

Previous work on the decay rate of \(W_N(f)\) in \([10]\) Section 5] shows that for wavelet-based networks (i.e., in every network layer the filters \(\Psi = \{\chi\} \cup \{g_{\lambda}\}_{\lambda \in A}\) are taken to be (specific) 1-D wavelets that constitute a Parseval frame, with \(\chi\) a low-pass filter) there exist \(\varepsilon > 0\) and \(a > 1\) (both constants unspecified) such that

\[
W_N(f) \leq \int_{\mathbb{R}} |\hat{f}(\omega)|^2 \left(1 - \left|\hat{r}_\Omega(\omega/a_{N-1})\right|^2\right) d\omega,
\]

for real-valued 1-D signals \(f \in L^2(\mathbb{R})\) and \(N \geq 2\), where \(\hat{r}_\Omega(\omega) := e^{-\omega^2} \cdot \mathbb{1}_{|\omega| < \mu}\). To see that this result indicates energy decay, Fig. 3 illustrates the influence of network depth \(N\) on the upper bound in \(9\). Specifically, we can see that increasing the network depth results in cutting out increasing amounts of energy of \(f\) and thereby making the upper bound in \(9\) decay as a function of \(N\). It is interesting to note that the upper bound on \(W_N(f) = \sum_{q \in \Lambda^n} \|(U[q]f)\|_2^2\) is independent of the wavelets generating the feature maps \(U[q]f, q \in \Lambda^n\). For scattering networks that employ, in every network layer, uniform covering filters \(\Psi = \{\chi\} \cup \{g_{\lambda}\}_{\lambda \in A} \subseteq L^2(\mathbb{R}) \cap L^2(\mathbb{R}^d)\) forming a Parseval frame (where \(\chi\), again, is a low-pass filter), exponential energy decay according to

\[
W_N(f) = \mathcal{O}(a^{-N}), \quad \forall f \in L^2(\mathbb{R}^d),
\]

for an unspecified \(a > 1\), was established in \([11]\) Proposition 3.3]. Moreover, \([10]\) Section 5] and \([11]\) Theorem 3.6 (a)] state—for the respective module-sequences—that \(9\) holds with \(A_\Omega = B_\Omega = 1\) and hence

\[
\|\Phi_{\Omega}(f)\|_2^2 = \|f\|_2^2.
\]

The first main goal of the present paper is to establish i) for \(d\)-dimensional complex-valued input signals that \(W_N(f)\) decays polynomially according to

\[
W_N(f) \leq B_N^2 \int_{\mathbb{R}^d} \left|\hat{f}(\omega)\right|^2 \left(1 - \left|\hat{r}_\Omega(\omega/a_{N-1})\right|^2\right) d\omega,
\]

for \(f \in L^2(\mathbb{R}^d)\) and \(N \geq 1\), where \(a = 1, d = 1, \text{and } \alpha = \log_2(\sqrt{d/(d-1/2)})\), for \(d \geq 2, B_N^2 = \prod_{k=1}^{N} \max\{1, B_k\}\), and \(\hat{r}_\Omega: \mathbb{R}^d \to \mathbb{R}, \hat{r}_\Omega(\omega) = (1 - |\omega|^4)^{l/4}, \text{with } l > [d/2] + 1, \text{for networks based on general filters } \{\chi\} \cup \{g_{\lambda}\}_{\lambda \in A}\) that satisfy mild analyticity and high-pass conditions and are allowed to be different in different network layers (with the proviso that \(\chi_n, n \in \mathbb{N}\), is of low-pass nature in a sense to be made precise), and ii) for \(d\)-D complex-valued input signals that \(9\) decays exponentially according to

\[
W_N(f) \leq \int_{\mathbb{R}} \left|\hat{f}(\omega)\right|^2 \left(1 - \left|\hat{r}_\Omega(\omega/a_{N-1})\right|^2\right) d\omega,
\]

for \(f \in L^2(\mathbb{R})\) and \(N \geq 1\), for networks that are based, in every network layer, on a broad family of wavelets, with the decay factor given explicitly as \(a = \frac{3}{2}\), or on a broad family of Weyl-Heisenberg filters \([9]\) Appendix B], with decay factor \(a = \frac{3}{2}\). Thanks to the right-hand side (RHS) of \(12\) and \(13\) not depending on the specific filters \(\{\chi\} \cup \{g_{\lambda}\}_{\lambda \in A}\), we will be able to establish—under smoothness assumptions on the input signal \(f\)—universal energy decay results. Specifically, particularizing the RHS expressions in \(12\) and \(13\) to Sobolev-class input signals \(f \in H^s(\mathbb{R}^d), s \geq 0\), where

\[
H^s(\mathbb{R}^d) = \{f \in L^2(\mathbb{R}^d) \mid \int_{\mathbb{R}^d} (1 + |\omega|^2)^s |\hat{f}(\omega)|^2 d\omega < \infty\},
\]

we show that \(12\) yields polynomial energy decay according to

\[
W_N(f) = \mathcal{O}(N^{-\frac{N(2s+\beta)}{2s+n+1}}), \quad \forall f \in H^s(\mathbb{R}^d),
\]

and \(13\) exponential energy decay

\[
W_N(f) = \mathcal{O}(a^{-N(2s+\beta)/2s+n+1}), \quad \forall f \in H^s(\mathbb{R}).
\]

The constant \(\beta > 0\) in \(14\) and \(15\) depends on the underlying Sobolev function \(\hat{f}(\omega) \in H^*(\mathbb{R}^d)\). The parameter \(s\) acts as a smoothness index; larger \(s\) leads to smoother \(f\) and faster decay of \(W_N(f)\). Sobolev spaces \(H^s(\mathbb{R}^d)\) contain a wide range of practically relevant signal classes such as, e.g.,

- the space of square-integrable functions \(L^2(\mathbb{R}^d) = H^0(\mathbb{R}^d)\),
- the space \(L^2_0(\mathbb{R}^d) := \{f \in L^2(\mathbb{R}^d) \mid \supp(\hat{f}) \subseteq B_l(0), L \geq 0, \text{of } L\text{-band-limited functions according to } L^2_0(\mathbb{R}^d) \subseteq H^s(\mathbb{R}^d), \text{for } L \geq 0 \text{and } s \geq 0\). This follows

Specifically, \(\beta > 0\) determines the decay of \(\hat{f}(\omega)\) as \(|\omega| \to \infty\) (see, e.g., [41] Sec. 6.2.1). For every \(f \in H^s(\mathbb{R}^d)\) there exist \(\beta, \mu, L > 0\) such that \(|\hat{f}(\omega)| \leq \mu(1 + |\omega|^2)^{-l/4}, \text{for } |\omega| \geq L, \text{where } L > 0 \text{ acts as an “effective bandwidth” of } f.\)
Our second central goal is to prove energy conservation according to (8) (which, as explained above, implies $N(\Phi_H) = \{0\}$) for the network configurations corresponding to the energy decay results (12) and (13). Finally, we provide handy estimates of the number of layers needed to have at least $(1 - \varepsilon) \cdot 100\%$ of the input signal energy be contained in the feature vector.

**IV. MAIN RESULTS**

Throughout the paper, we make the following assumptions on the filters $\{g_{\lambda_n}\}_{\lambda_n \in \Lambda_n}$.

**Assumption 1.** The $\{g_{\lambda_n}\}_{\lambda_n \in \Lambda_n}, n \in \mathbb{N}$, are analytic in the following sense: For every layer index $n \in \mathbb{N}$, for every $\lambda_n \in \Lambda_n$, there exists an orthonth $H_{A_{\lambda_n}} \subseteq \mathbb{R}^d$ with $A_{\lambda_n} \in O(d)$, such that

$$\text{supp}(g_{\lambda_n}) \subseteq H_{A_{\lambda_n}}.$$  

(16)

Moreover, there exists $\delta > 0$ so that

$$\sum_{\lambda_n \in \Lambda_n} |g_{\lambda_n}(\omega)|^2 = 0, \quad a.e. \ \omega \in B_\delta(0).$$  

(17)

In the 1-D case, i.e., for $d = 1$, Assumption 1 simply amounts to every filter $g_{\lambda_n}$ satisfying either

$$\text{supp}(g_{\lambda_n}) \subseteq (-\infty, -\delta) \quad \text{or} \quad \text{supp}(\hat{g}_{\lambda_n}) \subseteq [\delta, \infty),$$

which constitutes an “analyticity” and “high-pass” condition. For dimensions $d \geq 2$, Assumption 1 requires that every filter $g_{\lambda_n}$ be of high-pass nature and have a Fourier transform supported in a (not necessarily canonical) orthonth. Since the
The proof of statement iii) is based on

\[ A_n \leq |\hat{\chi}_n(\omega)|^2 + \sum_{\lambda_n \in \Lambda_n} |\hat{\chi}_n(\omega)|^2 \leq B_n, \quad \text{a.e. } \omega \in \mathbb{R}^d, \quad (18) \]

\[ \text{for all } f \in L^2(\mathbb{R}^d) \text{ and } N \geq 1, \]

\[ W_N(f) = O(B^{N}_{\Omega}N^{-\alpha(2+\beta)(2\alpha+\beta+1)}). \quad (21) \]

\[ 0 < A_{\Omega} := \lim_{N \to \infty} A^N_{\Omega} \leq B_{\Omega} := \lim_{N \to \infty} B^N_{\Omega} < \infty, \quad (22) \]

\[ \text{for all } f \in L^2(\mathbb{R}^d). \]

Proof. For the proofs of i) and ii), we refer to Appendices C and D respectively. The proof of statement iii) is based on two key ingredients. First, we establish—in Proposition 1 in Appendix E—that the feature extractor $\Phi_{\Omega}$ satisfies the energy decomposition identity

\[ A^N_{\Omega}\|f\|^2 \leq \sum_{n=0}^{N-1} \|\Phi^N_{\Omega}(f)\|^2 + W_N(f) \leq B^N_{\Omega}\|f\|^2, \quad (24) \]

for all $f \in L^2(\mathbb{R}^d)$ and all $N \geq 1$. Second, we show—in Proposition 2 in Appendix F—that the integral on the RHS of (20) goes to zero as $N \to \infty$ which, thanks to $\lim_{n \to \infty} B^N_{\Omega} = B_{\Omega} < \infty$, implies that $W_N(f) \to 0$ as $N \to \infty$. We note that while the decomposition (24) holds for general filters $\{g_{\lambda}\}_{\lambda \in \Lambda_n}$ satisfying the frame property (2), it is the upper bound (20) that makes use of the analyticity and high-pass conditions in Assumption 1. The final energy conservation result (23) is obtained by letting $N \to \infty$ in (24).

The strength of the results in Theorem 1 derives itself from the fact that the only condition we need to impose on the filters $\{g_{\lambda}\}_{\lambda \in \Lambda_n}$ is Assumption 1 which, as already mentioned, is met by a wide array of filters. Moreover, condition (22) is easily satisfied by normalizing the filters $\Psi_n$, $n \in \mathbb{N}$, appropriately (see, e.g., [9, Proposition 3]). We note that this normalization, when applied to filters that satisfy Assumption 1, yields filters that still meet Assumption 1.

The identity (21) establishes, upon normalization [9, Proposition 3] of the $\Psi_n$ to get $B_n \leq 1$, $n \in \mathbb{N}$, that the energy decay rate, i.e., the decay rate of $W_N(f)$, is at least polynomial in $N$. We hasten to add that (20) does not preclude the energy from decaying faster in practice. Finally, as already mentioned, increasing the smoothness index $s$ leads to faster energy decay irrespective of the filters employed in the network.

Underlying the energy conservation result (23) is the following demodulation effect induced by the modulus non-linearity in combination with the analyticity and high-pass properties of the filters $\{g_{\lambda}\}_{\lambda \in \Lambda_n}$. In every network layer, the spectral content of each individual feature map is moved to base-band (i.e., to low frequencies), where it is extracted by the low-pass output-generating atom $\chi_{n+1}$, see Fig. 5. The components not collected by $\chi_{n+1}$ (see Fig. 5 bottom row) are captured by the analytic high-pass filters $\{g_{\lambda_{n+1}}\}_{\lambda_{n+1} \in \Lambda_{n+1}}$ in the next layer and, thanks to the modulus non-linearity, again moved to low frequencies and extracted by $\chi_{n+2}$. Iterating this process ensures that the null-space of the feature vector (be it for the infinite-depth network or, as established in Section V for finite network depths) is trivial. It is interesting to observe that the sigmoid, the rectified linear unit, and the hyperbolic tangent non-linearities—all widely used in the deep learning literature—exhibit very different behavior in this regard, namely, they do not demodulate in the way the modulus non-linearity does [44, Fig. 6]. It is therefore unclear whether the proof machinery for energy conservation developed in this paper extends to these non-linearities or, for that matter, whether one gets energy decay and conservation at all.

The feature map energy decay result (21) relates to the feature vector energy conservation result (23) via the energy
Let \( \hat{\Theta} \) be the module-sequence with filters \( \Psi = \{\chi\} \cup \{g_j\}_{j \in \mathbb{Z} \setminus \{0\}} \) in every network layer. Then,
\[
W_N(f) = O\left(\left(\frac{5}{3}\right)^{-N(2+\delta)/(2\pi+2\delta)}\right).
\]

proof. See Appendix \( \square \)

**Appendix F**

This residual energy will eventually be collected in the infinite-depth feature vector \( \Phi_{x} \) so that no input signal energy is "lost" in the network. In Section \( \mathbb{F} \) we shall answer the question of how many layers are needed to absorb \((1-\varepsilon) \cdot 100\)% of the input signal energy.

The next result shows that, under additional structural assumptions on the filters \( \{g_{\lambda_n}\}_{n \in \Lambda} \), the guaranteed energy decay rate can be improved from polynomial to exponential. Specifically, we can get exponential energy decay for broad assumptions on the filters \( \{g_{\lambda_n}\}_{n \in \Lambda} \) ensures that—in every network layer—the spectral content of each individual feature map is moved to base-band (i.e., to low frequencies), where it is extracted by the (low-pass) output-generating filter \( \chi_{n+1} \).

**Appendix G**

**Theorem 2.** Let \( \hat{\Theta} : \mathbb{R} \to \mathbb{R} \), \( \hat{\Theta}(\omega) := (1-|\omega|)^{2\delta} \), with \( \delta > 0 \).

i) Wavelets: Let the mother and father wavelets \( \psi, \phi \in L^1(\mathbb{R}) \cap L^2(\mathbb{R}) \) satisfy \( \text{supp}(\psi) \subseteq [1/2, 2] \) and
\[
|\hat{\phi}(\omega)|^2 + \sum_{j=1}^{\infty} |\hat{\psi}(2^{-j}\omega)|^2 = 1, \quad a.e. \ \omega \geq 0.
\]

Moreover, let \( g_{j}(x) := 2^{j}\psi(2^{j}x), \) for \( x \in \mathbb{R}, \) \( j \geq 1 \), and \( g_{j}(x) := 2^{j}\psi(-2^{j}x), \) for \( x \in \mathbb{R}, \) \( j \leq -1 \), and set \( \chi(x) := \phi(x), \) for \( x \in \mathbb{R} \).

ii) Weyl-Heisenberg filters: For \( R \in \mathbb{R} \), let the functions \( g, \phi \in L^1(\mathbb{R}) \cap L^2(\mathbb{R}) \) satisfy \( \text{supp}(g) \subseteq [-R, R] \), \( g(\omega) = \hat{g}(\omega) \), for \( \omega \in \mathbb{R} \), and
\[
|\hat{g}(\omega)|^2 + \sum_{k=1}^{\infty} |\hat{g}(\omega-R(k+1))|^2 = 1, \quad a.e. \ \omega \geq 0.
\]
a.e. \( \omega \geq 0 \). Moreover, let \( g_{k}(x) := e^{2\pi ik(x)}g(x), \) for \( x \in \mathbb{R}, \) \( k \geq 1 \), and \( g_{k}(x) := e^{-2\pi ik(x)}g(x), \) for \( x \in \mathbb{R}, \) \( k \leq -1 \), and set \( \chi(x) := \phi(x), \) for \( x \in \mathbb{R} \).

