# Residual Hybrid Filterbanks

Vincent Lostanlen[1], Xiran Zhang[1], Daniel Haider[2], Mathieu Lagrange[1], Martin Ehler[3], Peter Balazs[2]
[1]*Nantes Université, École Centrale Nantes, CNRS, LS2N, UMR 6004, F-44000 Nantes, France*
[2]*Acoustics Research Institute, Austrian Academy of Sciences, A-1010 Vienna, Austria*
[3]*Faculty of Mathematics, University of Vienna, Vienna, Austria*

***Abstract*—A hybrid filterbanks is a convolutional neural network (convnet) whose learnable filters operate over the subbands of a non-learnable filterbank, which is designed from domain knowledge. While hybrid filterbanks have found successful applications in speech enhancement, our paper shows that they remain susceptible to large deviations of the energy response due to randomness of convnet weights at initialization. Against this issue, we propose a variant of hybrid filterbanks, by inspiration from residual neural networks (ResNets). The key idea is to introduce a shortcut connection at the output of each non-learnable filter, bypassing the convnet. We prove that the shortcut connection in a residual hybrid filterbank lowers the relative standard deviation of the energy response while the pairwise cosine distances between non-learnable filters contributes to preventing duplicate features.**

***Index Terms*—audio and speech processing, convolutional neural networks, filterbank analysis, hybrid deep learning, random matrix theory.**

## I. Introduction

Hybrid deep learning [1], also known as model-based deep learning [2], has a key role to play in speech and music technology. By integrating domain knowledge within deep neural networks, hybrid systems can enhance explainability, controllability and resource efficiency [3]. In this context, hybrid auditory filterbanks have recently been proposed for speech enhancement [4]. Formally:

**Definition I.1.** A *hybrid filterbank* composes non-learnable filters $\boldsymbol{\psi}_1 \dots \boldsymbol{\psi}_J \in \mathbb{C}^N$ with learnable filters $\boldsymbol{w}_1 \dots \boldsymbol{w}_J \in \mathbb{R}^T$, $T \leq N$. Its response to $\boldsymbol{x} \in \mathbb{R}^N$ is a double convolution: $(\boldsymbol{w}_j * \boldsymbol{\psi}_j * \boldsymbol{x})$.

While the $\boldsymbol{\psi}_j$'s are designed by expert knowledge and kept fixed, the $\boldsymbol{w}_j$'s are typically initialized at random and then iteratively updated by gradient descent. Yet, convnets for raw audio are known to be susceptible to numerical instabilities, particularly for $T > 2^J$ and for $\boldsymbol{x}$ having strong correlations at lags up to $T$ [5].

In this article, we present theoretical results for a specific setting of hybrid filterbanks, where we reduce numerical instabilities while retaining their capability for gradient-based optimization. Our problem is that the non-learned filter $\boldsymbol{\psi}_j$ may introduce long-range correlations into $(\boldsymbol{\psi}_j * \boldsymbol{x})$, causing random fluctuations of the output energy $\|\boldsymbol{w}_j * \boldsymbol{\psi}_j * \boldsymbol{x}\|_2^2$. Of course, these random fluctuations could be canceled by initializing the $\boldsymbol{w}_j$'s with zeros; but such a design choice would cause the hybrid layer to predict a constant, which is detrimental to gradient-based optimization [6, Chapter 8.4]. Formally:

**Definition I.2.** A *residual hybrid filterbank (RHF)* is a hybrid filterbank in which the filters $\boldsymbol{w}_1 \dots \boldsymbol{w}_J$ are initialized as i.i.d. random Gaussian filters: $\boldsymbol{w}_j \sim \mathcal{N}(\mu\boldsymbol{\delta}, \sigma^2\mathbf{I})$ of length $T$ where $\boldsymbol{\delta}$ is the Kronecker symbol, i.e., $\boldsymbol{\delta}[t] = 1$ if $t = 0$ and zero otherwise.

The work of V. Lostanlen and X. Zhang is supported by ANR project MuReNN (ANR-23-CE23-0007-01). D. Haider is recipient of a DOC Fellowship of the Austrian Academy of Sciences at the Acoustics Research Institute (A 26355). The work of P. Balazs is supported by the FWF projects LoFT (P 34624), NoMASP (P 34922) and Voice Prints (P 36446). To reproduce our numerical simulations, please visit: https://github.com/lostanlen/lostanlen2025ssp

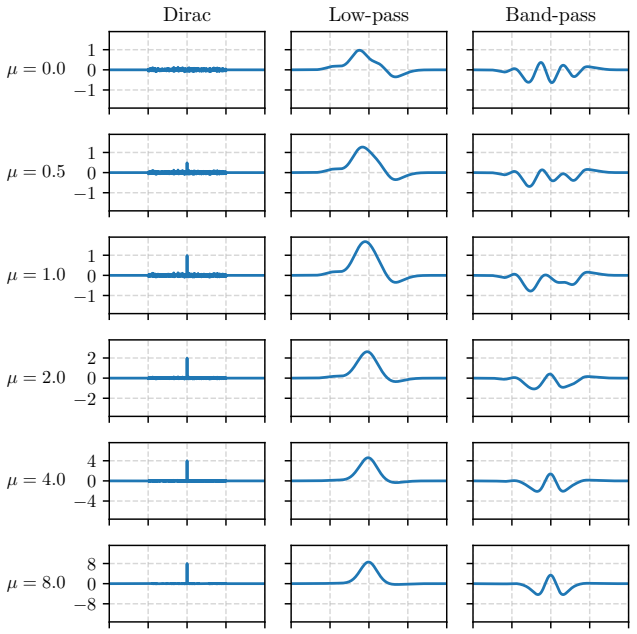


Fig. 1. Random samples of residual hybrid filters $(\boldsymbol{w} * \boldsymbol{\psi})$ where $\boldsymbol{w} \sim \mathcal{N}(\mu\boldsymbol{\delta}, \sigma^2\mathbf{I})$: see Definition I.2. Columns correspond to different designs of $\boldsymbol{\psi}$ while rows correspond to different values of the residual connection parameter $\mu$. Each x-axis tick denotes 512 time samples. Note the change in y-axis scaling across rows. $N = 2048$, $T = 1024$.