Let \( \Omega \) be the module-sequence with filters \( \Psi = \{\chi\} \cup \{g_k\}_{k \in \mathbb{Z} \setminus \{0\}} \) in every network layer. Then,
\[
W_N(f) = O\left(\left(\frac{3}{2}\right)^{-N(2+\delta)/(2\pi+2\delta)}\right).
\]

Proof. See Appendix \( \square \)
The conditions we impose on the mother and father wavelets \(\psi, \phi\) in Section 3.3.5, and those on the prototype function \(g\) and low-pass filter \(\phi\) in ii) by B-splines \([15]\) Section 1. Moreover, as shown in \([14]\) Theorem 3.1], the exponential energy decay results in (28) and (31) can be generalized to \(O(a^{-N})\) with arbitrary decay factor \(a > 1\) realized through suitable choice of the mother wavelet or the Weyl-Heisenberg prototype function.

We note that in the presence of pooling by sub-sampling (as defined in \([9]\) Eq. 9)], say with pooling factors \(S_n := S \in [1, a]\), for all \(n \in \mathbb{N}\), (where \(a = \frac{6}{7}\) in the wavelet and \(a = \frac{3}{2}\) in the Weyl-Heisenberg case) the effective decay factor in (28) and (31) becomes \(\frac{5}{7}\) and \(\frac{3}{7}\), respectively. Exponential energy decay is hence compatible with vertical translation invariance according to \([9]\) Theorem 1, albeit at the cost of a slower (exponential) decay rate. The proof of this statement is structurally very similar to that of Theorem 2 and will therefore not be given here. Finally, we note that the energy decay and conservation results in Theorems 1 and 2 are compatible with the feature extractor \(\Phi_\Omega\) being deformation-insensitive according to \([9]\) Theorem 2, simply by noting that \([9]\) Theorem 2] applies to general semi-discrete frames and general Lipschitz-continuous non-linearities.

We next put the results in Theorems 1 and 2 into perspective with respect to the literature.

Relation to \([10]\) Section 5]: The basic philosophy of our proof technique for (20), (23), (27), and (30) is inspired by the proof in \([10]\) Section 5], which establishes (9) and (11] for scattering networks based on certain wavelet filters and with 1-D real-valued input signals \(f \in L^2(\mathbb{R})\). Specifically, in \([10]\) Section 5], in every network layer, the filters \(\Psi_W = \{\chi\} \cup \{g_j\}_{j \in \mathbb{Z}}\) (where \(g_j(\omega) := 2^j \hat{\psi}(2^j \omega), j \in \mathbb{Z}\), for some mother wavelet \(\psi \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})\)) are 1-D functions satisfying the frame property (2) with \(A_n = B_n = 1\), \(n \in \mathbb{N}\), a mild analyticity condition (\([10]\) Eq. 5.5]) in the sense of \(\hat{g}_j(\omega)\), \(j \in \mathbb{Z}\), being larger for positive frequencies \(\omega\) than for the corresponding negative ones, and a vanishing moments condition \([10]\) Eq. 5.6]) which controls the behavior of \(\hat{\psi}(\omega)\) around the origin according to \(|\hat{\psi}(\omega)| \leq C |\omega|^{1+\varepsilon}\), for \(\omega \in \mathbb{R}\), for some \(C, \varepsilon > 0\). Similarly to the proof of (11] as given in \([10]\) Section 5], we base our proof of (23) on the energy decomposition identity (24) and on an upper bound on \(W_N(f)\) (see \([9]\) for the corresponding upper bound established in \([10]\) Section 5]) shown to go to zero as \(N \to \infty\).

The exponential energy decay results (21), (28), and (31) for \(f \in H^s(\mathbb{R}^d)\) are entirely new. The major differences between \([10]\) Section 5] and our results are i) the results in \([11]\) apply exclusively to filters satisfying the uniform covering condition such as, e.g., Weyl-Heisenberg filters with a band-limited prototype function \([11]\) Proposition 2.3], but do not apply to multi-scale filters such as wavelets, (\(\alpha\)-)curvelets, shearlets, and ridgelets (see \([11]\) Remark 2.2 (b)], ii) \([10]\) leaves the decay factor \(a > 1\) unspecified, whereas our results in (28) and (31) make the decay factor \(a\) explicit (namely, \(a = \frac{5}{3}\) in the wavelet case and \(a = \frac{3}{2}\) in the Weyl-Heisenberg case), iii) the exponential energy decay result in \([10]\] as established in \([11]\] is universal in \(f \in L^2(\mathbb{R}^d)\), i.e., the RHS of \([10]\] is independent of \(f \in L^2(\mathbb{R}^d)\), whereas our decay results in (21), (28), and (31) depend on the particular \(f\) through the parameter \(\beta\) (which determines the decay of \(\hat{f}(\omega)\) as \(|\omega| \to \infty\)), iv) the technical elements employed to arrive at the upper bounds on \(W_N(f)\), specifically, while the proof in \([11]\] makes explicit use of the uniform covering property of the filters, our proof of (20) is completely oblivious to the (algebraic) structure of the filters, (v) the assumptions imposed on the filters, i.e., the vanishing moments and uniform covering condition in \([11]\] Definition 2.1 (a)-(b)], in contrast to our Assumption 1, which is less restrictive, and thereby makes our results in Theorem 1 apply to general (possibly unstructured) filters that, in addition, can be different in different network layers.

V. Number of layers needed

DCNNs used in practice employ potentially hundreds of layers \([7]\]. Such network depths entail formidable computational challenges both in training and in operating the network. It is therefore important to understand how many layers are needed to have most of the input signal energy be contained in the feature vector. This will be done by considering Parseval frames in all layers, i.e., frames with frame bounds \(A_n = B_n = 1\), \(n \in \mathbb{N}\). The energy conservation result (23) then implies that the infinite-depth feature vector \(\Phi_\Omega(f) = \bigcup_{n=1}^{\infty} \Phi_\Omega^n(f)\) contains the entire input signal energy according to \(|\|\Phi_\Omega(f)\||^2 = \sum_{n=0}^{\infty} ||\Phi_\Omega^n(f)||^2 = ||f||_2^2\). Now,
the decomposition \((19)\) reveals that thanks to \(\lim_{N \to \infty} W_N(f) \to 0\), increasing the network depth \(N\) implies that the feature vector \(\bigcup_{n=0}^{N} \Phi^* \Omega(f)\) progressively contains a larger fraction of the input signal energy. We formalize the question on the number of layers needed by asking for bounds of the form

\[
(1 - \varepsilon) \leq \frac{\sum_{n=0}^{N} ||\Phi^* \Omega(f)||^2}{||f||^2} \leq 1, \tag{32}
\]

i.e., by determining the network depth \(N\) guaranteeing that at least \((1 - \varepsilon) \cdot 100\%\) of the input signal energy are captured by the corresponding depth-\(N\) feature vector \(\bigcup_{n=0}^{N} \Phi^* \Omega(f)\). Moreover, \((32)\) ensures that the depth-\(N\) feature extractor \(\bigcup_{n=0}^{N} \Phi^* \Omega(f)\) exhibits a trivial null-space.

The following results establish handy estimates of the number \(N\) of layers needed to guarantee \((32)\). For pedagogical reasons, we start with the case of band-limited input signals and then proceed to a more general statement pertaining to Sobolev functions \(H^s(\mathbb{R}^d)\).

**Corollary 1.**

i) Let \(\Omega\) be the module-sequence \((1)\) with filters \(\{g_{\lambda_n}\}_{\lambda_n \in \Lambda_n}\) satisfying the conditions in Assumption \(7\) and let the corresponding frame bounds be \(A_n = B_n = 1, n \in \mathbb{N}\). Let \(\delta > 0\) be the radius of the spectral gap \(B_\delta(0)\) left by the filters \(\{g_{\lambda_n}\}_{\lambda_n \in \Lambda_n}\) according to \((17)\). Furthermore, let \(l > |d/2| + 1, \varepsilon \in (0, 1), \alpha\) as defined in \((19)\), and \(f \in L^2(\mathbb{R}^d)\) B-band-limited. If

\[
N \geq \left( \frac{L}{(1 - (1 - \varepsilon) \frac{1}{\pi}) \delta} \right)^{1/\alpha} - 1 \tag{33}
\]

then \((32)\) holds.

ii) Assume that the conditions in Theorem 2(i) and ii) hold. For the wavelet case, let \(a = \frac{5}{4}\) and \(\delta = 1\) (where \(\delta\) corresponds to the radius of the spectral gap left by the wavelets \(\{g_j\}_{j \in \mathbb{Z}\setminus \{0\}}\) ). For the Weyl-Heisenberg case, let \(a = \frac{2}{3}\) and \(\delta = \rho\) (here, \(\delta\) corresponds to the radius of the spectral gap left by the Weyl-Heisenberg filters \(\{g^*_k\}_{k \in \mathbb{Z}\setminus \{0\}}\) ). Moreover, let \(l > 1, \varepsilon \in (0, 1), \) and \(f \in L^2(\mathbb{R})\) B-band-limited. If

\[
N \geq \left[ \log_a \left( \frac{L}{(1 - (1 - \varepsilon) \frac{1}{\pi}) \delta} \right) \right], \tag{34}
\]

then \((32)\) holds in both cases.

**Proof.** See Appendix \([1]\). \(\Box\)

Corollary 1 nicely shows how the description complexity of the signal class under consideration, namely the bandwidth \(L\) and the dimension \(d\) through the decay exponent \(\alpha\) defined in \((19)\) determine the number \(N\) of layers needed. Specifically, \((33)\) and \((34)\) show that larger bandwidths \(L\) and larger dimension \(d\) render the input signal \(f\) more “complex”, which requires deeper networks to capture most of the energy of \(f\). The dependence of the lower bounds in \((33)\) and \((34)\) on the network properties, through the module-sequence \(\Omega\), is through the decay factor \(a > 1\) and the radius \(\delta\) of the spectral gap left by the filters \(\{g_{\lambda_n}\}_{\lambda_n \in \Lambda_n}\).

The following numerical example provides qualitative insights on the influence of the parameter \(\varepsilon\) on \((33)\) and \((34)\).

Specifically, we set \(L = 1, \delta = 1, d = 1\) (which implies \(\alpha = 1\), see \((19)\), \(l = 1.0001\), and show in Table I the number \(N\) of layers needed according to \((33)\) and \((34)\) for different values of \(\varepsilon\). The results show that 95% of the input signal energy are contained in the first 8 layers in the wavelet case and in the first 10 layers in the Weyl-Heisenberg case. We can therefore conclude that in practice a relatively small number of layers is needed to have most of the input signal energy be contained in the feature vector. In contrast, for general filters, where we can guarantee polynomial energy decay only, \(N = 39\) layers are needed to absorb 95% of the input signal energy. We hasten to add, however, that \((20)\) simply guarantees polynomial energy decay and does not preclude the energy from decaying faster in practice.

We proceed with the estimates for Sobolev-class input signals.

**Corollary 2.**

i) Let \(\Omega\) be the module-sequence \((1)\) with filters \(\{g_{\lambda_n}\}_{\lambda_n \in \Lambda_n}\) satisfying the conditions in Assumption \(7\) and let the corresponding frame bounds be \(A_n = B_n = 1, n \in \mathbb{N}\). Let \(\delta > 0\) be the radius of the spectral gap \(B_\delta(0)\) left by the filters \(\{g_{\lambda_n}\}_{\lambda_n \in \Lambda_n}\) according to \((17)\). Furthermore, let \(l > |d/2| + 1, \varepsilon \in (0, 1), \alpha\) as defined in \((19)\), and \(f \in H^s(\mathbb{R}^d)\setminus \{0\}\). If

\[
N \geq \left[ \left( \frac{\max\{L, \gamma\}}{(1 - (1 - \varepsilon) \frac{1}{\pi}) \delta} \right)^{1/\alpha} - 1 \right], \tag{35}
\]

where

\[
\gamma := \left( \frac{\mu^2 \vol^{d-1}(\partial B_\delta(0))}{(2s + \beta)(1 - \sqrt{1 - \varepsilon}) \|f\|_2^2} \right)^{\frac{1}{2s + \beta}}, \tag{36}
\]

then \((32)\) holds. Here, the parameters \(\beta, \mu, L > 0\) depend on \(f\) according to

\[
|\hat{f}(\omega)| \leq \frac{\mu}{(1 + |\omega|)^{\frac{d+1}{2} + \frac{s}{2} + \frac{\beta}{2} + \frac{\alpha}{2}}}, \quad \forall |\omega| \leq L.
\]

ii) Assume that the conditions in Theorem 2(i) and ii) hold. For the wavelet case, let \(a = \frac{5}{4}\) and \(\delta = 1\) (where \(\delta\) corresponds to the radius of the spectral gap left by the wavelets \(\{g_j\}_{j \in \mathbb{Z}\setminus \{0\}}\) ). For the Weyl-Heisenberg case, let \(a = \frac{2}{3}\) and \(\delta = \rho\) (here, \(\delta\) corresponds to the radius of the spectral gap left by the Weyl-Heisenberg filters \(\{g^*_k\}_{k \in \mathbb{Z}\setminus \{0\}}\) ). Furthermore, let \(l > 1, \varepsilon \in (0, 1), \) and \(f \in H^s(\mathbb{R})\setminus \{0\}\). If

\[
N \geq \left[ \log_a \left( \frac{\max\{L, \gamma\}}{(1 - (1 - \varepsilon) \frac{1}{\pi}) \delta} \right) \right], \tag{37}
\]
where

\[ \gamma := \left( \frac{2\mu^2}{(2s + \beta)(1 - \sqrt{1 - \varepsilon})} \right) \frac{1}{d + \mu}, \]

(38)

then (32) holds. Here, the parameters \( \beta, \mu, L > 0 \) depend on \( f \) according to

\[ |f(\omega)| \leq \frac{\mu}{(1 + |\omega|^2)^{\frac{s}{2} + \frac{1}{4} + \frac{\beta}{4}}}, \quad \forall |\omega| \geq L. \]

Proof. See Appendix \[ \square \]

As already mentioned in Section \[ \[ \] \] Sobolev spaces \( H^s(\mathbb{R}^d) \) contain a wide range of practically relevant signal classes. The results in Corollary \[ \[ \] \] therefore provide—for a wide variety of input signals—a picture of how many layers are needed to have most of the input signal energy be contained in the feature vector.

The width of the networks considered throughout the paper is, in principle, infinite as the sets \( \Lambda_n \) need to be countably infinite in order to guarantee that the frame property (2) is satisfied. As Sobolev-type input signals exhibit spectral decay, the number of “operationally significant nodes” will, however, be finite in practice. For a treatment of this aspect as well as results on depth-width tradeoffs, the interested reader is referred to [44].

**Appendix A**

**A Feature Extractor with a Non-trivial Kernel**

We show, by way of example, that employing filters \( \Psi_n \) which satisfy the frame property (2) alone does not guarantee a trivial null-space for the feature extractor \( \Phi_{\Omega} \). Specifically, we construct a feature extractor \( \Phi_{\Omega} \) based on filters satisfying (2) and a corresponding function \( f \neq 0 \) with \( f \in \mathcal{N}(\Phi_{\Omega}) \).