The term "residual" is chosen in reference to residual networks (ResNets) [7]. Indeed, the nonzero expected value of $\boldsymbol{w}_j$ may be interpreted as an identity mapping which is weighted by the parameter $\mu$: $(\mu\boldsymbol{\delta} + \boldsymbol{w}) * \boldsymbol{\psi} = (\mu\boldsymbol{\psi} + \boldsymbol{w} * \boldsymbol{\psi})$. However, an important difference is that ResNet blocks typically contain one or several nonlinearities before the identity mapping, whereas the response of an RHF is linear with respect to the input signal $\boldsymbol{x}$.

Figure 1 illustrates the effect of the residual connection onto random samples of a hybrid filter $(\boldsymbol{w} * \boldsymbol{\psi})$, for $\boldsymbol{\psi}$ being a Dirac impulse (left); a low-pass filter (center); and a band-pass filter (right). We observe that larger values of $\mu$ improve the temporal localization of $(\boldsymbol{w} * \boldsymbol{\psi})$ and reduce the relative effect of random initialization.

Section II gives exact formulae for the expected value and variance of $\|\boldsymbol{w}_j * \boldsymbol{\psi}_j\|_2^2$, and derives bounds for its relative standard deviation. Section III gives exact formulae for the expected area of the parallelogram with sides $(\boldsymbol{w}_j * \boldsymbol{\psi}_j)$ and $(\boldsymbol{w}_{j'} * \boldsymbol{\psi}_{j'})$ so as to measure feature diversity. Section IV supports our main findings with a numerical simulation. We conclude the article in Section V and defer the proofs to an appendix (Section VI).

## II. Moments of the energy response

**Definition II.1.** The circular *cross-correlation* between $\boldsymbol{y}$ and $\boldsymbol{y}'$ is

$$\mathbf{R}_{(\boldsymbol{y},\boldsymbol{y}')}[\tau] = \sum_{n=0}^{N-1} \overline{\boldsymbol{y}[n]}\boldsymbol{y}'[n+\tau] = \sum_{n=0}^{N-1} \overline{\boldsymbol{y}[n-\tau]}\boldsymbol{y}'[n], \tag{1}$$

where $\boldsymbol{y},\boldsymbol{y}' \in \mathbb{C}^N$, the variable $\tau$ is an integer lag, the overline denotes complex conjugation, and indexing is understood modulo $N$. The circular *autocorrelation* of $\boldsymbol{y}$ is $\mathbf{R}_{\boldsymbol{y}} = \mathbf{R}_{(\boldsymbol{y},\boldsymbol{y})}$.

**Proposition II.2.** Given $\boldsymbol{y} \in \mathbb{C}^N$ and $\boldsymbol{w} \sim \mathcal{N}(\mu\boldsymbol{\delta}, \sigma^2\mathbf{I})$ of length $T$:

$$\mathbb{E}\big[\|\boldsymbol{w} * \boldsymbol{y}\|_2^2\big] = \big(\mu^2 + \sigma^2 T\big)\|\boldsymbol{y}\|^2. \tag{2}$$

**Proposition II.3.** Given $\boldsymbol{y} \in \mathbb{C}^N$ and $\boldsymbol{w} \sim \mathcal{N}(\mu\boldsymbol{\delta}, \sigma^2\mathbf{I})$ of length $T$:

$$\begin{aligned}\mathbb{V}\big[\|\boldsymbol{w} * \boldsymbol{y}\|_2^2\big] &= 4\mu^2\sigma^2 \sum_{\tau=0}^{T-1} |\mathbf{R}_{\boldsymbol{y}}[\tau]|^2 \\ &\quad + 2\sigma^4 \sum_{\tau=-T}^{T} (T - |\tau|)|\mathbf{R}_{\boldsymbol{y}}[\tau]|^2.\end{aligned} \tag{3}$$

*Note.* The proposition above generalizes [5, Proposition II. 1] to complex values of $\boldsymbol{y}$ and to nonzero values of $\mu$.