Our example employs, in every network layer, filters \( \Psi = \{\chi\} \cup \{g_k \}_{k \in \mathbb{Z}} \) that satisfy the Littlewood-Paley condition (18) with \( A = B = 1 \), and where \( g_0 \) is such that \( \hat{g}_0(\omega) = 1 \), for \( \omega \in B_1(0) \), and arbitrary else (of course, as long as the Littlewood-Paley condition (18) with \( A = B = 1 \) is satisfied). We emphasize that no further restrictions are imposed on the filters \( \{\chi\} \cup \{g_k \}_{k \in \mathbb{Z}} \), specifically \( \chi \) need not be of low-pass nature and the filters \( \{g_k \}_{k \in \mathbb{Z}} \) may be structured (such as wavelets [9] Appendix B) or unstructured (such as random filters [48], [49]), as long as they satisfy the Littlewood-Paley condition (18) with \( A = B = 1 \). Now, consider the input signal \( f \in L^2(\mathbb{R}^d) \) according to

\[ \hat{f}(\omega) := (1 - |\omega|)^s_+, \quad \omega \in \mathbb{R}^d, \]

with \( l > |d/2| + 1 \). Then \( f \ast g_0 = f \), owing to \( \text{supp}(\hat{f}) = B_1(0) \) and \( \hat{g}_0(\omega) = 1 \), for \( \omega \in B_1(0) \). Moreover, \( \hat{f} \) is a positive definite radial basis function [50] Theorem 6.20] and hence by [50] Theorem 6.18] \( f(x) \geq 0, x \in \mathbb{R}^d \), which, in turn, implies \( \|f\| = f \). This yields

\[ U[g_0^N]f = \cdots \|f \ast g_0 \ast g_0 | \cdots g_0 | = f, \]

for \( g_0^N := (0, 0, \ldots., 0) \in \mathbb{Z}^N \) and \( N \in \mathbb{N} \). Owing to the energy decomposition identity (24), together with \( A^N_1 = B^N_1 = 1 \), \( N \in \mathbb{N} \), which, in turn, is by \( A_n = B_n = 1, n \in \mathbb{N} \), we have

\[ \|f\|_2^2 = \sum_{n=0}^{N-1} ||\Phi^n_{\Omega}(f)||^2 + W_N(f) = \sum_{n=0}^{N-1} ||\Phi^n_{\Omega}(f)||^2 + \sum_{q \in \mathbb{Z}^N \setminus \{0^N\}} ||U[q]f||_2^2, \]

for \( N \in \mathbb{N} \). This implies

\[ \sum_{n=0}^{N-1} ||\Phi^n_{\Omega}(f)||^2 + \sum_{q \in \mathbb{Z}^N \setminus \{0^N\}} ||U[q]f||_2^2 = 0. \]

(39)

As both terms in (39) are positive, we can conclude that \( \sum_{n=0}^{N=0} ||\Phi^n_{\Omega}(f)||^2 = 0, N \in \mathbb{N} \), and thus \( ||\Phi_{\Omega}(f)||^2 = \sum_{n=0}^{N=0} ||\Phi^n_{\Omega}(f)||^2 = 0 \). Since \( ||\Phi_{\Omega}(f)||^2 = 0 \) implies \( \Phi_{\Omega}(f) = 0 \), we have constructed a non-zero \( f \), namely

\[ f(x) = \int_{\mathbb{R}^d} (1 - |\omega|)^s_+ e^{2\pi i (x, \omega)} \omega, \]

that maps to the all-zeros feature vector, i.e., \( f \in \mathcal{N}(\Phi_{\Omega}) \).

The point of this example is the following. Owing to the nature of \( \hat{g}_0(\omega) \) (namely, \( \hat{g}_0(\omega) = 1 \), for \( \omega \in B_1(0) \)) and the Littlewood-Paley condition

\[ |\hat{\chi}(\omega)|^2 + \sum_{k \in \mathbb{Z}} |\hat{g}_k(\omega)|^2 = 1, \quad \text{a.e.} \ \omega \in \mathbb{R}^d, \]

it follows that neither the output-generating filter \( \chi \) nor any of the other filters \( g_k, k \in \mathbb{Z} \), can have spectral support in \( B_1(0) \). Consequently, the only non-zero contribution to the feature vector can come from

\[ U[g_0^N]f = f, \]

which, however, thanks to \( \text{supp}(\hat{f}) = B_1(0) \), is spectrally disjoint from the output-generating filter \( \chi \). Therefore, \( \Phi_{\Omega}(f) \) will be identically equal to 0. Assumption 1 disallows this situation as it forces the filters \( g_k, k \in \mathbb{Z} \), to be of high-pass nature which, in turn, implies that \( \chi \) must have low-pass characteristics. The punch-line of our general results on energy conservation, be it for finite \( N \) or for \( N \to \infty \), is that Assumption 1 in combination with the frame property and the modulus non-linearity prohibit a non-trivial null-space in general.

**Appendix B**

**Sobolev Smoothness of Cartoon Functions**

Cartoon functions, introduced in [35], satisfy mild decay properties and are piecewise continuously differentiable apart from curved discontinuities along Lipschitz-continuous hypersurfaces. This function class has been widely adopted in the literature [15], [19], [26], [36], [37] as a standard model for natural images such as, e.g., images of handwritten digits [38] (see Fig. 4). We proceed to the formal definition of cartoon functions.

**Definition 1.** The function \( f : \mathbb{R}^d \to \mathbb{C} \) is referred to as a cartoon function if it can be written as \( f = f_1 + I_D f_2 \), where \( D \subseteq \mathbb{R}^d \) is a compact Lipschitz domain with boundary of finite length, i.e., \( \text{vol}^{d-1}(\partial D) < \infty \), and \( f_i \in L^2(\mathbb{R}^d) \cap C^2(\mathbb{R}^d, \mathbb{C}), \)
Lemma 1. Let $f_i$ satisfy the decay condition
\[
|\nabla f_i(x)| \leq C(x)^{-d}, \quad i = 1, 2,
\]
for some $C > 0$ (not depending on $f_1, f_2$). Furthermore, we denote by
\[
C^K_{\text{CART}} := \{f_1 + 1_D f_2 \mid f_i \in L^2(\mathbb{R}^d) \cap C^2(\mathbb{R}^d, \mathbb{C}), \ i = 1, 2, \ \ n \ \ \ \ n
|\nabla f_i(x)| \leq K(x)^{-d}, \ \ vol^{d-1}(\partial D) \leq K, \ |f_2|_\infty \leq K\}
\]
the class of cartoon functions of size $K > 0$.

Even though cartoon functions are in general discontinuous, they still admit Sobolev smoothness. The following result formalizes this statement.

Lemma 1. Let $K > 0$. Then $C^K_{\text{CART}} \subseteq H^s(\mathbb{R}^d)$, for all $s \in (0, 1/2)$.

Proof. Let $(f_1 + 1_D f_2) \in C^K_{\text{CART}}$. We first establish $1_D \in H^s(\mathbb{R}^d)$, for all $s \in (0, 1/2)$. To this end, we define the Sobolev-Slobodeckii semi-norm $[51]$ Section 2.1.2
\[
|f|_{H^s} := \left( \int_{\mathbb{R}^d} \left( \int_{\mathbb{R}^d} \frac{|f(x) - f(y)|^2}{|x-y|^{2+d}} \, dy \right)^{1/s} \right)^{1/2},
\]
and note that, thanks to $[51]$ Section 2.1.2, $1_D \in H^s(\mathbb{R}^d)$ if $|1_D|_{H^s} < \infty$. We have
\[
|1_D|_{H^s} = \int_{\mathbb{R}^d} \left( \int_{\mathbb{R}^d} \frac{|1_D(x) - 1_D(y)|^2}{|x-y|^{2+d}} \, dy \right)^{1/2} \, dx,
\]
where we employed the change of variables $t = x - y$. Next, we note that the function
\[
h(x) := |1_D(x) - 1_D(x-t)|^2
\]
satisfies $h(x) = 1$, for $x \in S$, where
\[
S := \{x \in \mathbb{R}^d \mid x \in D \text{ and } x - t \notin D\}
\]
and $h(x) = 0$, for $x \in \mathbb{R}^d \setminus S$. Since $S \subseteq (\partial D + B_{|t|}(0))$, where $(\partial D + B_{|t|}(0))$ is a tubular neighborhood of diameter $2|t|$ around the boundary $\partial D$ of $D$ (see Fig. 6), we have
\[
\int_{\mathbb{R}^d} |1_D(x) - 1_D(x-t)|^2 \, dx = \int_{\mathbb{R}^d} h(x) \, dx = \int_S \, dx = vol^d(S) \leq \min\{2|t|vol^{d-1}(\partial D), 2vol^d(D)\}.
\]
Let $R > 0$. We have
\[
|1_D|_{H^s} \leq \int_{\mathbb{R}^d} \frac{g(t)}{|t|^{2+d}} \, dt + \int_{B_R(0)} \frac{g(t)}{|t|^{2+d}} \, dt \leq \int_{\mathbb{R}^d \setminus B_R(0)} \frac{2vol^d(D)}{|t|^{2+d}} \, dt + \int_{B_R(0)} \frac{2vol^{d-1}(\partial D)}{|t|^{2+d-1}} \, dt
\]
\[
= 2vol^d(D) vol^{d-1}(\partial B_1(0)) \int_R^\infty \frac{r^{-2s+1}}{r} \, dr = I_1 + I_2,
\]
for fixed $f \in L^2(\mathbb{R}^d) \cap C^2(\mathbb{R}^d, \mathbb{C})$, constituting a bounded linear operator on $H^s(\mathbb{R}^d)$, for all $s \in (0, 1/2)$, i.e., $|hf|_{H^s} \leq C|h|_{H^s}$, for some $C > 0$, and the last inequality follows from $|1_D|_{H^s} < \infty$, for $s \in (0, 1/2)$, established above, and $|f_1|_{H^s} < \infty$, for $s \in (0, 1/2)$, which is by $f_1 \in L^2(\mathbb{R}^d) \cap C^2(\mathbb{R}^d, \mathbb{C})$. This completes the proof. □
for $N \in \mathbb{N}$, where
\[
C_n^{N+1} := \prod_{k=n}^{n+N-1} \max\{1, B_k\}.
\]
Setting $n = 1$ in (41) and noting that $C_n^N = B_n^N$ yields the desired result (20). We proceed by induction over the path length $\ell(q) := N$, for $q = (\lambda_n, \lambda_{n+1}, \ldots, \lambda_{n+N-1}) \in \Lambda_n \times \Lambda_{n+1} \times \cdots \times \Lambda_{n+N-1}$. Starting with the base case $N = 1$, we have
\[
\sum_{q \in \Lambda_n} \|U[q]f\|^2 = \sum_{\lambda_n \in \Lambda_n} \|f * g_{\lambda_n}\|^2 = \int_{\mathbb{R}^d} \sum_{\lambda_n \in \Lambda_n} \|\hat{g}_{\lambda_n}(\omega)(f(\omega))^2d\omega
\]
\[
\leq B_n \int_{\mathbb{R}^d \setminus B_1(0)} \|\hat{f}(\omega)\|^2d\omega \leq \max\{1, B_n\} \int_{\mathbb{R}^d} \|\hat{f}(\omega)\|^2d\omega,
\]
for $N \in \mathbb{N}$, where (42) is by Parseval’s formula, (43) is thanks to (17) and (18), and (44) is due to supp$(\hat{f}) \subseteq B_1(0)$ and $0 \leq \hat{r}(\omega) \leq 1$, for $\omega \in \mathbb{R}^d$. The inductive step is established as follows. Let $N > 1$ and suppose that (41) holds for all paths $q$ of length $\ell(q) = N - 1$, i.e.,
\[
\sum_{q \in \Lambda_N \times \Lambda_{N+1} \times \cdots \times \Lambda_{N+N-2}} \|U[q]f\|^2 \leq C_n^{N+N-2} \int_{\mathbb{R}^d} \|\hat{f}(\omega)\|^2 \left(1 - \left|\hat{r}\left(\frac{\omega}{N-1}\right)\right|^2\right)d\omega,
\]
for $n \in \mathbb{N}$. We start by noting that every path $\tilde{q} \in \Lambda_n \times \Lambda_{n+1} \times \cdots \times \Lambda_{N+N-1}$ of length $\ell(\tilde{q}) = N$, with arbitrary starting index $n$, can be decomposed into a path $q \in \Lambda_{n+1} \times \cdots \times \Lambda_{N+N-1}$ of length $\ell(q) = N - 1$ and an index $\lambda_n \in \Lambda_n$ according to $\tilde{q} = (\lambda_n, q)$. Thanks to (4) we have $U[q] = U[\lambda_n, q] = U[q]U[\lambda_n]$, which yields
\[
\sum_{q \in \Lambda_n \times \Lambda_{n+1} \times \cdots \times \Lambda_{N+N-1}} \|U[q]f\|^2 = \sum_{\lambda_n \in \Lambda_n} \sum_{q \in \Lambda_{n+1} \times \cdots \times \Lambda_{N+N-1}} \|U[q]U[\lambda_n]f\|^2,
\]
for $n \in \mathbb{N}$. We proceed by examining the inner sum on the RHS of (46). Invoking the induction hypothesis (45) with $n$ replaced by $(n + 1)$ and employing Parseval’s formula, we get
\[
\sum_{q \in \Lambda_{n+1} \times \cdots \times \Lambda_{N+N-1}} \|U[q]U[\lambda_n]f\|^2 \leq C_n^{N+n-1} \int_{\mathbb{R}^d} \|U[\lambda_n]f(\omega)\|^2 \left(1 - \left|\hat{r}\left(\frac{\omega}{N-1}\right)\right|^2\right)d\omega
\]
\[
= C_n^{N+n-1} \int_{\mathbb{R}^d} \left\|U[\lambda_n]f(\omega)\right\|^2 \left(1 - \left|\hat{r}\left(\frac{\omega}{N-1}\right)\right|^2\right)d\omega
\]
\[
= C_n^{N+n-1} \left(\|f * g_{\lambda_n}\|^2 - \|f * g_{\lambda_n} * r_{l, N-1, -1, \alpha, \delta}\|^2\right),
\]
for $n \in \mathbb{N}$, where $r_{l, N-1, -1, \alpha, \delta}(x)$ is the inverse Fourier transform of $\hat{r}\left(\frac{\omega}{N-1}\right)$. Next, we note that $\hat{r}\left(\frac{\omega}{N-1}\right)$ is a positive definite radial basis function [50, Theorem 6.20] and hence by [50, Theorem 6.18] $r_{l, N-1, -1, \alpha, \delta}(x) \geq 0$, for $x \in \mathbb{R}^d$. Furthermore, it follows from Lemma 2 stated below, that for $\{\nu_{\lambda_n}\}_{\lambda_n \in \Lambda_n} \subseteq \mathbb{R}^d$, we have
\[
\|f * g_{\lambda_n} * r_{l, N-1, -1, \alpha, \delta}\|^2 \geq \|f * (M_{\lambda_n}r_{l, N-1, -1, \alpha, \delta})\|^2.
\]
Here, we note that choosing the modulation factors $\{\nu_{\lambda_n}\}_{\lambda_n \in \Lambda_n} \subseteq \mathbb{R}^d$ appropriately (see (52) below) will be key in establishing the inductive step.

**Lemma 2.** [Lemma 2.7] Let $f, g \in L^2(\mathbb{R}^d)$ with $g(x) \geq 0$, for $x \in \mathbb{R}^d$. Then, $\|f * g\|^2 \geq \|f * (M_{\omega}g)\|^2$, for $\omega \in \mathbb{R}^d$.

Inserting (47) and (48) into the inner sum on the RHS of (46) yields
\[
\sum_{q \in \Lambda_{n+1} \times \cdots \times \Lambda_{N+N-1}} \|U[q]f\|^2 \leq C_n^{N+n-1} \sum_{\lambda_n \in \Lambda_n} \left(\|f * g_{\lambda_n}\|^2 - \|f * g_{\lambda_n} * r_{l, N-1, -1, \alpha, \delta}\|^2\right)
\]
\[
= C_n^{N+n-1} \int_{\mathbb{R}^d} \|\hat{f}(\omega)\|^2 \left(1 - \left|\hat{r}\left(\frac{\omega}{N-1}\right)\right|^2\right)d\omega, \quad \forall N \in \mathbb{N},
\]
where we applied Parseval’s formula together with $\sum_{\omega} = T_{\omega} f$, for $f \in L^2(\mathbb{R}^d)$, and $\omega \in \mathbb{R}^d$, and set
\[
h_{n,N,\alpha,\delta}(\omega) := \sum_{\lambda_n \in \Lambda_n} \left|\hat{g}_{\lambda_n}(\omega)\right|^2 \left(1 - \left|\hat{r}\left(\frac{\omega - \nu_{\lambda_n}}{(N-1)\alpha\delta}\right)\right|^2\right).
\]
The key step is now to establish—by judiciously choosing $\{\nu_{\lambda_n}\}_{\lambda_n \in \Lambda_n} \subseteq \mathbb{R}^d$—the upper bound
\[
h_{n,N,\alpha,\delta}(\omega) \leq \max\{1, B_n\} \left(1 - \left|\hat{r}\left(\frac{\omega}{N\alpha\delta}\right)\right|^2\right),
\]
for $\omega \in \mathbb{R}^d$, which upon noting that $C_n^{N+n-1} = \max\{1, B_n\} C_n^{N+N-1}$ yields (41) and thereby completes the proof. We start by defining $H_{\lambda_n}$, for $\lambda_n \in \Lambda_n$, to be the orthonormal supporting $g_{\lambda_n}$, i.e., supp$(g_{\lambda_n}) \subseteq H_{\lambda_n}$, where $\lambda_n \in O(d)$ (see Assumption 1). Furthermore, for $\lambda_n \in \Lambda_n$, we choose the modulation factors according to
\[
\nu_{\lambda_n} := A_{\lambda_n} \nu \in \mathbb{R}^d,
\]
for the components of $\nu \in \mathbb{R}^d$ are given by $\nu_k := (1 + 2^{-1/2})^k$, for $k \in \{1, \ldots, d\}$. Invoking (16) and (17), we get
\[
h_{n,N,\alpha,\delta}(\omega) = \sum_{\lambda_n \in \Lambda_n} \left|\hat{g}_{\lambda_n}(\omega)\right|^2 \left(1 - \left|\hat{r}\left(\frac{\omega - \nu_{\lambda_n}}{(N-1)\alpha\delta}\right)\right|^2\right)
\]
\[
= \sum_{\lambda_n \in \Lambda_n} \left|\hat{g}_{\lambda_n}(\omega)\right|^2 \left|S_{\lambda_n, \delta}(\omega)\right|^2 \left(1 - \left|\hat{r}\left(\frac{\omega - \nu_{\lambda_n}}{(N-1)\alpha\delta}\right)\right|^2\right),
\]
for $\omega \in \mathbb{R}^d$, where $S_{\lambda_n, \delta} := H_{\lambda_n} \setminus B_3(0)$. For the first canonical orthon $H = \{x \in \mathbb{R}^d \mid x_k \geq 0, \ k = 1, \ldots, d\}$, we show in Lemma 3 below that
\[
\left|\hat{r}\left(\frac{\omega - \nu_{\lambda_n}}{(N-1)\alpha\delta}\right)\right| \geq \left|\hat{r}\left(\frac{\omega}{N\alpha\delta}\right)\right|,
\]
for $\omega \in H \setminus B_3(0)$ and $N \geq 2$. This will allow us to deduce
\[
\left|\hat{r}\left(\frac{\omega - \nu_{\lambda_n}}{(N-1)\alpha\delta}\right)\right| \geq \left|\hat{r}\left(\frac{\omega}{N\alpha\delta}\right)\right|.
\]
for \( \omega \in S_{\lambda_{n},d}, \lambda_n \in \Lambda_n, \) and \( N \geq 2, \) where \( S_{\lambda_{n},d} = H_{A_{\lambda_n}} \backslash B_0(0) \), simply by noting that

\[
\left| \hat{r}_l \left( \frac{\omega - \nu_{\lambda_{n}}}{(N - 1) \alpha \delta} \right) \right|^l = \left( 1 - \left| \frac{A_{\lambda_n}(\omega' - \nu)}{(N - 1) \alpha \delta} \right| \right)^l
\]

\[
= \left( 1 - \left| \frac{\omega' - \nu}{(N - 1) \alpha \delta} \right| \right)^l = \left| \hat{r}_l \left( \frac{\omega' - \nu}{(N - 1) \alpha \delta} \right) \right|^l
\]

\[
\geq \left| \hat{r}_l \left( \frac{\omega'}{N \alpha \delta} \right) \right|^l = \left( 1 - \left| \frac{\omega'}{N \alpha \delta} \right| \right)^l
\]

\[
\geq \left( 1 - \frac{A_{\lambda_n} \omega'}{N \alpha \delta} \right)^l
\]

\[
(56)
\]

\[
(57)
\]

\[
(58)
\]

for \( \omega = A_{\lambda_n} \omega' \in H_{A_{\lambda_n}} \backslash B_3(0), \) where \( \omega' \in H \backslash B_3(0). \) Here, (56) and (58) are thanks to \( |\omega| = |A_{\lambda_n} \omega|, \) which is by \( A_{\lambda_n} \in O(d), \) and the inequality in (57) is due to (54). Insertion of (55) into (53) then yields

\[
h_{n,N,\alpha,\omega}(\omega) \leq \sum_{\lambda_n \in \Lambda_n} \left| \hat{g}_{\lambda_n}(\omega) \right|^2 \left( 1 - \left| \frac{\omega}{N \alpha \delta} \right| \right)^2
\]

\[
(59)
\]

\[
\leq \sum_{\lambda_n \in \Lambda_n} \left| \hat{g}_{\lambda_n}(\omega) \right|^2 \left( 1 - \left| \frac{\omega}{N \alpha \delta} \right| \right)^2
\]

\[
(60)
\]

for \( \omega \in \mathbb{R}^d, \) where in (59) we employed Assumption 1 and (60) is thanks to (18). This establishes (51) and completes the proof of (20) for \( \alpha = \log_2(\sqrt{d/(d - 1/2)}), d \geq 1. \)

It remains to show (54), which is accomplished through the following lemma.

Lemma 3. Let \( \alpha := \log_2(\sqrt{d/(d - 1/2)}) \), \( \hat{r}_l : \mathbb{R}^d \rightarrow \mathbb{R}, \)

\[
\hat{r}_l(\omega) := (1 - |\omega|)^l, \text{ with } l > \lfloor d/2 \rfloor + 1, \text{ and define } \nu = \mathbb{R}^d \text{ to have components } \nu_k = (1 + 2^{1/2})^\delta_\nu, \text{ for } k \in \{1, \ldots, d\}. \text{ Then,}
\]

\[
\left| \hat{r}_l \left( \frac{\omega - \nu}{(N - 1) \alpha \delta} \right) \right| \geq \left| \hat{r}_l \left( \frac{\omega}{N \alpha \delta} \right) \right|
\]

\[
(61)
\]

for \( \omega \in H \backslash B_0(0) \) and \( N \geq 2. \)

Proof. The key idea of the proof is to employ a monotonicity argument. Specifically, thanks to \( \hat{r}_l \) monotonically decreasing in \( |\omega|, \) i.e., \( \hat{r}_l(\omega_1) \geq \hat{r}_l(\omega_2), \) for \( \omega_1, \omega_2 \in \mathbb{R}^d \) with \( |\omega_2| \geq |\omega_1|, \)

\[
(61)
\]

for \( \omega \in H \backslash B_0(0) \) and \( N \geq 2. \) We first note that for \( \omega \in H \backslash B_0(0) \) with \( |\omega| > N^\alpha \delta, \)

\[
(61)
\]

is trivially satisfied as the RHS of (61) equals zero (owing to \( |\omega| > N^\alpha \delta \)) and together with \( \hat{r}_l(\omega) \geq B_0(0). \) It hence suffices to prove (62) for \( \omega \in H \) with \( \delta \leq |\omega| \leq N^\alpha \delta. \) To this end, fix \( \tau \in \sqrt{\delta, N^\alpha \delta}, \) and define the spherical segment \( \Xi_\tau := \{ \omega \in H \mid |\omega| = \tau \}. \) We then have

\[
\kappa_N(\omega) = \tau^2 \left| \frac{N - 1}{N} \right|^{2\alpha} - |\omega - \nu|^2
\]

\[
(63)
\]

negative (owing to \( \alpha = \log_2(\sqrt{d/(d - 1/2)}) > 0, \) for \( d \geq 1). \) Therefore, thanks to \( p_{\nu}, N \geq 2, \) beeing concave, establishing \( p_{\nu}(\delta) \geq 0, \) and \( p_{\nu}(N^\alpha \delta) \geq 0, \) for \( N \geq 2, \) implies \( p_{\nu}(\tau) \geq 0, \) for \( \tau \in \sqrt{\delta, N^\alpha \delta} \) and \( N \geq 2 \) (see Fig. 8), and thus (62), which completes the proof. It remains to show that \( p_{\nu}(\delta) \geq 0 \) and \( p_{\nu}(N^\alpha \delta) \geq 0, \) both for \( N \geq 2. \) We have

\[
p_{\nu}(\delta) = \delta^2 \left( \left| \frac{N - 1}{N} \right|^{2\alpha} - \frac{2(2 + 1/2) - (1 + 2^{-1/2})^2}{d} \right)
\]

\[
\geq \delta^2 \left( 2^{-2\alpha} - d - 1/2 \right) = 0,
\]

\[
(64)
\]

where the inequality in (64) is by

\[
n \mapsto \left| \frac{N - 1}{N} \right|^{2\alpha}, \quad N \geq 2.
\]
where, again, appropriate choice of the modulation factors \( \{\nu_{\lambda_n}\}_{\lambda_n \in \Lambda_n} \subseteq \mathbb{R}^d \) will be key in establishing the inductive step. We start by defining \( \Lambda_n^{+} \) to be the set of indices \( \lambda_n \in \Lambda_n \) such that \( \text{supp}(\chi_{\lambda_n}) \subseteq [\delta, \infty) \), and \( \Lambda_n^{-} \) to be the set of indices \( \lambda_n \in \Lambda_n \) so that \( \text{supp}(\chi_{\lambda_n}) \subseteq (-\infty, -\delta) \) (see Assumption \ref{assumption:pseudoband}). Clearly, \( \Lambda_n = \Lambda_n^{+} \cup \Lambda_n^{-} \). Moreover, we define the modulation factors according to \( \nu_{\lambda_n} := \alpha \), for \( \lambda_n \in \Lambda_n^{+} \), and \( \nu_{\lambda_n} := -\delta \), for \( \lambda_n \in \Lambda_n^{-} \). We then get

\[
h_{n,N,\alpha,\delta}(\omega) = \sum_{\lambda_n \in \Lambda_n} |\hat{g}_{\lambda_n}(\omega)|^2 \left(1 - \left|\hat{r}_i\left(\frac{\omega - \nu_{\lambda_n}}{(N-1)\delta}\right)\right|^2\right)
\]

for \( \nu_{\lambda_n} \in \mathbb{R}^d \), \( \hat{r}_i \) is quadratic in \( \nu_{\lambda_n} \), and the coefficient of the \( \nu_{\lambda_n} \) term is \( \frac{d}{\sqrt{d/\delta}} \). We proceed to sharpen, for \( d \geq 1 \), and complete the proof of statement i) in Theorem \ref{theorem:main_theorem}. We start by defining \( \Lambda_n^{+} \) to be the set of indices \( \lambda_n \in \Lambda_n \) such that \( \text{supp}(\chi_{\lambda_n}) \subseteq [\delta, \infty) \), and \( \Lambda_n^{-} \) to be the set of indices \( \lambda_n \in \Lambda_n \) so that \( \text{supp}(\chi_{\lambda_n}) \subseteq (-\infty, -\delta) \) (see Assumption \ref{assumption:pseudoband}). Clearly, \( \Lambda_n = \Lambda_n^{+} \cup \Lambda_n^{-} \). Moreover, we define the modulation factors according to \( \nu_{\lambda_n} := \alpha \), for \( \lambda_n \in \Lambda_n^{+} \), and \( \nu_{\lambda_n} := -\delta \), for \( \lambda_n \in \Lambda_n^{-} \). We then get

\[
h_{n,N,\alpha,\delta}(\omega) = \sum_{\lambda_n \in \Lambda_n} |\hat{g}_{\lambda_n}(\omega)|^2 \left(1 - \left|\hat{r}_i\left(\frac{\omega - \nu_{\lambda_n}}{(N-1)\delta}\right)\right|^2\right)
\]

for \( \nu_{\lambda_n} \in \mathbb{R}^d \), \( \hat{r}_i \) is quadratic in \( \nu_{\lambda_n} \), and the coefficient of the \( \nu_{\lambda_n} \) term is \( \frac{d}{\sqrt{d/\delta}} \). We proceed to sharpen, for \( d \geq 1 \), and complete the proof of statement i) in Theorem \ref{theorem:main_theorem}. We start by defining \( \Lambda_n^{+} \) to be the set of indices \( \lambda_n \in \Lambda_n \) such that \( \text{supp}(\chi_{\lambda_n}) \subseteq [\delta, \infty) \), and \( \Lambda_n^{-} \) to be the set of indices \( \lambda_n \in \Lambda_n \) so that \( \text{supp}(\chi_{\lambda_n}) \subseteq (-\infty, -\delta) \) (see Assumption \ref{assumption:pseudoband}). Clearly, \( \Lambda_n = \Lambda_n^{+} \cup \Lambda_n^{-} \). Moreover, we define the modulation factors according to \( \nu_{\lambda_n} := \alpha \), for \( \lambda_n \in \Lambda_n^{+} \), and \( \nu_{\lambda_n} := -\delta \), for \( \lambda_n \in \Lambda_n^{-} \). We then get

\[
h_{n,N,\alpha,\delta}(\omega) = \sum_{\lambda_n \in \Lambda_n} |\hat{g}_{\lambda_n}(\omega)|^2 \left(1 - \left|\hat{r}_i\left(\frac{\omega - \nu_{\lambda_n}}{(N-1)\delta}\right)\right|^2\right)
\]

for \( \nu_{\lambda_n} \in \mathbb{R}^d \), \( \hat{r}_i \) is quadratic in \( \nu_{\lambda_n} \), and the coefficient of the \( \nu_{\lambda_n} \) term is \( \frac{d}{\sqrt{d/\delta}} \). We proceed to sharpen, for \( d \geq 1 \), and complete the proof of statement i) in Theorem \ref{theorem:main_theorem}.
Proof. We first note that for \( \omega > N\delta \), (74) is trivially satisfied as the RHS of (74) equals zero (owing to \( \frac{1}{N\delta} > 1 \)) together with \( \text{supp}(\hat{\rho}) \subseteq B(0) \). It hence suffices to prove (74) for \( \delta \leq \omega \leq N\delta \). The key idea of the proof is to employ a monotonicity argument. Specifically, to \( \hat{\rho} \) monotonically decreasing in \( |\omega| \), i.e., \( \hat{\rho}(\omega_1) \geq \hat{\rho}(\omega_2) \), for \( \omega_1, \omega_2 \in \mathbb{R} \) with \( |\omega_2| \geq |\omega_1| \), (74) can be established simply by showing that
\[
\left| \frac{\omega - \delta}{(N - 1)\delta} \right| \leq \left| \frac{\omega}{N\delta} \right|, \quad \forall \omega \in [\delta, N\delta], \forall N \geq 2,
\]
which, by \( \omega \in [\delta, N\delta] \), is equivalent to
\[
\frac{\omega - \delta}{(N - 1)\delta} \leq \frac{\omega}{N\delta}, \quad \forall \omega \in [\delta, N\delta], \forall N \geq 2. \tag{75}
\]
Rearranging terms in (75), we get \( \omega \leq N\delta \), for \( \omega \in [\delta, N\delta] \) and \( N \geq 2 \), which completes the proof. \( \square \)

Remark 1. What makes the improved exponent \( \alpha \) possible in the 1-D case is the absence of rotated orthants. Specifically, for \( d = 1 \), the filters \( \{g_{\lambda n}\}_{\lambda_n \in \mathbb{A}_N} \) satisfy either \( \text{supp}(\hat{g}_{\lambda_n}) \subseteq (-\infty, -\delta] \) or \( \text{supp}(\hat{g}_{\lambda_n}) \subseteq [\delta, \infty) \), i.e., the support sets \( \text{supp}(\hat{g}_{\lambda_n}) \) are located in one of the two half-spaces.