**Theorem II.4.** *Given a residual hybrid filterbank with $\boldsymbol{w}_1 \dots \boldsymbol{w}_J \sim \mathcal{N}(\mu\boldsymbol{\delta}, \sigma^2\mathbf{I})$ i.i.d. and nonzero $\boldsymbol{\psi}_1 \dots \boldsymbol{\psi}_J$, for all $j$:*

$$\frac{2}{T}\left(1 - \frac{\mu^4}{(\mu^2+\sigma^2 T)^2}\right) \leq \frac{\mathbb{V}\big[\|\boldsymbol{w}_j * \boldsymbol{\psi}_j\|_2^2\big]}{\mathbb{E}\big[\|\boldsymbol{w}_j * \boldsymbol{\psi}_j\|_2^2\big]^2} \leq 2\left(1 - \frac{\mu^4}{(\mu^2+\sigma^2 T)^2}\right). \tag{4}$$

*Furthermore, the lower bound is reached if and only if there exist $n_0 \in \mathbb{Z}_N$ and $c \neq 0$ such that $\boldsymbol{\psi}[n_0] = c$ and $\boldsymbol{\psi}[n] = 0$ for all $n \neq n_0$.*

## III. Expected bivector magnitudes

**Definition III.1.** Given $\boldsymbol{y},\boldsymbol{y}' \in \mathbb{C}^N$, their *bivector magnitude* is

$$|\boldsymbol{y} \wedge \boldsymbol{y}'| = \sqrt{\|\boldsymbol{y}\|_2^2\|\boldsymbol{y}'\|_2^2 - |\langle \boldsymbol{y}|\boldsymbol{y}'\rangle|^2}, \tag{5}$$

i.e., the square root of the determinant of the Gram matrix associated to the indexed family $(\boldsymbol{y},\boldsymbol{y}')$.

*Note.* Geometrically, $|\boldsymbol{y} \wedge \boldsymbol{y}'|$ is the area under the parallelogram with sides $\boldsymbol{y}$ and $\boldsymbol{y}'$. After normalization by $\mathbb{E}[\|\boldsymbol{y}\|^2]\mathbb{E}[\|\boldsymbol{y}'\|^2]$, it can be interpreted as a quantitative measure of feature diversity.

**Proposition III.2.** Given $\boldsymbol{y},\boldsymbol{y}' \in \mathbb{C}^N$ and $\boldsymbol{w},\boldsymbol{w}' \sim \mathcal{N}(\mu\boldsymbol{\delta}, \sigma^2\mathbf{I})$ i.i.d.:

$$\mathbb{E}\big[\langle \boldsymbol{w} * \boldsymbol{y}|\boldsymbol{w}' * \boldsymbol{y}'\rangle\big] = \mu^2\langle \boldsymbol{y}|\boldsymbol{y}'\rangle \tag{6}$$

where the bracket notation is the Hermitian inner product over $\mathbb{C}^N$.

**Proposition III.3.** Given $\boldsymbol{y},\boldsymbol{y}' \in \mathbb{C}^N$ and $\boldsymbol{w},\boldsymbol{w}' \sim \mathcal{N}(\mu\boldsymbol{\delta}, \sigma^2\mathbf{I})$ i.i.d.:

$$\begin{aligned}\mathbb{V}\big[\langle \boldsymbol{w} * \boldsymbol{y}|\boldsymbol{w}' * \boldsymbol{y}'\rangle\big] &= 2\mu^2\sigma^2 \sum_{\tau=0}^{T-1} |\mathbf{R}_{(\boldsymbol{y}',\boldsymbol{y})}[\tau]|^2 \\ &\quad + \sigma^4 \sum_{\tau=-T}^{T} \big(T - |\tau|\big)|\mathbf{R}_{(\boldsymbol{y}',\boldsymbol{y})}[\tau]|^2\end{aligned} \tag{7}$$

**Proposition III.4.** Given a residual hybrid filterbank with $\boldsymbol{w}_1 \dots \boldsymbol{w}_J \sim \mathcal{N}(\mu\boldsymbol{\delta}, \sigma^2\mathbf{I})$ i.i.d. and $\boldsymbol{\psi}_1 \dots \boldsymbol{\psi}_J$, for all $j \neq j'$:

$$\begin{aligned}\mathbb{E}\big[|(\boldsymbol{w}_j * \boldsymbol{\psi}_j) \wedge (\boldsymbol{w}_{j'} * \boldsymbol{\psi}_{j'})|^2\big] &= \mu^4\left(\|\boldsymbol{\psi}_j\|_2^2\|\boldsymbol{\psi}_{j'}\|_2^2 - |\langle \boldsymbol{\psi}_j|\boldsymbol{\psi}_{j'}\rangle|^2\right) \\ &\quad + 2\mu^2\sigma^2 \sum_{\tau=0}^{T-1}\left(\|\boldsymbol{\psi}_j\|_2^2\|\boldsymbol{\psi}_{j'}\|_2^2 - |\mathbf{R}_{(\boldsymbol{\psi}_{j'},\boldsymbol{\psi}_j)}[\tau]|^2\right) \\ &\quad + \sigma^4 \sum_{|\tau|<T}\big(T - |\tau|\big)\left(\|\boldsymbol{\psi}_j\|_2^2\|\boldsymbol{\psi}_{j'}\|_2^2 - |\mathbf{R}_{(\boldsymbol{\psi}_{j'},\boldsymbol{\psi}_j)}[\tau]|^2\right)\end{aligned} \tag{8}$$

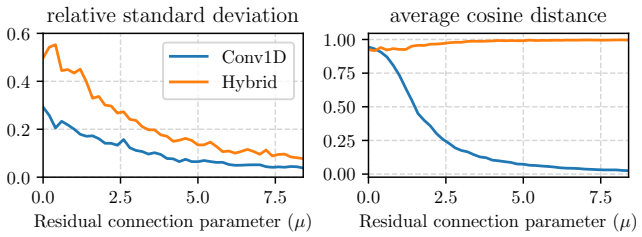


Fig. 2. Statistics of residual hybrid filterbanks with $J = 2$ filters as functions of the parameter $\mu$. Blue, lower curves: $\boldsymbol{\psi}_1$ and $\boldsymbol{\psi}_2$ are Dirac impulses, as in a plain "Conv1D" layer. Orange, upper curves: $\boldsymbol{\psi}_1$ and $\boldsymbol{\psi}_2$ are orthogonal low-pass and band-pass filters. Left: relative standard deviation of energy (lower is better): see Theorem II.4. Right: average cosine distance between the filters (higher is better): see Theorem III.5. $N = 32$, $T = 16$, 100 i.i.d. trials.