APPENDIX D

PROOF OF STATEMENT II) IN THEOREM 1

We need to show that there exist \( C > 0 \) (that is independent of \( N \)) and \( N_0 \in \mathbb{N} \) such that
\[
W_N(f) \leq C B_0^N \int_{\mathbb{R}^d} N^{-\frac{d(2s+\beta)}{2(2s+\beta+d)}} , \quad \forall N \geq N_0,
\]
for \( f \in H^s(\mathbb{R}^d) \), where \( \beta > 0 \) depends on \( f \). Let us start by noting that for \( f \in H^s(\mathbb{R}^d) \), by \cite{41} Sec. 6.2.1) there exist \( \beta, \mu, L > 0 \) (all depending on \( f \)) such that
\[
|\hat{f}(\omega)| \leq \frac{\mu}{(1 + |\omega|^2)^{\frac{d+\beta}{2} + \frac{\beta}{2}}}, \quad \forall |\omega| \geq L. \tag{76}
\]
The key idea of the proof is to split the integral on the RHS of (20), i.e.,
\[
\int_{\mathbb{R}^d} |\hat{f}(\omega)|^2 \left( 1 - |\hat{\rho}(\omega/N\delta)| \right)^2 d\omega,
\]
into integrals over the sets \( B_{R_N}(0) \) and \( \mathbb{R}^d \setminus B_{R_N}(0) \), where
\[
R_N := N^{\frac{\alpha}{2(2s+\beta+d)}}, \quad N \in \mathbb{N},
\]
and to establish that
\[
\int_{B_{R_N}(0)} z(\omega) d\omega \leq 2 l\delta^{-1} \| f \|_2^2 N^{-\frac{d(2s+\beta)}{2(2s+\beta+d)}}, \tag{77}
\]
for \( N \geq 1 \), and
\[
\int_{\mathbb{R}^d \setminus B_{R_N}(0)} z(\omega) d\omega \leq \frac{\mu^2}{2s + \beta} \text{vol}^{d-1}(\partial B_1(0)) N^{-\frac{d(2s+\beta)}{2(2s+\beta+d)}}, \tag{78}
\]
for \( N \geq N_0 := \left[ L^{\frac{2(2s+\beta+1)}{\alpha}} \right] \). This then implies
\[
\int_{\mathbb{R}^d} z(\omega) d\omega \leq C N^{1 - \frac{\alpha}{2(2s+\beta)}} N^{-\frac{d(2s+\beta)}{2(2s+\beta+d)}},
\]
for \( N \geq \max \{1, N_0\} = N_0 \), where
\[
C := 2 \max \left\{ 2 l\delta^{-1} \| f \|_2^2, \text{vol}^{d-1}(\partial B_1(0))(2s + \beta)^{-1} \right\},
\]
and thereby completes the proof. It remains to establish (77) and (78). We start with (77) and note that
\[
(1 - 2|\omega|) \leq (1 - |\omega|)^d, \quad \omega \in \mathbb{R}^d, \tag{79}
\]
where \( l > [d/2] + 1 \), see Fig. 9. This implies
\[
1 - |\hat{\rho}(\omega/N\delta)|^2 = 1 - (1 - |\omega|/N\delta)^2 \leq \frac{2l}{N\delta |\omega|}, \quad \forall \omega \in \mathbb{R}^d,
\]
which when employed in \( \int_{B_{R_N}(0)} z(\omega) d\omega \) yields (77) thanks to
\[
\int_{B_{R_N}(0)} z(\omega) d\omega \leq \frac{2 l\delta^{-1}}{N\alpha \delta} \int_{B_{R_N}(0)} |\hat{f}(\omega)|^2 \frac{\omega}{|\omega|} d\omega \leq R_N \int_{B_{R_N}(0)} \frac{\| f \|_2^2}{2 l\delta^{-1} \| f \|_2^2 N^{-\frac{d(2s+\beta)}{2(2s+\beta+d)}}},
\]
for \( N \geq 1 \), where we used Parseval’s formula and \( R_N = N^{\frac{\alpha}{2(2s+\beta+d)}} \). Next, we establish (78) and start by noting that \( 1 - |\hat{\rho}(\omega/N\delta)|^2 \leq 1 \), for \( \omega \in \mathbb{R}^d \), which is by \( 0 \leq \hat{\rho}(\omega) \leq 1 \), for \( \omega \in \mathbb{R}^d \). Moreover, we have
\[
R_N = N^{\frac{\alpha}{2(2s+\beta+d)}} \geq N_0^{\frac{\alpha}{2(2s+\beta+d)}} = \left( \left[ L^{\frac{2(2s+\beta+1)}{\alpha}} \right]^{\frac{2(2s+\beta+d)}{2(2s+\beta+d)}} \right) \geq L,
\]
for \( N \geq N_0 \), which by (76) implies
\[
|\hat{f}(\omega)| \leq \frac{\mu}{(1 + |\omega|^2)^{\frac{d+\beta}{2} + \frac{\beta}{2}}}, \quad \forall |\omega| \geq R_N. \tag{80}
\]
Employing (80) in \( \int_{B_{R_N}(0)} z(\omega) d\omega \) now yields
\[
\int_{B_{R_N}(0)} z(\omega) d\omega \leq \int_{B_{R_N}(0)} \frac{\mu^2}{(1 + |\omega|^2)^{\frac{d+\beta}{2} + \frac{\beta}{2}}} d\omega = \int_{R^d \setminus B_{R_N}(0)} \frac{\mu^2}{(1 + r^2)^{\frac{d+\beta}{2} + \frac{\beta}{2}}} dr,
\]
where in (81) we introduced polar coordinates. Moreover, \( 1 + r^2)^{\frac{d+\beta}{2} + \frac{\beta}{2}} \geq r^{2s + d + \beta} \), for \( r \geq 0 \), which when employed in (81) yields (78) owing to
\[
\int_{\mathbb{R}^d \setminus B_{R_N}(0)} z(\omega) d\omega \leq \frac{\mu^2}{2s + \beta} R_N^{2s - \beta} \int_{R^d} r^{2s - \beta - 1} dr = \frac{\mu^2}{2s + \beta} R_N^{2s - \beta} \text{vol}^{d-1}(\partial B_1(0)) \int_{R^d} r^{2s - \beta - 1} dr,
\]
and
\[
\text{vol}^{d-1}(\partial B_1(0)) = N^{-\frac{\alpha}{2(2s+\beta)}},
\]
for \( N \geq N_0 \).
where in the last step we used \( R_N = N^{\frac{d}{d+1}} \). This completes the proof.

APPENDIX E

PROPOSITION 1

Proposition 1. Let \( \Omega \) be the module-sequence \( \{1\} \). Then,

\[
A_\Omega^N \|f\|_2^2 \leq \sum_{n=0}^{N-1} \|\Phi_\Omega^n(f)\|_2^2 + W_N(f) \leq B_\Omega^N \|f\|_2^2, \tag{82}
\]

for all \( f \in L^2(\mathbb{R}^d) \) and all \( N \geq 1 \), where \( A_\Omega^N = \prod_{k=1}^{N-1} \min\{1, A_k\} \) and \( B_\Omega^N = \prod_{k=1}^{N} \max\{1, B_k\} \).

Proof. We proceed by induction across \( N \) and start with the base case \( N = 1 \) which follows directly from the frame property (2) according to

\[
A_\Omega^1 \|f\|_2^2 = \min\{1, A_1\} \|f\|_2^2 \leq A_1 \|f\|_2^2 \leq B_1 \|f\|_2^2, \forall f \in L^2(\mathbb{R}^d).
\]

The inductive step is obtained as follows. Let \( N > 1 \) and suppose that (82) holds for \( N - 1 \), i.e.,

\[
A_\Omega^{N-1} \|f\|_2^2 \leq \sum_{n=0}^{N-2} \|\Phi_\Omega^n(f)\|_2^2 + W_{N-1}(f) \leq B_\Omega^{N-1} \|f\|_2^2, \forall f \in L^2(\mathbb{R}^d). \tag{83}
\]

We start by noting that

\[
\sum_{n=0}^{N-1} \|\Phi_\Omega^n(f)\|_2^2 + W_N(f) = \sum_{n=0}^{N-2} \|\Phi_\Omega^n(f)\|_2^2 + \sum_{q \in \Lambda^{N-1}} \|U[q]f\|_2^2 \geq \sum_{q \in \Lambda^N} \|U[q]f\|_2^2,
\]

and proceed by examining the third term on the RHS of (84). Every path \( \tilde{q} \in \Lambda^N = \Lambda_1 \times \cdots \times \Lambda_N - \Lambda_1 \times \Lambda_N \) of length \( N \) can be decomposed into a path \( q \in \Lambda^{N-1} \) of length \( N - 1 \) and an index \( \lambda_N \in \Lambda_N \) according to \( \tilde{q} = (q, \lambda_N) \). Thanks to (2) we have \( U[q] = U[\lambda_NU[q]] = U_N[\lambda_N]U[q] \), which yields

\[
\sum_{q \in \Lambda^N} \|U[q]f\|_2^2 = \sum_{q \in \Lambda^{N-1}} \sum_{\lambda_N \in \Lambda_N} \|U[q]f\|_2^2 \geq \sum_{\lambda_N \in \Lambda_N} \|U[q]f\|_2^2 \tag{85}
\]

Substituting the third term on the RHS of (84) by (85) and rearranging terms, we obtain

\[
\sum_{n=0}^{N-1} \|\Phi_\Omega^n(f)\|_2^2 + W_N(f) = \sum_{n=0}^{N-2} \|\Phi_\Omega^n(f)\|_2^2 + \sum_{q \in \Lambda^{N-1}} \|U[q]f\|_2^2 \geq \sum_{\lambda_N \in \Lambda_N} \|U[q]f\|_2^2 = \rho_N(U[q]f).
\]

Thanks to the frame property (2) and \( U[q]f \in L^2(\mathbb{R}^d) \), which is by [9, Eq. 16], we have \( A_N \|U[q]f\|_2^2 \leq \rho_N(U[q]f) \leq B_N \|U[q]f\|_2^2 \), and thus

\[
\min\{1, A_N\} \left( \sum_{n=0}^{N-2} \|\Phi_\Omega^n(f)\|_2^2 + W_{N-1}(f) \right) \leq \sum_{n=0}^{N-1} \|\Phi_\Omega^n(f)\|_2^2 + W_N(f) \leq \max\{1, B_N\} \left( \sum_{n=0}^{N-2} \|\Phi_\Omega^n(f)\|_2^2 + W_{N-1}(f) \right), \tag{86}
\]

where we employed the identity \( \sum_{q \in \Lambda^{N-1}} \|U[q]f\|_2^2 = W_{N-1}(f) \). Invoking the induction hypothesis (83) in (86) and (87) and noting that \( A_\Omega^N = \min\{1, A_N\} A_\Omega^{N-1} \) and \( B_\Omega^N = \max\{1, B_N\} B_\Omega^{N-1} \) completes the proof. \( \square \)

APPENDIX F

PROPOSITION 2

Proposition 2. Let \( \tilde{r}_i : \mathbb{R}^d \rightarrow \mathbb{R}, \tilde{r}_i(\omega) := (1 - |\omega|)^l_i \), with \( l > |d/2| + 1 \), and let \( \alpha := \log_2(\sqrt{d/(d-1/2)}) \). Then, we have

\[
\lim_{N \rightarrow \infty} \int_{\mathbb{R}^d} \left| \tilde{f}(\omega) \right|^2 \left( 1 - \left| \tilde{r}_i \left( \frac{\omega}{N^{\alpha}} \right) \right|^2 \right) \text{d}\omega = 0, \tag{88}
\]

for all \( f \in L^2(\mathbb{R}^d) \).

Proof. We start by setting

\[
d_{N, \alpha, \delta}(\omega) := \left( 1 - \left| \tilde{r}_i \left( \frac{\omega}{N^{\alpha}} \right) \right|^2 \right), \quad \omega \in \mathbb{R}^d, N \in \mathbb{N}.
\]

Let \( f \in L^2(\mathbb{R}^d) \). For every \( \varepsilon > 0 \) there exists \( R > 0 \) such that

\[
\int_{\mathbb{R}^d \setminus B_R(0)} \left| \tilde{f}(\omega) \right|^2 \text{d}\omega \leq \varepsilon/2,
\]

where \( B_R(0) \) denotes the closed ball of radius \( R \) centered at the origin. Next, we employ Dini’s theorem [52, Theorem 7.3] to show that \( d_{N, \alpha, \delta} \) converges to zero uniformly on \( B_R(0) \). To this end, we note that (i) \( d_{N, \alpha, \delta} \) is continuous as a composition of continuous functions, (ii) \( d_{N, \alpha, \delta}(\omega) = 0 \), for \( \omega = 1 \), is clearly continuous, (iii) \( d_{N, \alpha, \delta}(\omega) \leq d_{N-1, \alpha, \delta}(\omega) \), for \( \omega \in \mathbb{R}^d \) and \( N \in \mathbb{N} \), and (iv) \( d_{N, \alpha, \delta} \) converges to \( z_0 \) pointwise on \( B_R(0) \), i.e.,

\[
\lim_{N \rightarrow \infty} d_{N, \alpha, \delta}(\omega) = z_0(\omega) = 0, \quad \forall \omega \in \mathbb{R}^d.
\]

This allows us to conclude that there exists \( N_0 \in \mathbb{N} \) (that depends on \( \varepsilon \)) such that \( d_{N, \alpha, \delta}(\omega) \leq \frac{\varepsilon}{2\|f\|_2^2} \), for \( \omega \in B_R(0) \) and \( N \geq N_0 \), and we therefore get

\[
\int_{\mathbb{R}^d \setminus B_R(0)} \left| \tilde{f}(\omega) \right|^2 d_{N, \alpha, \delta}(\omega) \text{d}\omega \leq \int_{\mathbb{R}^d \setminus B_R(0)} \left| \tilde{f}(\omega) \right|^2 d_{N_0, \alpha, \delta}(\omega) \text{d}\omega \leq \varepsilon/2
\]

and

\[
\int_{B_R(0)} \left| \tilde{f}(\omega) \right|^2 d_{N_0, \alpha, \delta}(\omega) \text{d}\omega \leq \varepsilon/2 + \frac{\varepsilon}{2\|f\|_2^2} \| \tilde{f} \|_2^2 = \varepsilon,
\]

as desired. \( \square \)
where in the last step we employed Parseval’s formula. Since \( \varepsilon > 0 \) was arbitrary, we have (88), which completes the proof. \( \square \)

**Appendix G**

**Proof of Theorem 2**

We start by establishing (27) in statement i). The structure of the proof is similar to that of the proof of statement i) in Theorem 1 specifically we perform induction over \( N \). Starting with the base case \( N = 1 \), we first note that \( \text{supp}(\hat{\psi}) \subseteq [1/2, 2] \), \( \hat{g}_j(\omega) = \hat{\psi}(2^{-j}\omega) \), \( j \geq 1 \), and \( \hat{g}_j(\omega) = \hat{\psi}(-2^{-j}\omega) \), \( j \leq 1 \), all by assumption, imply

\[
\text{supp}(\hat{g}_j) = \text{supp}(\hat{\psi}(2^{-j} \cdot)) \subseteq [2^{-j-1}, 2^{j+1}] \}, \quad j \geq 1, \quad (89)
\]

and

\[
\text{supp}(\hat{g}_j) = \text{supp}(\hat{\psi}(2^{-j} \cdot)) \subseteq [-2^{j+1}, -2^{j-1}] \}, \quad j \leq -1
\]

for \( j \leq -1 \). We then get

\[
W_1(f) = \sum_{j \in \mathbb{Z} \setminus \{0\}} \|f * g_j\|^2_2 \leq \int_{\mathbb{R}} \sum_{j \in \mathbb{Z} \setminus \{0\}} |\hat{g}_j(\omega)|^2 |\hat{f}(\omega)|^2 d\omega = \int_{\mathbb{R}} \sum_{j = 1} \|\hat{\psi}(2^{-j} \cdot)\|^2 |\hat{f}(\omega)|^2 d\omega + \sum_{j = -1} |\hat{\psi}(-2^{-j} \cdot)|^2 |\hat{f}(\omega)|^2 d\omega = \int_{\mathbb{R}} \sum_{j = 1} |\hat{\psi}(2^{-j} \cdot)|^2 |\hat{f}(\omega)|^2 d\omega + \sum_{j = -1} |\hat{\psi}(-2^{-j} \cdot)|^2 |\hat{f}(\omega)|^2 d\omega = \int_{\mathbb{R}} |\hat{\psi}(2^{-j} \cdot)|^2 |\hat{f}(\omega)|^2 d\omega + \int_{\mathbb{R}} |\hat{\psi}(-2^{-j} \cdot)|^2 |\hat{f}(\omega)|^2 d\omega \leq \int_{\mathbb{R}} |\hat{\psi}(2^{-j} \cdot)|^2 |\hat{f}(\omega)|^2 d\omega + \int_{\mathbb{R}} |\hat{\psi}(-2^{-j} \cdot)|^2 |\hat{f}(\omega)|^2 d\omega \leq \int_{\mathbb{R}} |\hat{\psi}(\omega)|^2 (1 - |\hat{f}(\omega)|^2) d\omega = \left( 1 - \left| \hat{f}(\omega) \right|^2 \right) \right)^2, \quad (90)
\]

where (91) is by Parseval’s formula, and (92) is thanks to (89) and (90). The inequality in (93) is owing to (26), and (94) is due to \( \text{supp}(\hat{f}) \subseteq [-1, 1] \) and \( 0 \leq |\hat{f}(\omega)| \leq 1 \), for \( \omega \in \mathbb{R} \). The inductive step is obtained as follows. Let \( N > 1 \) and suppose that (27) holds for \( N - 1 \), i.e.,

\[
W_{N-1}(f) = \int_{\mathbb{R}} |\hat{\psi}(\omega)|^2 \left( 1 - \left| \hat{f}(\omega) \right|^2 \right) ^2 d\omega, \quad (95)
\]

for \( f \in L^2(\mathbb{R}) \). We start by noting that every path \( \tilde{q} \in (\mathbb{Z} \setminus \{0\})^N \) of length \( N \) can be decomposed into a path \( q \in (\mathbb{Z} \setminus \{0\})^{N-1} \) of length \( N - 1 \) and an index \( j \in \mathbb{Z} \setminus \{0\} \) according to \( \tilde{q} = (j, q) \). Thanks to (4) we have \( U[|q] = U(j, q) = U[q]U[j] \), which yields

\[
W_N(f) = \sum_{j \in \mathbb{Z} \setminus \{0\}} \sum_{q \in (\mathbb{Z} \setminus \{0\})^{N-1}} |||U[q]|(U[j]f)||^2_2 = \sum_{j \in \mathbb{Z} \setminus \{0\}} W_{N-1}(U[j]f). \quad (96)
\]