**Theorem III.5.** *Given a residual hybrid filterbank with $\boldsymbol{w}_1 \dots \boldsymbol{w}_J \sim \mathcal{N}(\mu\boldsymbol{\delta}, \sigma^2\mathbf{I})$ i.i.d. and nonzero $\boldsymbol{\psi}_1 \dots \boldsymbol{\psi}_J$, for all $j \neq j'$:*

$$\lim_{\mu/\sigma \to +\infty} \frac{\mathbb{E}\big[|(\boldsymbol{w}_j * \boldsymbol{\psi}_j) \wedge (\boldsymbol{w}_{j'} * \boldsymbol{\psi}_{j'})|^2\big]}{\mathbb{E}\big[\|\boldsymbol{w}_j * \boldsymbol{\psi}_j\|_2^2\|\boldsymbol{w}_{j'} * \boldsymbol{\psi}_{j'}\|_2^2\big]} = 1 - \frac{|\langle \boldsymbol{\psi}_j|\boldsymbol{\psi}_{j'}\rangle|^2}{\|\boldsymbol{\psi}_j\|_2^2\|\boldsymbol{\psi}_{j'}\|_2^2}. \tag{9}$$

*Note.* The right-hand side in the equation above is the cosine distance between vectors $\boldsymbol{\psi}_j$ and $\boldsymbol{\psi}_{j'}$. This distance is minimal if and only if $\boldsymbol{\psi}_j = \boldsymbol{\psi}_{j'}$ and maximal if and only if $\boldsymbol{\psi}_j$ and $\boldsymbol{\psi}_{j'}$ are orthogonal.

## IV. Numerical simulation

For simplicity, we consider a residual hybrid filterbank with only $J = 2$ filters: $\boldsymbol{\psi}_1$ is a low-pass filter and $\boldsymbol{\psi}_2$ is a band-pass filter, as seen in the center and right columns of Figure 1. As a baseline, we also construct a plain convolutional layer or "Conv1D" for short. The Conv1D layer is obtained by setting $\boldsymbol{\psi}_1 = \boldsymbol{\psi}_2 = \boldsymbol{\delta}$. For both filterbanks, we measure this relative standard deviation by estimating the expected value and variance of energy over 100 i.i.d. trials.

In Figure 2 (left), we observe that, in a non-residual setting, making the non-learned filters $\boldsymbol{\psi}_j$'s autocorrelated (i.e., non-Dirac) increases the typical random fluctuations of energy. This is in accordance with Theorem II.4, which states that, for $\mu = 0$, Conv1D is optimal in terms of residual standard deviation of energy. However, we may compensate for it via residual connections (see Theorem II.4): with our filterbank design, setting $\mu = 2.5$ leads to a relative standard deviation that is on par with Conv1D with $\mu = 0$. Crucially, this shift in expected value does not come at the cost of feature diversity: as seen in Figure 2 (right), on average over 100 i.i.d. trials, the cosine distance between $(\boldsymbol{w}_1 * \boldsymbol{\psi}_1)$ and $(\boldsymbol{w}_2 * \boldsymbol{\psi}_2)$ goes up with $\mu$ for our low-pass/band-pass pair, even so it goes down with $\mu$ for Conv1D. This is in accordance with Theorem III.5.

## V. Conclusion

We have presented a simple solution against excessive random fluctuations of energy in hybrid filterbanks. On the theoretical side, it remains to be seen how our characterization of feature diversity scales beyond $J = 2$ filters: the distribution of the condition number of residual hybrid filters with random Gaussian weights remain an open question [8]. On the practical side, residual connections could be paired with dilated convolutions with $\boldsymbol{w}_j$'s, as in multiresolution neural networks (MuReNN) [9].

## VI. Lemmata and Proofs

**Lemma VI.1.** *Given $\boldsymbol{y},\boldsymbol{y}' \in \mathbb{C}^N$ and $\boldsymbol{w},\boldsymbol{w}' \in \mathbb{R}^T$:*

$$\left\langle \boldsymbol{w} * \boldsymbol{y} \,\middle|\, \boldsymbol{w}' * \boldsymbol{y}' \right\rangle = \boldsymbol{w}^{\mathsf{T}} \cdot \mathbf{Q}_T(\boldsymbol{y},\boldsymbol{y}') \cdot \boldsymbol{w}' \tag{10}$$

*where $\mathbf{Q}_T(\boldsymbol{y},\boldsymbol{y}')[t,t'] = \mathbf{R}_{(\boldsymbol{y}',\boldsymbol{y})}[t'-t]$ for every $0 \le t,t' < T$.*

*Note.* In the circular cross-correlation (Definition II.1), we conjugate the first argument to enable its interpretation as inverse Fourier transform of cross-power spectral density. However, in the Hermitian dot product, we conjugate the second term so as to follow the convention of sesquilinear forms. Such a mismatch explains why Lemma VI.1 permutes indices $j$ and $j'$ when defining $\mathbf{Q}_T$.

*Proof.* We write the circular convolution $(\boldsymbol{w} * \boldsymbol{y})$ as the matrix–vector product $\mathbf{C}_T(\boldsymbol{y}) \cdot \boldsymbol{w}$ where

$$\mathbf{C}_T(\boldsymbol{y}) = \begin{pmatrix} \boldsymbol{y}[0] & \boldsymbol{y}[N-1] & \cdots & \boldsymbol{y}[N-T+1] \\ \boldsymbol{y}[1] & \boldsymbol{y}[0] & \cdots & \boldsymbol{y}[N-T+2] \\ \vdots & \vdots & & \vdots \\ \boldsymbol{y}[N-2] & \boldsymbol{y}[N-3] & \cdots & \boldsymbol{y}[N-T-1] \\ \boldsymbol{y}[N-1] & \boldsymbol{y}[N-2] & \cdots & \boldsymbol{y}[N-T] \end{pmatrix}$$

has entries $\mathbf{C}_T(\boldsymbol{y})[n,t] = \boldsymbol{y}[n-t]$ for $0 \le n < N$ and $0 \le t < T$. Likewise: $(\boldsymbol{w}' * \boldsymbol{x}') = \mathbf{C}_T(\boldsymbol{y}') \cdot \boldsymbol{w}'$. Hence the bilinear form:

$$\begin{aligned} \left\langle \boldsymbol{w} * \boldsymbol{y} \,\middle|\, \boldsymbol{w}' * \boldsymbol{y}' \right\rangle &= (\boldsymbol{w}' * \boldsymbol{y}')^{\mathsf{H}} \cdot (\boldsymbol{w} * \boldsymbol{y}) \\ &= (\mathbf{C}_T(\boldsymbol{y}') \cdot \boldsymbol{w}')^{\mathsf{H}} \cdot (\mathbf{C}_T(\boldsymbol{y}) \cdot \boldsymbol{w}) \\ &= (\boldsymbol{w}')^{\mathsf{T}} \cdot \left(\mathbf{C}_T(\boldsymbol{y}')^{\mathsf{H}} \cdot \mathbf{C}_T(\boldsymbol{y})\right) \cdot \boldsymbol{w} \\ &= \boldsymbol{w}^{\mathsf{T}} \cdot \left(\mathbf{C}_T(\boldsymbol{y}')^{\mathsf{H}} \cdot \mathbf{C}_T(\boldsymbol{y})\right)^{\mathsf{T}} \cdot \boldsymbol{w}', \end{aligned} \tag{11}$$

where $\mathsf{T}$ (resp. $\mathsf{H}$) denote transpose (resp. conjugate transpose). We recognize the definition of circular cross-correlation:

$$\begin{aligned} \left(\mathbf{C}_T(\boldsymbol{y}')^{\mathsf{H}} \cdot \mathbf{C}_T(\boldsymbol{y})\right)^{\mathsf{T}}[t,t'] &= \sum_{n=0}^{N-1} \overline{\mathbf{C}_T(\boldsymbol{y}')[t',n]} \mathbf{C}_T(\boldsymbol{y})[n,t] \\ &= \sum_{n=0}^{N-1} \overline{\boldsymbol{y}'[n-t']} \boldsymbol{y}[n-t] \\ &= \sum_{m=0}^{N-1} \overline{\boldsymbol{y}'[m-(t'-t)]} \boldsymbol{y}[m] \\ &= \mathbf{R}_{(\boldsymbol{y}',\boldsymbol{y})}[t'-t]. \end{aligned} \tag{12}$$

Defining the above as $\mathbf{Q}_T(\boldsymbol{y},\boldsymbol{y}')[t,t']$ concludes the proof. ∎

*Proof of Proposition II.2.* By the cyclic property of the trace, Lemma VI.1 with $\boldsymbol{w} = \boldsymbol{w}'$ and $\boldsymbol{y} = \boldsymbol{y}'$ yields

$$\begin{aligned} \|\boldsymbol{w} * \boldsymbol{y}\|_2^2 &= \boldsymbol{w}^{\mathsf{T}} \cdot \mathbf{Q}_T(\boldsymbol{y},\boldsymbol{y}) \cdot \boldsymbol{w} \\ &= \operatorname{Tr}\left(\boldsymbol{w}^{\mathsf{T}} \cdot \mathbf{Q}_T(\boldsymbol{y},\boldsymbol{y}) \cdot \boldsymbol{w}\right) \\ &= \operatorname{Tr}\left(\mathbf{Q}_T(\boldsymbol{y},\boldsymbol{y}) \cdot \boldsymbol{w} \cdot \boldsymbol{w}^{\mathsf{T}}\right). \end{aligned} \tag{13}$$

Thus, by linearity of the expected value and by definition of $\mathbf{Q}_T(\boldsymbol{y},\boldsymbol{y})$:

$$\begin{aligned} \mathbb{E}\left[\|\boldsymbol{w} * \boldsymbol{y}\|_2^2\right] &= \operatorname{Tr}\left(\mathbf{Q}_T(\boldsymbol{y},\boldsymbol{y}) \cdot \mathbb{E}\left[\boldsymbol{w}_j \cdot \boldsymbol{w}_j^{\mathsf{T}}\right]\right) \\ &= \operatorname{Tr}\left(\mathbf{Q}_T(\boldsymbol{y},\boldsymbol{y}) \cdot (\mu^2 \boldsymbol{\delta}\boldsymbol{\delta}^{\top} + \sigma^2 \mathbf{I})\right) \\ &= \mu^2 \operatorname{Tr}\left(\boldsymbol{\delta}^{\top} \cdot \mathbf{Q}_T(\boldsymbol{y},\boldsymbol{y}) \cdot \boldsymbol{\delta}\right) + \sigma^2 \operatorname{Tr}\left(\mathbf{Q}_T(\boldsymbol{y},\boldsymbol{y})\right) \\ &= (\mu^2 + \sigma^2 T) \mathbf{R}_{(\boldsymbol{y},\boldsymbol{y})}[0]. \end{aligned} \tag{14}$$