We proceed by examining the term \( W_{N-1}(U[j]f) \) inside the sum on the RHS of (96). Invoking the induction hypothesis (95) and employing Parseval’s formula, we get

\[
W_{N-1}(U[j]f) \leq \int_{\mathbb{R}} |\hat{U[j]f}(\omega)|^2 \left( 1 - \left| \hat{f}(\frac{\omega - \nu_j}{(5/3)^{N-2}}) \right|^2 \right) d\omega = \left( \left| \hat{f}(\omega) \right|^2 - \left| \hat{f}(\omega - \nu_j) \right|^2 \right) \right)^2, \quad (97)
\]

where \( \nu_j = \frac{2^j \eta}{5/3} \) is the inverse Fourier transform of \( \hat{f}(\frac{\omega - \nu_j}{(5/3)^{N-2}}) \). Next, we note that \( \hat{f}(\frac{\omega - \nu_j}{(5/3)^{N-2}}) \) is a positive definite radial basis function \cite{50} Theorem 6.20 and hence by \cite{50} Theorem 6.18 \( \nu_j \in (\mathbb{Z} \setminus \{0\}) \subseteq \mathbb{R} \). Choosing the modulation factors \( \{\nu_j\}_{j \in \mathbb{Z} \setminus \{0\}} \subseteq \mathbb{R} \) appropriately (see (102) below) will be key in establishing the inductive step. Using (97) and (98) to upper-bound the term \( W_{N-1}(U[j]f) \) inside the sum on the RHS of (96) yields

\[
W_N(f) \leq \sum_{j \in \mathbb{Z} \setminus \{0\}} \left( \left| \hat{f}(\omega) \right|^2 - \left| \hat{f}(\omega - \nu_j) \right|^2 \right) \right)^2, \quad (99)
\]

where

\[
h_{l,N-2}(\omega) \leq \left( 1 - \left| \hat{f}(\omega) \right|^2 \right) \right)^2, \quad \forall \omega \in \mathbb{R}, \quad (101)
\]

In (99) we employed Parseval’s formula together with \( \mathcal{M}_\omega f = T_\omega f \), for \( f \in L^2(\mathbb{R}) \) and \( \omega \in \mathbb{R} \). The key step is now to establish—by judiciously choosing \( \{\nu_j\}_{j \in \mathbb{Z} \setminus \{0\}} \subseteq \mathbb{R} \)—the upper bound

\[
h_{l,N-2}(\omega) \leq \left( 1 - \left| \hat{f}(\frac{\omega - \nu_j}{(5/3)^{N-2}}) \right|^2 \right) \right)^2, \quad \forall \omega \in \mathbb{R}, \quad (102)
\]

which then yields (27) and thereby completes the proof. To this end, we set \( \eta := \frac{5}{2} \).

\[
\nu_j := \frac{2^j \eta}{5/3}, \quad j \geq 1, \quad \nu_j := -\frac{2^j \eta}{5/3}, \quad j \leq -1, \quad (102)
\]

and note that it suffices to prove (101) for \( \omega \geq 0 \), as

\[
h_{l,N-2}(\omega) = \sum_{j \in \mathbb{Z} \setminus \{0\}} \left| \hat{g}_j(\omega) \right|^2 \left( 1 - \left| \hat{f}(\omega - \nu_j) \right|^2 \right) \right)^2, \quad (103)
\]

\[
= \sum_{j = -1} \left| \hat{g}_j(\omega) \right|^2 \left( 1 - \left| \hat{f}(\omega + \nu_j) \right|^2 \right) \right)^2, \quad (104)
\]

\[
= \sum_{j = 1} \left| \hat{g}_j(\omega) \right|^2 \left( 1 - \left| \hat{f}(\omega - \nu_j) \right|^2 \right) \right)^2, \quad (105)
\]

\[
= h_{l,N-2}(\omega), \quad \forall \omega \geq 0. \quad (105)
\]
Here, \( \text{(103)} \) is thanks to \( \hat{g}_j(-\omega) = 0 \), for \( j \geq 1 \) and \( \omega \geq 0 \), which is by \( \text{(89)} \), and \( \text{(105)} \) is owing to \( \hat{g}_j(\omega) = 0 \), for \( j \leq -1 \) and \( \omega \geq 0 \), which is by \( \text{(90)} \). Moreover, in \( \text{(103)} \) we used \( \hat{r}_1(-\omega) = \hat{r}_1(\omega) \), for \( \omega \in \mathbb{R} \), and \( \text{(104)} \) is thanks to
\[
\hat{g}_{-j}(-\omega) = \hat{\psi}(2^{-j-1}\omega) = \hat{\psi}(2^{-j}\omega) = \hat{g}_j(\omega),
\]
for \( \omega \in \mathbb{R} \) and \( j \geq 1 \), as well as
\[
\nu_{-j} = -2^j \eta = -\nu_j, \quad j \geq 1.
\]
Now, let \( \omega \in [0, 1] \), and note that
\[
\begin{align*}
\hat{h}_{1,N-2}(\omega) &= \sum_{j \in \mathbb{Z} \setminus \{0\}} |\hat{g}_j(\omega)|^2 \left( 1 - |\hat{r}_1\left(\frac{\omega - \nu_j}{(5/3)^{N-2}}\right)|^2 \right) \\
&\leq 1 - |\hat{r}_1\left(\frac{\omega - \eta}{(5/3)^{N-1}}\right)|^2, \quad \forall N \geq 2,
\end{align*}
\]
(106)
where the second equality in \( \text{(106)} \) is simply a consequence of \( \hat{g}_j(\omega) = 0 \), for \( j \in \mathbb{Z} \setminus \{0\} \) and \( \omega \in [0, 1] \), which, in turn, is by \( \text{(89)} \) and \( \text{(90)} \). The inequality in \( \text{(106)} \) is thanks to \( 0 \leq \hat{r}_1(\omega) \leq 1 \), for \( \omega \in \mathbb{R} \). Next, let \( \omega \in [1, 2] \). Then, we have
\[
\begin{align*}
\hat{h}_{1,N-2}(\omega) &= |\hat{g}_1(\omega)|^2 \left( 1 - |\hat{r}_1\left(\frac{\omega - 2\eta}{(5/3)^{N-2}}\right)|^2 \right) \\
&\leq |\hat{g}_1(\omega)|^2 \left( 1 - |\hat{r}_1\left(\frac{\omega - 2\eta}{(5/3)^{N-2}}\right)|^2 \right) \\
&+ \left( 1 - |\hat{g}_1(\omega)|^2 \right) \left( 1 - |\hat{r}_1\left(\frac{\omega - \eta}{(5/3)^{N-2}}\right)|^2 \right) \\
&= 1 - |\hat{r}_1\left(\frac{\omega - \eta}{(5/3)^{N-2}}\right)|^2 + |\hat{g}_1(\omega)|^2 \left| \hat{r}_1\left(\frac{\omega - \eta}{(5/3)^{N-2}}\right) \right|^2 \\
&- |\hat{r}_1\left(\frac{\omega - 2\eta}{(5/3)^{N-2}}\right)\right|^2,
\end{align*}
\]
(107)
where \( \text{(107)} \) is thanks to \( \hat{g}_j(\omega) = 0 \), for \( j \in \mathbb{Z} \setminus \{0, 1\} \) and \( \omega \in [1, 2] \), which, in turn, is by \( \text{(89)} \) and \( \text{(90)} \). Moreover, \( \text{(108)} \) is owing to
\[
|\hat{g}_1(\omega)|^2 \in [0, 1],
\]
which, in turn, is by \( \text{(26)} \) and \( 0 \leq \hat{r}_1(\omega) \leq 1 \), for \( \omega \in \mathbb{R} \). Next, fix \( j \geq 2 \) and let \( \omega \in [2^{j-1}, 2^j] \). Then, we have
\[
\begin{align*}
\hat{h}_{1,N-2}(\omega) &= |\hat{g}_j(\omega)|^2 \left( 1 - |\hat{r}_1\left(\frac{\omega - 2^j \eta}{(5/3)^{N-2}}\right)|^2 \right) \\
&+ \left( 1 - |\hat{g}_j(\omega)|^2 \right) \left( 1 - |\hat{r}_1\left(\frac{\omega - 2^j \eta}{(5/3)^{N-2}}\right)|^2 \right) \\
&= \left( 1 - |\hat{g}_j(\omega)|^2 \right) \left( \hat{r}_1\left(\frac{\omega - 2^j \eta}{(5/3)^{N-2}}\right) \right)^2 \\
&\leq 1 - |\hat{r}_1\left(\frac{\omega - 2^j \eta}{(5/3)^{N-2}}\right)|^2 + |\hat{g}_j(\omega)|^2 \left| \hat{r}_1\left(\frac{\omega - 2^j \eta}{(5/3)^{N-2}}\right) \right|^2 \\
&- |\hat{r}_1\left(\frac{\omega - 2^j \eta}{(5/3)^{N-2}}\right)|^2,
\end{align*}
\]
(110)
where \( \text{(110)} \) is thanks to \( \hat{g}_{j'}(\omega) = 0 \), for \( j' \in \mathbb{Z} \setminus \{0, j, j - 1\} \) and \( \omega \in [2^{j-1}, 2^j] \), which, in turn, is by \( \text{(89)} \) and \( \text{(90)} \), and ii) \( |\hat{\phi}(\omega)|^2 + |\hat{g}_{-j}(\omega)|^2 + |\hat{g}_j(\omega)|^2 = 1 \), (112)
for \( \omega \in [2^{j-1}, 2^j] \), which is a consequence of the Littlewood-Paley condition \( \text{(26)} \) and of \( \text{(89)} \) and \( \text{(90)} \). It follows from
\[
\hat{f}(\omega) \leq \frac{\mu}{\left(1 + |\omega|^2\right)^{\frac{3}{2} + \frac{1}{4} + \frac{1}{4}}, \quad \forall |\omega| \geq L.
\]
(113)
Without loss of generality, we can assume that $L \geq 1$ (if $L < 1$, the upper bound (113) clearly holds for $\omega \in \mathbb{R}$ with $|\omega| \geq 1 > L$). The key idea of the proof is to split the integral on the RHS of (27), i.e.,

$$
\int_{\mathbb{R}} \left| \hat{f}(\omega) \right|^2 \left(1 - \left| \hat{r}_1 \left( \frac{\omega}{(5/3)^{N-1}} \right) \right|^2 \right) d\omega,
$$

into integrals over the sets $B_{R_N}(0)$ and $\mathbb{R} \setminus B_{R_N}(0)$, where

$$
R_N := (5/3)^{\frac{N-1}{2+\beta}}, \quad N \in \mathbb{N},
$$

and to establish that

$$
\int_{B_{R_N}(0)} z(\omega) d\omega \leq 2\|f\|_2^2 \left(\frac{(N-1)(2+\beta)}{2+\beta+1}\right)^2 \omega \left(\frac{5/3}{(5/3)^{N-1}} \right)^{2l}, \quad (114)
$$

for $N \geq 1$, and

$$
\int_{\mathbb{R} \setminus B_{R_N}(0)} z(\omega) d\omega \leq \frac{2\mu^2}{2s + \beta} \frac{(N-1)(2+\beta)}{2+\beta+1}, \quad (115)
$$

for $N \geq N_0 := \left\lceil (2s+\beta+1) \log_{5/3}(L+1) \right\rceil$. This then implies

$$
\int_{\mathbb{R}} z(\omega) d\omega \leq C (5/3)^{-\frac{N+2+\beta}{2+\beta+1}},
$$

for $N \geq \max\{1, N_0\} = N_0$, where

$$
C := 4 \cdot (5/3)^{2+\beta+1} \max\left\{ \|f\|_2^2, \mu^2 (2s + \beta)^{-1} \right\},
$$

and thereby completes the proof. It remains to establish and note that it suffices to establish (121) for

$$
\int_{\mathbb{R} \setminus B_{R_N}(0)} z(\omega) d\omega \text{ yields } (114) \text{ thanks to }
$$

$$
\int_{B_{R_N}(0)} z(\omega) d\omega \leq \frac{2l}{(5/3)^{N-1}} \int_{B_{R_N}(0)} |\hat{f}(\omega)|^2 |\omega| d\omega \leq R_N
$$

for $N \geq 1$, where we used Parseval’s formula as well as $R_N = (5/3)^{\frac{N-1}{2+\beta}}$. Next, we establish (115) and start by noting that $1 - \left| \hat{r}_1 \left( \frac{\omega}{(5/3)^{N-1}} \right) \right|^2 \leq 1$, for $\omega \in \mathbb{R}$, which is by $0 \leq \hat{r}_1(\omega) \leq 1$, for $\omega \in \mathbb{R}$. Moreover, we have

$$
R_N = (5/3)^{\frac{N-1}{2+\beta}} \geq (5/3)^{\frac{N-1}{2+\beta} \leq (2s+\beta+1) \log_{5/3}(L+1)-1},
$$

$$
= (5/3)^{\log_{5/3}(L)}, \quad \forall N \geq N_0,
$$

which by (113) implies

$$
|\hat{f}(\omega)| \leq \frac{\mu}{(1 + |\omega|^2)^{\frac{1}{2} + \frac{\beta}{4}}}, \quad \forall |\omega| \geq R_N, \quad (116)
$$

Employing (116) in $\int_{\mathbb{R} \setminus B_{R_N}(0)} z(\omega) d\omega$ now yields

$$
\int_{\mathbb{R} \setminus B_{R_N}(0)} z(\omega) d\omega \leq \int_{\mathbb{R} \setminus B_{R_N}(0)} \frac{\mu^2}{(1 + |\omega|^2)^{s+\beta/2}} d\omega
$$

$$
= 2\mu^2 \int_{R_N}^\infty \left(1 + r^2\right)^{s+\beta/2} dr, \quad (117)
$$

where in the last step we introduced polar coordinates. Moreover,

$$
(1 + r^2)^{s+\beta/2} \geq r^{s+\beta}, \quad r \geq 0,
$$

which when employed in (117) yields (115) owing to

$$
\int_{\mathbb{R} \setminus B_{R_N}(0)} z(\omega) d\omega \leq 2\mu^2 \int_{R_N}^\infty r^{-2s-\beta-1} dr
$$

$$
= \frac{2\mu^2}{2s + \beta} \frac{(5/3)^{2s+\beta+1}}{2+\beta},$$

where in the last step we used $R_N = (5/3)^{\frac{N-1}{2+\beta}}$. This completes the proof of statement i).

We proceed to the proof of statement ii), again, effected by induction over $N$. Specifically, we first establish (30) by employing the same arguments as those leading to (99) with $(5/3)^{N-2}$ replaced by $(3/2)^{N-2} R$. With this replacement $h_{1,N-2}$ in (100) becomes

$$
h_{1,N-2}(\omega) := \sum_{k \in \mathbb{Z} \setminus \{0\}} |\hat{g}_k(\omega)|^2 \left(1 - \left| \hat{r}_1 \left( \frac{\omega - \nu_k}{(3/2)^{N-2} R} \right) \right|^2 \right), \quad (118)
$$

while, again, judicious choice of the modulation factors $\{\nu_k\}_{k \in \mathbb{Z} \setminus \{0\}} \subseteq \mathbb{R}$ (see (122) below) will be key in establishing the inductive step. Here, we note that the functions $\hat{g}_k$ in (118) satisfy $\hat{g}_k(\omega) = \hat{g}(\omega - R(k + 1))$, for $k \geq 1$, $\hat{g}_k(\omega) = \hat{g}(\omega + R(|k| + 1))$, for $k \leq -1$, both by assumption, as well as

$$
\text{supp}(\hat{g}_k) = \text{supp}(\hat{g}(\cdot - R(k + 1))) \subseteq [Rk, R(k + 2)], \quad k \geq 1, \quad (119)
$$

and

$$
\text{supp}(\hat{g}_k) = \text{supp}(\hat{g}(\cdot + R(|k| + 1))) \subseteq [-R(|k| + 2), -R|k|], \quad k \leq -1, \quad (120)
$$

where (119) and (120) follow from

$$
\text{supp}(\hat{g}) \subseteq [-R, R],
$$

which again is by assumption. It remains to establish the equivalent of (101), namely

$$
h_{1,N-2}(\omega) \leq 1 - \left| \hat{r}_1 \left( \frac{\omega}{(3/2)^{N-2} R} \right) \right|^2, \quad \forall \omega \in \mathbb{R}, \quad (121)
$$

To this end, we set $\eta := \frac{3}{2} R$, $\nu_k := Rk + \eta$, $\forall k \geq 1$, $\nu_k := -\nu_{|k|}$, $\forall k \leq -1$, (122) and note that it suffices to establish (121) for $\omega \geq 0$, thanks to
\[ h_{l,N-2}(\omega) = \sum_{k \in \mathbb{Z} \setminus \{0\}} |g_k(\omega)|^2 \left( 1 - \left| \tilde{r}_l \left( \frac{\omega - \nu_k}{(3/2)^2 N^{-2} R} \right) \right|^2 \right) \]
\[ = \sum_{k \geq -1} |g_k(\omega)|^2 \left( 1 - \left| \tilde{r}_l \left( \frac{\omega + \nu_k}{(3/2)^2 N^{-2} R} \right) \right|^2 \right) \tag{123} \]
\[ = \sum_{k \geq 1} |g_{-k}(\omega)|^2 \left( 1 - \left| \tilde{r}_l \left( \frac{\omega + \nu_k}{(3/2)^2 N^{-2} R} \right) \right|^2 \right) \]
\[ = \sum_{k \geq 1} |g_{-k}(\omega)|^2 \left( 1 - \left| \tilde{r}_l \left( \frac{\omega - \nu_k}{(3/2)^2 N^{-2} R} \right) \right|^2 \right) \tag{124} \]
\[ = h_{l,N-2}(\omega), \quad \forall \omega \geq 0. \tag{125} \]

Here, (123) follows from \( \tilde{g}_k(\omega) = 0 \), for \( k \geq 1 \) and \( \omega \geq 0 \), which, in turn, is by (119), and (125) is owing to \( \tilde{g}_0(\omega) = 0 \), for \( k \leq -1 \) and \( \omega \geq 0 \), which is by (120). Moreover, in (123) we used \( \tilde{r}_l(\omega) = \tilde{r}_l(\omega) \), for \( \omega \in \mathbb{R} \), and (124) is thanks to \( \nu_{-k} = -\nu_k \), for \( k \geq 1 \), and
\[ \tilde{g}_{-k}(\omega) = \tilde{g}_{-k}(\omega + \mathcal{R}(-k + 1)) = \tilde{g}_{-k}(\omega + \mathcal{R}(k + 1)) = \tilde{g}_k(\omega), \quad \forall \omega \in \mathbb{R}, \forall k \geq 1, \]
where we used \( \tilde{g}(\omega) = \tilde{g}(\omega) \), for \( \omega \in \mathbb{R} \), which is by assumption. Now, let \( \omega \in [0, R] \), and note that
\[ h_{l,N-2}(\omega) = 0 \leq 1 - \left| \tilde{r}_l \left( \frac{\omega}{(3/2)^2 N^{-1} R} \right) \right|^2, \tag{126} \]
for \( N \geq 2 \), where the equality in (126) is a consequence of (119) and (120), and the inequality is thanks to \( 0 \leq \tilde{r}_l(\omega) \leq 1 \), for \( \omega \in \mathbb{R} \). Next, let \( \omega \in [R, 2R) \), and then we have
\[ h_{l,N-2}(\omega) = |g_1(\omega)|^2 \left( 1 - \left| \tilde{r}_l \left( \frac{\omega - \nu_1}{(3/2)^2 N^{-2} R} \right) \right|^2 \right) \tag{127} \]
\[ \leq |g_1(\omega)|^2 \left( 1 - \left| \tilde{r}_l \left( \frac{\omega - \nu_1}{(3/2)^2 N^{-2} R} \right) \right|^2 \right) \]
\[ + (1 - |g_1(\omega)|^2) \left( 1 - \left| \tilde{r}_l \left( \frac{\omega - \eta}{(3/2)^2 N^{-2} R} \right) \right|^2 \right) \geq 0, \tag{128} \]
\[ = 1 - \left| \tilde{r}_l \left( \frac{\omega - \eta}{(3/2)^2 N^{-2} R} \right) \right|^2 + |g_1(\omega)|^2 \left( \left| \tilde{r}_l \left( \frac{\omega - \eta}{(3/2)^2 N^{-2} R} \right) \right|^2 \right) \]
\[ - \left| \tilde{r}_l \left( \frac{\omega - \nu_1}{(3/2)^2 N^{-2} R} \right) \right|^2 \right) \tag{129} \]
where (127) is thanks to \( \tilde{g}_k(\omega) = 0 \), for \( k \in \mathbb{Z} \setminus \{0\} \) and \( \omega \in [0, R] \), which, in turn, is by (119) and (120). Moreover, (128) is owing to \( |g_1(\omega)|^2 \leq 1 \), which, in turn, is by (29), and \( 0 \leq \tilde{r}_l(\omega) \leq 1 \), for \( \omega \in \mathbb{R} \). Next, fix \( k \geq 2 \), and let \( \omega \in [Rk, R(k+1)] \). Then, we have
\[ h_{l,N-2}(\omega) = |g_k(\omega)|^2 \left( 1 - \left| \tilde{r}_l \left( \frac{\omega - \nu_k}{(3/2)^2 N^{-2} R} \right) \right|^2 \right) \tag{130} \]
\[ + |g_{k-1}(-\omega)|^2 \left( 1 - \left| \tilde{r}_l \left( \frac{\omega - \nu_{k-1}}{(3/2)^2 N^{-2} R} \right) \right|^2 \right) \]
\[ \leq 1 - \left| \tilde{r}_l \left( \frac{\omega - \nu_{k-1}}{(3/2)^2 N^{-2} R} \right) \right|^2 + |g_k(\omega)|^2 \left( \left| \tilde{r}_l \left( \frac{\omega - \nu_{k-1}}{(3/2)^2 N^{-2} R} \right) \right|^2 \right) \]
\[ - \left| \tilde{r}_l \left( \frac{\omega - \nu_k}{(3/2)^2 N^{-2} R} \right) \right|^2 \right) \tag{131} \]
where (130) is thanks to \( \tilde{g}_k(\omega) = 0 \), for \( k \in \mathbb{Z} \setminus \{0, k-1\} \) and \( \omega \in [Rk, R(k+1)] \), and \( \omega \in [Rk, R(k+1)] \), which, in turn, is by (119) and (120), and ii)
\[ |\tilde{\omega}(\omega)|^2 + |g_{k-1}(\omega)|^2 + |g_k(\omega)|^2 = 1, \tag{132} \]
for \( \omega \in [Rk, R(k+1)] \), which is a consequence of the Littlewood-Paley condition (29) and of (119) and (120). It follows from (129) and (131) that for \( k \geq 1 \), we have
\[ h_{l,N-2}(\omega) \leq 1 - \left| \tilde{r}_l \left( \frac{\omega - \nu_{k-1}}{(3/2)^2 N^{-2} R} \right) \right|^2 + |g_k(\omega)|^2 \left( \left| \tilde{r}_l \left( \frac{\omega - \nu_{k-1}}{(3/2)^2 N^{-2} R} \right) \right|^2 \right) \]
\[ = s(\omega), \]
of statement ii) in Theorem I in Appendix D. We start by noting that (31) amounts to the existence of $C > 0$ (that is independent of $N$) and $N_0 \in \mathbb{N}$ such that
$$ W_N(f) \leq C(3/2)^{-N/(2r+\beta)} \frac{1}{x^{r+\beta}}, \quad \forall N \geq N_0, $$
for $f \in H^s(\mathbb{R})$, where $\beta > 0$ depends on $f$. Now, for $f \in H^s(\mathbb{R})$, by [41, Sec. 6.2.1] there exist $\beta, \mu, L > 0$ (all depending on $f$) such that
$$ |\hat{f}(\omega)| \leq \frac{\mu}{(1 + |\omega|^2)^{\frac{1}{2} + \frac{\mu}{4} + \frac{\beta}{4}}}, \quad \forall |\omega| \geq L. \quad (133) $$
Without loss of generality, we can assume that $L \geq 1$ (if $L < 1$, the upper bound (133) clearly holds for $\omega \in \mathbb{R}$ with $|\omega| \geq 1 > L$). The key idea of the proof is to split the integral on the RHS of (30), i.e.,
$$ \int_{\mathbb{R}} |\hat{f}(\omega)|^2 \left(1 - \left|\hat{r}_l \left(\frac{\omega}{(3/2)^N-1}R\right)\right|^2\right) d\omega, $$
into integrals over the sets $B_{R_N}(0)$ and $\mathbb{R} \setminus B_{R_N}(0)$, where
$$ R_N := (3/2)^{-N/(2r+\beta)}, \quad N \in \mathbb{N}, $$
and to establish that
$$ \int_{B_{R_N}(0)} z(\omega) \, d\omega \leq 2\|f\|_2^2 R^{-1}(3/2)^{-N/(2r+\beta)}, \quad (134) $$
for $N \geq 1$, and
$$ \int_{\mathbb{R} \setminus B_{R_N}(0)} z(\omega) \, d\omega \leq \frac{2\mu^2}{2s + \beta} (3/2)^{-N/(2r+\beta)}, \quad (135) $$
for $N \geq N_0 := \lceil(2s+\beta+1) \log_3(2L+1) \rceil$. This then implies
$$ \int_{\mathbb{R}} z(\omega) \, d\omega \leq C \{1, N_0\} = N_0, $$
where
$$ C := 4 \cdot (3/2)^{2r+\beta} \max\left\{\|f\|_2^2 R^{-1}, \mu^2 (2s + \beta)^{-1}\right\}, $$
and thereby completes the proof. It remains to establish (134) and (135). We start with (134) and note that
$$ (1 - 2|\omega|) \leq (1 - |\omega|)^{2l}, \quad \forall \omega \in \mathbb{R}, $$
for $l > 1$, see Fig. B in Appendix D. This implies
$$ 1 - \left|\hat{r}_l \left(\frac{\omega}{(3/2)^N-1}R\right)\right|^2 \leq 1 - \left(1 - \frac{\omega}{(3/2)^N-1}R\right)^{2l} \leq \frac{2l}{(3/2)^N-1} |\omega|, \quad \forall \omega \in \mathbb{R}, $$
which when employed in $\int_{B_{R_N}(0)} z(\omega) \, d\omega$ yields (134) thanks to
$$ \int_{B_{R_N}(0)} z(\omega) \, d\omega \leq \frac{2l}{(3/2)^N-1} \int_{B_{R_N}(0)} \left|\hat{f}(\omega)\right|^2 |\omega| \, d\omega \leq \frac{2lR_N}{(3/2)^N-1} \|f\|_2^2 \left(\frac{3}{2}\right)^{-N/(2r+\beta)}, $$
for $N \geq 1$, where we used Parseval’s formula as well as $R_N = (3/2)^{-N/(2r+\beta)}$. Next, we establish (135) and start by noting that $1 - \left|\hat{r}_l \left(\frac{\omega}{(3/2)^N-1}R\right)\right|^2 \leq 1$, for $\omega \in \mathbb{R}$, which is by
$$ 0 \leq \hat{r}_l(\omega) \leq 1, \quad \forall \omega \in \mathbb{R}. $$
Moreover, we have
$$ R_N = (3/2)^{-N/(2r+\beta)} \geq (3/2)^{-N/(2r+\beta+1)}, \quad (136) $$
for $N \geq N_0$, which by (133) implies
$$ |\hat{f}(\omega)| \leq \frac{\mu}{(1 + |\omega|^2)^{\frac{1}{2} + \frac{\mu}{4} + \frac{\beta}{4}}}, \quad \forall |\omega| \geq R_N. $$
Employing (136) in $\int_{\mathbb{R} \setminus B_{R_N}(0)} z(\omega) \, d\omega$ now yields
$$ \int_{\mathbb{R} \setminus B_{R_N}(0)} z(\omega) \, d\omega \leq \int_{\mathbb{R} \setminus B_{R_N}(0)} \frac{\mu^2}{1 + |\omega|^2} d\omega \leq 2\mu^2 \int_{R_N} \frac{1}{(1 + r^2)^{s + \frac{\mu}{4} + \frac{\beta}{4}}} r \, dr, \quad (137) $$
where in the last step we introduced polar coordinates. Moreover,
$$ (1 + r^2)^{s + \frac{\mu}{4} + \frac{\beta}{4}} \geq 2^{s+\beta+1}, \quad r \geq 0, $$
which when employed in (137) yields (135) owing to
$$ \int_{\mathbb{R} \setminus B_{R_N}(0)} z(\omega) \, d\omega \leq \frac{2\mu^2}{2s + \beta} R^{-2s-\beta} = \frac{2\mu^2}{2s + \beta} (3/2)^{-N/(2r+\beta+1)}, $$
where in the last step we used $R_N = (3/2)^{-N/(2r+\beta)}$. This completes the proof of statement ii).

APPENDIX H
PROOF OF COROLLARY I
We start with statement i) and note that $A^*_N = B^*_N = 1$, $N \in \mathbb{N}$, by assumption. Let $f \in L^2(\mathbb{R})$ with
$$ \text{supp}(\hat{f}) \subseteq B_L(0). $$
Then, by Proposition I in Appendix E together with
$$ \lim_{N \to \infty} W_N(f) = 0, \quad \forall f \in L^2(\mathbb{R}), $$
which follows from Proposition 2 in Appendix F, we have
$$ \|f\|_2^2 = \|\Phi_{\Omega}(f)\|_2^2 = \sum_{n=0}^{\infty} \|\Phi_{\Omega}^n(f)\|_2^2 $$
$$ \geq \sum_{n=0}^{N} \|\Phi_{\Omega}^n(f)\|_2^2 = \|f\|_2^2 - W_N(1)(f), \quad (138) $$
$$ \geq \int_{\mathbb{R}^d} \left|\hat{f}(\omega)\right|^2 \left|\hat{r}_l \left(\frac{\omega}{(N + 1)^{\alpha \delta}}\right)\right|^2 d\omega, \quad (139) $$
$$ = \int_{B_L(0)} \left|\hat{f}(\omega)\right|^2 \left|\hat{r}_l \left(\frac{\omega}{(N + 1)^{\alpha \delta}}\right)\right|^2 d\omega, \quad (140) $$
where (138) is by the lower bound in (82), (139) is thanks to Parseval’s formula and (20), and (140) follows from $f$ being $L$-band-limited. Next, thanks to $\hat{r}_l$ monotonically decreasing in $|\omega|$, we get
$$ \left|\hat{r}_l \left(\frac{\omega}{(N + 1)^{\alpha \delta}}\right)\right|^2 \geq \left|\hat{r}_l \left(\frac{L}{(N + 1)^{\alpha \delta}}\right)\right|^2, \quad (141) $$
for $\omega \in B_L(0)$. Employing (141) in (140), we obtain...
\[ \|f\|_2^2 \geq \|\hat{r}_i \left( \frac{L}{(N+1)\alpha\delta} \right)\|^2 \|f\|_2^2 = \left( 1 - \frac{L}{(N+1)\alpha\delta} \right)^{2l} \|f\|_2^2 \]

where in (143) we used Parseval’s formula, the equality in (142) is due to \( L \leq (N+1)\alpha\delta \), which, in turn, is by \( \delta \leq \frac{\alpha}{2} \), and the inequality in (143) is also by \( \delta \leq \frac{\alpha}{2} \) (upon rearranging terms). This establishes (32) and thereby completes the proof.