Replacing $\mathbf{R}_{(\boldsymbol{y},\boldsymbol{y})}[0]$ by $\|\boldsymbol{y}\|_2^2$ concludes the proof. ∎

*Proof of Proposition II.3.* By Proposition II.2:

$$\begin{aligned} \mathbb{V}\left[\|\boldsymbol{w} * \boldsymbol{y}\|_2^2\right] &= \mathbb{E}\left[\left(\|\boldsymbol{w} * \boldsymbol{y}\|_2^2\right)^2\right] - \mathbb{E}\left[\|\boldsymbol{w} * \boldsymbol{y}\|_2^2\right]^2 \\ &= \mathbb{E}\left[\|\boldsymbol{w} * \boldsymbol{y}\|_2^4\right] - (\mu^2 + T\sigma^2)^2 \|\boldsymbol{y}\|_2^4. \end{aligned} \tag{15}$$

By applying Lemma VI.1 with $\boldsymbol{w}' = \boldsymbol{w}$ and $\boldsymbol{y}' = \boldsymbol{y}$, and by noting that $\boldsymbol{w}$ is real-valued, the term $\|\boldsymbol{w} * \boldsymbol{y}\|_2^4$ can be expanded as:

$$\begin{aligned} \|\boldsymbol{w} * \boldsymbol{y}\|_2^4 &= \left(\|\boldsymbol{w} * \boldsymbol{y}\|_2^2\right)\left(\|\boldsymbol{w} * \boldsymbol{y}\|_2^2\right) \\ &= \left(\sum_{0 \le t,t' < T} \mathbf{R}_{\boldsymbol{y}}[t'-t] \boldsymbol{w}[t] \boldsymbol{w}[t']\right) \overline{\left(\sum_{0 \le u,u' < T} \mathbf{R}_{\boldsymbol{y}}[u'-u] \boldsymbol{w}[u] \boldsymbol{w}[u']\right)} \\ &= \left(\sum_{0 \le t,t' < T} \mathbf{R}_{\boldsymbol{y}}[t'-t] \boldsymbol{w}[t] \boldsymbol{w}[t']\right) \left(\sum_{0 \le u,u' < T} \overline{\mathbf{R}_{\boldsymbol{y}}[u'-u]} \boldsymbol{w}[u] \boldsymbol{w}[u']\right) \\ &= \sum_{t,t',u,u'} \mathbf{R}_{\boldsymbol{y}}[t'-t] \overline{\mathbf{R}_{\boldsymbol{y}}[u'-u]} \boldsymbol{w}[t] \boldsymbol{w}[t'] \boldsymbol{w}[u] \boldsymbol{w}[u'], \end{aligned} \tag{16}$$

where the indices $t,t',u,u'$ range between zero and $(T-1)$. Hence:

$$\mathbb{E}\left[\|\boldsymbol{w} * \boldsymbol{y}\|_2^4\right] = \sum_{t,t',u,u'} \mathbf{R}_{\boldsymbol{y}}[t'-t] \overline{\mathbf{R}_{\boldsymbol{y}}[u'-u]} \mathbb{E}\left[\boldsymbol{w}[t] \boldsymbol{w}[t'] \boldsymbol{w}[u] \boldsymbol{w}[u']\right] \tag{17}$$

We distinguish 12 cases in the sum above:

(i) if $t = t' = u = u' = 0$, $\mathbb{E}\left[\boldsymbol{w}[t]^4\right]$ is the kurtosis of $\boldsymbol{w}[0]$; i.e., $\mu^4 + 6\mu^2\sigma^2 + 3\sigma^4$. This term is weighted by $|\mathbf{R}_{\boldsymbol{y}}[0]|^2$.
(ii) if $t = t' = u = u' \neq 0$, $\mathbb{E}\left[\boldsymbol{w}[t]^4\right]$ is the kurtosis of $\boldsymbol{w}[t]$; i.e., $3\sigma^4$. There are $(T-1)$ such terms, weighted by $|\mathbf{R}_{\boldsymbol{y}}[0]|^2$.
(iii) if $0 = t = t' \neq u = u'$, $\mathbb{E}\left[\boldsymbol{w}[t]^2\right] = \mu^2 + \sigma^2$ and $\mathbb{E}\left[\boldsymbol{w}[u]^2\right] = \sigma^2$. There are $(T-1)$ such terms, weighted by $|\mathbf{R}_{\boldsymbol{y}}[0]|^2$.
(iv) if $0 = u = u' \neq t = t'$, we refer back to (iii).
(v) if $0 \neq u = u' \neq t = t' \neq 0$, $\mathbb{E}\left[\boldsymbol{w}[t]^2\right] = \mathbb{E}\left[\boldsymbol{w}[u]^2\right] = \sigma^2$. There are $(T-1)(T-2)$ such terms, weighted by $|\mathbf{R}_{\boldsymbol{y}}[0]|^2$.
(vi) if $0 = t = u \neq t' = u'$, $\mathbb{E}\left[\boldsymbol{w}[t]^2\right] = \mu^2 + \sigma^2$ and $\mathbb{E}\left[\boldsymbol{w}[t']^2\right] = \sigma^2$. These terms are weighted by $|\mathbf{R}_{\boldsymbol{y}}[t']|^2$ for $0 \le t' < T$.
(vii) if $0 = t' = u' \neq t = u$, $\mathbb{E}\left[\boldsymbol{w}[t]^2\right] = \sigma^2$ and $\mathbb{E}\left[\boldsymbol{w}[t']^2\right] = \mu^2 + \sigma^2$. These terms are weighted by $|\mathbf{R}_{\boldsymbol{y}}[-t]|^2$ for $0 < t < T$.
(viii) if $0 = t = u' \neq t' = u$, we refer back to (vi).
(ix) if $0 = t' = u \neq t = u'$, we refer back to (vii).
(x) if $0 \neq t = u \neq t' = u' \neq 0$, $\mathbb{E}\left[\boldsymbol{w}[t]^2\right] = \mathbb{E}\left[\boldsymbol{w}[u]^2\right] = \sigma^2$. These terms are weighted by $|\mathbf{R}_{\boldsymbol{y}}[t'-t]|^2$ for $0 < t \neq t' < T$.
(xi) if $0 \neq t = u' \neq t' = u \neq 0$, we refer back to (x).
(xii) otherwise, the term equals zero.

The autocorrelation has Hermitian symmetry: $\mathbf{R}_{\boldsymbol{y}}[-\tau] = \overline{\mathbf{R}_{\boldsymbol{y}}[\tau]}$. Hence: $|\mathbf{R}_{\boldsymbol{y}}[-\tau]|^2 = |\mathbf{R}_{\boldsymbol{y}}[\tau]|^2$. Equation (17) becomes:

$$\begin{aligned} \mathbb{E}\left[\|\boldsymbol{w} * \boldsymbol{y}\|_2^4\right] =& (\mu^4 + 6\mu^2\sigma^2 + 3\sigma^4)|\mathbf{R}_{\boldsymbol{y}}[0]|^2 && \text{(i)} \\ &+ 3(T-1)\sigma^4 |\mathbf{R}_{\boldsymbol{y}}[0]|^2 && \text{(ii)} \\ &+ 2(T-1)(\mu^2\sigma^2 + \sigma^4)|\mathbf{R}_{\boldsymbol{y}}[0]|^2 && \text{(iii) and (iv)} \\ &+ (T-1)(T-2)\sigma^4 |\mathbf{R}_{\boldsymbol{y}}[0]|^2 && \text{(v)} \\ &+ 4(\mu^2\sigma^2 + \sigma^4) \sum_{0 < \tau < T} |\mathbf{R}_{\boldsymbol{y}}[\tau]|^2 && \text{(vi) to (ix)} \\ &+ 2\sigma^4 \sum_{0 < t \neq t' < T} |\mathbf{R}_{\boldsymbol{y}}[t'-t]|^2 && \text{(x) and (xi)} \\ =& \alpha\mu^4 + \beta\mu^2\sigma^2 + \gamma\sigma^4 \end{aligned} \tag{18}$$

where $\alpha = |\mathbf{R}_{\boldsymbol{y}}[0]|^2 = \|\boldsymbol{y}\|_2^4$,

$$\begin{aligned} \beta &= 6|\mathbf{R}_{\boldsymbol{y}}[0]|^2 + 2(T-1)|\mathbf{R}_{\boldsymbol{y}}[0]|^2 + 4 \sum_{0 < \tau < T} |\mathbf{R}_{\boldsymbol{y}}[\tau]|^2 \\ &= 2T\|\boldsymbol{y}\|_2^4 + 4 \sum_{\tau=0}^{T-1} |\mathbf{R}_{\boldsymbol{y}}[\tau]|^2, \end{aligned} \tag{19}$$

$$\begin{aligned}\gamma &= \big((3+3(T-1)+2(T-1)+(T-1)(T-2)\big)|\mathbf{R}_{\boldsymbol{y}}[0]|^2 \\ &\quad + 4\sum_{0<\tau<T}|\mathbf{R}_{\boldsymbol{y}}[\tau]|^2 + 2\sum_{0<t\neq t'<T}|\mathbf{R}_{\boldsymbol{y}}[t'-t]|^2 \\ &= T^2\|\boldsymbol{y}\|_2^4 + 2\sum_{t=0}^{T-1}\sum_{t'=0}^{T-1}|\mathbf{R}_{\boldsymbol{y}}[t'-t]|^2. \end{aligned} \tag{20}$$

Given an integer $\tau$ such that $|\tau|<T$, there are $(T-|\tau|)$ pairs $(t,t')$ in the double sum above such that $t'-t=\tau$. Thus:

$$\gamma = T^2\sigma^4\|\boldsymbol{y}\|_2^4 + 2\sigma^4\sum_{\tau=-T}^{T}\big(T-|\tau|\big)|\mathbf{R}_{\boldsymbol{y}}[\tau]|^2 \tag{21}$$

Using Equation (18) and Proposition II.2, we rewrite Equation (15) as:

$$\mathbb{V}\big[\|\boldsymbol{w}*\boldsymbol{y}\|_2^2\big] = \alpha\mu^4 + \beta\mu^2 + \gamma - (\mu^2+T\sigma^2)^2\|\boldsymbol{y}\|_2^4. \tag{22}$$

After having replaced $\alpha$, $\beta$, and $\gamma$ by their definitions, we cancel the term $(\mu^2+T\sigma^2)^2\|\boldsymbol{y}\|_2^4$, which completes the proof. ■