The proof of statement ii) is very similar to that of statement i). Again, we start by noting that \( A_0^N = B_0^N = 1 \), \( N \in \mathbb{N} \), by assumption. Let \( f \in L^2(\mathbb{R}) \) with \( \text{supp}(f) \subseteq B_0(0) \). Then, by Proposition 1 in Appendix I together with \( \lim_{N \to \infty} W_N(f) = 0 \), for \( f \in L^2(\mathbb{R}) \), we have

\[ \|f\|_2^2 = \|\Phi_0(f)\|^2 = \sum_{n=0}^{\infty} \|\Phi_n(\frac{\omega}{a^N\delta})\|^2 \]

\[ \geq \sum_{n=0}^{N} \|\Phi_n(f)\|^2 = \|f\|_2^2 - W_{N+1}(f) \]

\[ \geq \int_{B^2} |\hat{f}(\omega)|^2 \hat{r}_i \left( \frac{\omega}{a^N\delta} \right)^2 d\omega \]

where (144) is by the lower bound in (82). (145) is thanks to Parseval’s formula and (27) and (30), and (146) follows from \( f \) being \( L \)-band-limited. Next, thanks to \( \hat{r}_i \) monotonically decreasing in \( \omega \), we get

\[ \hat{r}_i \left( \frac{\omega}{a^N\delta} \right)^2 \geq \hat{r}_i \left( \frac{L}{a^N\delta} \right)^2, \quad \forall \omega \in B_L(0). \]

Employing (147) in (146) yields

\[ \|f\|_2^2 \geq \|\hat{r}_i \left( \frac{L}{a^N\delta} \right)\|^2 \|f\|_2^2 = \left( 1 - \frac{L}{a^N\delta} \right)^{2l} \|f\|_2^2 \]

where in (148) we used Parseval’s formula, the equality in (149) is by \( L \leq a^N\delta \), which, in turn, is by (34), and the inequality in (149) is also due to (34) (upon rearranging terms). This establishes (32) and thereby completes the proof of ii).

### Appendix I

**Proof of Corollary 2**

The proof is very similar to that of Corollary 1 in Appendix I. We start with statement i) and note that \( H^s(\mathbb{R}^d) \subseteq L^2(\mathbb{R}^d) \), \( s \geq 0 \), which follows from

\[ \|f\|_2^2 = \int_{\mathbb{R}^d} |\hat{f}(\omega)|^2 d\omega \]

\[ \leq \int_{\mathbb{R}^d} |\hat{f}(\omega)|^2 (1 + |\omega|^2)^s d\omega < \infty, \]

for \( f \in H^s(\mathbb{R}^d) \), where we used Parseval’s formula. Next, let \( f \in H^s(\mathbb{R}^d \setminus \{0\}) \) and \( \varepsilon \in (0, 1) \) and note that, by (141), Sec.

\[ f(\omega) \leq \frac{\mu}{(1 + |\omega|^2)^{1+\frac{1}{4}+\tau}}, \quad \forall \omega \geq L. \]

Moreover, the constant \( \tau := \max\{L, \gamma\} \), where

\[ \gamma := \left( \frac{\mu^2 \text{vol}^{d-1}(\partial B_1(0))}{(2s + \beta)(1 - \sqrt{1 - \varepsilon}) \|f\|_2^2} \right)^{\frac{1}{2s+\beta}}, \]

is finite thanks to \( 0 < \|f\|_2 < \infty \), which, in turn, is by \( f \neq 0 \) and (150). Then, we have

\[ \int_{\mathbb{R}^d \setminus B_r(0)} |\hat{f}(\omega)|^2 d\omega \]

\[ \leq \int_{\mathbb{R}^d \setminus B_r(0)} \frac{\mu^2}{(1 + |\omega|^2)^{1+\frac{1}{4}+\tau} d\omega} \]

\[ = \mu^2 \text{vol}^{d-1}(\partial B_1(0)) \int_{r}^{\infty} \frac{\tau^{d-1}}{d\omega} \]

\[ \leq \mu^2 \text{vol}^{d-1}(\partial B_1(0)) \int_{r}^{(2s + \beta)} \frac{\tau^{d-1}}{d\omega} \]

\[ \leq (1 - \sqrt{1 - \varepsilon}) \|f\|_2^2. \]

Here, (153) follows from (151) and (152), in (154) we changed to polar coordinates, (155) is owing to

\[ (1 + r^2)^{1+\frac{1}{4}+\tau} \geq r^{2s+\beta}, \quad r \in \mathbb{R}, \]

and (156) is by (152). Next, we note that \( A_0^N = B_0^N = 1 \), \( N \in \mathbb{N} \), by assumption. By Proposition 1 in Appendix I together with \( \lim_{N \to \infty} W_N(f) = 0 \), for \( f \in L^2(\mathbb{R}^d) \), which follows from Proposition 2 in Appendix I, we have

\[ \|f\|_2^2 = \|\Phi_0(f)\|^2 = \sum_{n=0}^{\infty} \|\Phi_n(f)\|^2 \]

\[ \geq \sum_{n=0}^{N} \|\Phi_n(f)\|^2 = \|f\|_2^2 - W_{N+1}(f) \]

\[ \geq \int_{\mathbb{R}^d} |\hat{f}(\omega)|^2 \hat{r}_i \left( \frac{\omega}{(N+1)\alpha\delta} \right)^2 d\omega \]

\[ = \int_{B_{(N+1)\alpha\delta}(0)} |\hat{f}(\omega)|^2 \hat{r}_i \left( \frac{\omega}{(N+1)\alpha\delta} \right)^2 d\omega, \]

where (157) is by the lower bound in (82), (158) is thanks to Parseval’s formula and (20), and (159) follows from \( \text{supp}(\hat{r}_i) \subseteq [0, 1] \). Next, thanks to

\[ \frac{\tau}{(N+1)\alpha\delta} \geq \tau, \]

by (35) and (36), we have

\[ \int_{B_{(N+1)\alpha\delta}(0)} |\hat{f}(\omega)|^2 \hat{r}_i \left( \frac{\omega}{(N+1)\alpha\delta} \right)^2 d\omega \]

\[ \geq \int_{B_r(0)} |\hat{f}(\omega)|^2 \hat{r}_i \left( \frac{\omega}{(N+1)\alpha\delta} \right)^2 d\omega \]

\[ \geq \left| \hat{r}_i \left( \frac{\tau}{(N+1)\alpha\delta} \right) \right|^2 \int_{B_r(0)} |\hat{f}(\omega)|^2 d\omega, \]

\[ \geq \left( \frac{\mu^2 \text{vol}^{d-1}(\partial B_1(0))}{(2s + \beta)(1 - \sqrt{1 - \varepsilon}) \|f\|_2^2} \right)^{\frac{1}{2s+\beta}}, \]

\[ \geq \left( \frac{\mu^2 \text{vol}^{d-1}(\partial B_1(0))}{(2s + \beta)(1 - \sqrt{1 - \varepsilon}) \|f\|_2^2} \right)^{\frac{1}{2s+\beta}}, \]

\[ \geq \left( \frac{\mu^2 \text{vol}^{d-1}(\partial B_1(0))}{(2s + \beta)(1 - \sqrt{1 - \varepsilon}) \|f\|_2^2} \right)^{\frac{1}{2s+\beta}}, \]
where the last step is because \( \hat{r}_1 \) is monotonically decreasing in \( |\omega| \). Employing (160) in (159), we get
\[
\|f\|^2 \geq \left| \hat{r}_1 \left( \frac{\tau}{(N+1)^{\alpha}} \right) \right|^2 \int_{B_r(0)} |\hat{f}(\omega)|^2 d\omega
\]
\[
= \left| \hat{r}_1 \left( \frac{\tau}{(N+1)^{\alpha}} \right) \right|^2 \left( \int_{|\omega|} |\hat{f}(\omega)|^2 d\omega - \int_{|\omega|} |\hat{f}(\omega)|^2 d\omega \right)
\]
\[
\geq \left| \hat{r}_1 \left( \frac{\tau}{(N+1)^{\alpha}} \right) \right|^2 \sqrt{1-\varepsilon} \|f\|^2
\]
\[= \left( 1 - \frac{\tau}{(N+1)^{\alpha}} \right)^{2l} \sqrt{1-\varepsilon} \|f\|^2 \]
\[
= \left( 1 - \frac{\tau}{(N+1)^{\alpha}} \right)^{2l} \sqrt{1-\varepsilon} \|f\|^2
\]
(161)
(162)
where the inequality in (161) is by Parseval’s formula and (162). The equality in (162) is due to \((N+1)^{\alpha} \geq \tau\), which, in turn, is by (55) and (36), and the last inequality is also due to (55) and (36). This establishes (32) and thereby completes the proof of i).

The proof of statement ii) is very similar to that of statement i). Let \( f \in H^s(R) \setminus \{0\} \) and \( \varepsilon \in (0, 1) \) and note that, by [41 Sec. 6.2.1], there exist \( \beta, \mu, L > 0 \) (all depending on \( f \)) such that
\[
|\hat{f}(\omega)| \leq \frac{\mu}{(1 + |\omega|^2)^{s+\frac{1}{2} + \frac{\gamma}{2}}}, \quad \forall |\omega| \geq L. \quad \text{(163)}
\]
Again, the constant \( \tau := \max\{L, \gamma\} \), where
\[
\gamma := \left( \frac{2\mu^2}{(2s+\beta)(1-\sqrt{1-\varepsilon})\|f\|_2^2} \right)^{1/(2s+\beta+1)}, \quad \text{(164)}
\]
is finite thanks to \( 0 < \|f\|_2 < \infty \), which, in turn, is by \( f \neq 0 \) and (150). Then, we have
\[
\int_{R \setminus B_r(0)} |\hat{f}(\omega)|^2 d\omega \leq \int_{R \setminus B_r(0)} \frac{\mu^2}{(1 + |\omega|^2)^{s+\frac{1}{2} + \frac{\gamma}{2}}} d\omega \quad \text{(165)}
\]
\[
= \frac{2\mu^2}{\gamma} \int_0^{\frac{1}{\sqrt{1-\varepsilon}}} \frac{1}{(1 + r^2)^{s+\frac{1}{2} + \frac{\gamma}{2}}} dr \quad \text{(166)}
\]
\[
= \frac{2\mu^2}{\gamma} \int_0^{\frac{1}{\sqrt{1-\varepsilon}}} r^{-2s+\beta+1} dr \quad \text{(167)}
\]
\[
\leq \left( 1 - \sqrt{1-\varepsilon} \right)\|f\|_2^2 \quad \text{(168)}
\]
Here, (165) follows from (163) and (164), in (166) we introduced polar coordinates, (167) is owing to
\[
(1 + r^2)^{s+\frac{1}{2} + \frac{\gamma}{2}} \geq r^{2s+1+\beta}, \quad r \in \mathbb{R},
\]
and the step leading to (168) is by (164). Next, we note that \( A_N^H = B_N^S = 1, N \in \mathbb{N} \), by assumption. By Proposition 1 in Appendix E together with
\[
\lim_{N \to \infty} W_N(f) = 0, \quad \forall f \in L^2(R),
\]
we have
\[
\|f\|^2 = \|\Phi_g(f)\|^2 = \sum_{n=0}^{\infty} \|\Phi_g^n(f)\|^2
\]
\[
\geq \sum_{n=0}^{N} \|\Phi_g^n(f)\|^2 = \|f\|^2 - W_{N+1}(f) \quad \text{(169)}
\]
\[
\geq \int_{R \setminus B_r(0)} |\hat{f}(\omega)|^2 |\hat{r}_1 \left( \frac{\omega}{a^N \delta} \right)|^2 d\omega
\]
\[
= \int_{B_r(0)} |\hat{f}(\omega)|^2 |\hat{r}_1 \left( \frac{\omega}{a^N \delta} \right)|^2 d\omega \quad \text{(170)}
\]
\[
\geq \left( 1 - \frac{\tau}{a^{N} \delta} \right)^{2l} \sqrt{1-\varepsilon} \|f\|^2 \quad \text{(171)}
\]
where (169) is by the lower bound in (82), (170) is thanks to Parseval’s formula and (77) and (30), and (171), where \( \xi := a^{N} \delta \), follows from \( \text{supp}(\hat{r}_1) \subseteq [0, 1] \). Next, thanks to \( \xi = a^{N} \delta \geq \tau \), by (37) and (38), we have
\[
\int_{B_r(0)} |\hat{f}(\omega)|^2 |\hat{r}_1 \left( \frac{\omega}{a^{N} \delta} \right)|^2 d\omega
\]
\[
\geq \int_{B_r(0)} |\hat{f}(\omega)|^2 |\hat{r}_1 \left( \frac{\omega}{a^{N} \delta} \right)|^2 d\omega
\]
\[
\geq \left( 1 - \frac{\tau}{a^{N} \delta} \right)^{2l} \sqrt{1-\varepsilon} \|f\|^2 \quad \text{(172)}
\]
where the last step is because \( \hat{r}_1 \) is monotonically decreasing in \( |\omega| \). Employing (172) in (171), we get
\[
\|f\|^2 \geq \left| \hat{r}_1 \left( \frac{\tau}{a^{N} \delta} \right) \right|^2 \int_{R \setminus B_r(0)} |\hat{f}(\omega)|^2 d\omega
\]
\[
= \left| \hat{r}_1 \left( \frac{\tau}{a^{N} \delta} \right) \right|^2 \left( \int_{R \setminus B_r(0)} |\hat{f}(\omega)|^2 d\omega - \int_{R \setminus B_r(0)} |\hat{f}(\omega)|^2 d\omega \right)
\]
\[
\geq \left| \hat{r}_1 \left( \frac{\tau}{a^{N} \delta} \right) \right|^2 \sqrt{1-\varepsilon} \|f\|^2 \quad \text{(173)}
\]
\[
= \left( 1 - \frac{\tau}{a^{N} \delta} \right)^{2l} \sqrt{1-\varepsilon} \|f\|^2 \quad \text{(174)}
\]
where the inequality in (173) is by Parseval’s formula and (168). The equality in (174) is due to \( a^{N} \delta \geq R \), which, in turn, is by (37) and (38), and the inequality in (174) is also due to (37) and (38). This establishes (32) and thereby completes the proof of ii).

References

[1] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” Nature, vol. 521, pp. 436–444, 2015.
[2] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. MIT Press, 2016. [http://www.deeplearningbook.org]
[3] Y. Bengio, A. Courville, and P. Vincent, “Representation learning: A review and new perspectives,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 35, no. 8, pp. 1798–1828, 2013.
[4] Y. LeCun, B. Boser, J. Denker, D. Henderson, R. Howard, W. Hubbard, and L. Jackel, “Handwritten digit recognition with a back-propagation network,” in Proc. of International Conference on Neural Information Processing Systems (NIPS), pp. 396–404, 1990.
[5] D. E. Rumelhart, G. Hinton, and R. J. Williams, “Learning internal representations by error propagation,” in Parallel distributed processing: Explorations in the microstructure of cognition (3). L. McClelland and D. E. Rumelhart, eds.), pp. 318–362, MIT Press, 1986.
[6] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” in Proc. of the IEEE, pp. 2278–2324, 1998.
He received the 2001 IEEE Signal Processing Society Young Author Best Paper Award, the 2006 IEEE Communications Society Leonard G. Abraham Best Paper Award, the 2010 Vodafone Innovations Award, the ETH "Golden Owl" Teaching Award, is a Fellow of the IEEE, a 2011 EURASIP Fellow, was a Distinguished Lecturer (2013-2014) of the IEEE Information Theory Society, an Erwin Schrödinger Fellow (1999-2001) of the Austrian National Science Foundation (FWF), was included in the 2014 Thomson Reuters List of Highly Cited Researchers in Computer Science, and is the 2016 Padovani Lecturer of the IEEE Information Theory Society. He served as an associate editor of the IEEE Transactions on Information Theory, the IEEE Transactions on Signal Processing, the IEEE Transactions on Wireless Communications, and the EURASIP Journal on Applied Signal Processing. He was editor-in-chief of the IEEE Transactions on Information Theory during the period 2010-2013. He served on the editorial board of the IEEE Signal Processing Magazine and is currently on the editorial boards of “Foundations and Trends in Networking” and “Foundations and Trends in Communications and Information Theory”. He was TPC co-chair of the 2008 IEEE International Symposium on Information Theory and the 2016 IEEE Information Theory Workshop and serves on the Board of Governors of the IEEE Information Theory Society. He has been a delegate of the president of ETH Zurich for faculty appointments since 2008.