**Lemma VI.2.** *Given $\boldsymbol{y}\in\mathbb{C}^N$ and $\tau\in\mathbb{Z}$, $|\mathbf{R}_{\boldsymbol{y}}[\tau]|\leq\mathbf{R}_{\boldsymbol{y}}[0]$.*

*Proof.* Given $\tau$, we express $\mathbf{R}_{\boldsymbol{y}}[\tau]$ as an inverse discrete Fourier transform and apply the triangular inequality, yielding:

$$|\mathbf{R}_{\boldsymbol{y}}[\tau]| \leq \frac{1}{N}\sum_{\omega=0}^{N-1}\Big||\widehat{\boldsymbol{y}}[\omega]|^2 e^{2\pi i\omega\tau/N}\Big| = \frac{1}{N}\sum_{\omega=0}^{N-1}|\widehat{\boldsymbol{y}}[\omega]|^2. \tag{23}$$

On the right hand side, we recognize the inverse discrete Fourier transform of $|\widehat{\boldsymbol{y}}|^2$ at lag zero, which concludes the proof. ■

*Proof of Theorem II.4.* Given $j$, we apply Proposition III.3 with $\boldsymbol{w}=\boldsymbol{w}_j$ and $\boldsymbol{y}=\boldsymbol{\psi}_j$. On one hand, for every $\tau>0$, $|\mathbf{R}_{\boldsymbol{\psi}_j}[\tau]|^2\geq 0$. Thus:

$$\mathbb{V}\big[\|\boldsymbol{w}_j*\boldsymbol{\psi}_j\|_2^2\big] \geq 4\mu^2\sigma^2|\mathbf{R}_{\boldsymbol{\psi}_j}[0]|^2 + 2\sigma^4 T|\mathbf{R}_{\boldsymbol{\psi}_j}[0]|^2. \tag{24}$$

We observe that $2\mu^2\sigma^2+\sigma^4T^2 = (\mu^2+\sigma^2T)^2-\mu^4$ and replace $|\mathbf{R}_{\boldsymbol{\psi}_j}[0]|^2$ by $\|\boldsymbol{\psi}_j\|_2^4$. Hence, after dividing by $T/2$:

$$\mathbb{V}\big[\|\boldsymbol{w}_j*\boldsymbol{\psi}_j\|_2^2\big] \geq \frac{2}{T}\big((\mu^2+\sigma^2T)^2-\mu^4\big)\|\boldsymbol{\psi}_j\|_2^4. \tag{25}$$

By applying Proposition II.2 with $\boldsymbol{w}=\boldsymbol{w}_j$ and $\boldsymbol{y}=\boldsymbol{\psi}_j$, we recognize the squared expected value: $\mathbb{E}\big[\|\boldsymbol{w}_j*\boldsymbol{\psi}_j\|_2^2\big]^2 = (\mu^2+\sigma^2T)^2\|\boldsymbol{\psi}_j\|_2^4$, yielding the lower bound. On the other hand, Lemma VI.2 yields:

$$\mathbb{V}\big[\|\boldsymbol{w}_j*\boldsymbol{\psi}_j\|^2\big] \leq 4\mu^2\sigma^4\sum_{\tau=0}^{T-1}|\mathbf{R}_{\boldsymbol{\psi}_j}[0]|^2 + 2\sigma^4\sum_{\tau=-T}^{T}\big(T-|\tau|\big)|\mathbf{R}_{\boldsymbol{\psi}_j}[0]|^2. \tag{26}$$

The first sum has $T$ equal terms. For the second sum, we compute:

$$\sum_{\tau=-T}^{T}\big(T-|\tau|\big) = T + 2\sum_{\tau=1}^{T}\big(T-\tau\big) = T + 2\frac{T(T-1)}{2} = T^2. \tag{27}$$

Thus, by replacing $|\mathbf{R}_{\boldsymbol{\psi}_j}[0]|^2$ by $\|\boldsymbol{\psi}_j\|_2^4$, and similarly to Equation (25):

$$\begin{aligned}\mathbb{V}\big[\|\boldsymbol{w}_j*\boldsymbol{\psi}_j\|_2^2\big] &\leq (4\mu^2\sigma^2T + 2\sigma^4T^2)\|\boldsymbol{\psi}_j\|_2^4 \\ &\leq 2\big((\mu^2+\sigma^2T)^2-\mu^4\big)\|\boldsymbol{\psi}_j\|_2^4. \end{aligned} \tag{28}$$

Dividing by $\mathbb{E}\big[\|\boldsymbol{w}_j*\boldsymbol{\psi}_j\|_2^2\big]^2$ concludes the proof.

Moreover, for $\boldsymbol{\psi}_j : n\mapsto c\boldsymbol{\delta}[n-n_0]$, we compute $\mathbf{R}_{\boldsymbol{\psi}_j}=\boldsymbol{\tau}$ and check the lower bound via Equations (24) and (25). Conversely, the existence of multiple nonzero elements in $\boldsymbol{\psi}_j$ implies that $|\widehat{\boldsymbol{\psi}_j}|^2$ is nonconstant, and thus (sim. Lemma VI.2) that $\mathbf{R}_{\boldsymbol{\psi}_j}$ has at least one nonzero coefficient for some nonzero lag $\tau$, leading to a contradiction. ■

*Proof of Proposition III.2.* Lemma VI.1 yields:

$$\langle\boldsymbol{w}*\boldsymbol{y}|\boldsymbol{w}'*\boldsymbol{y}'\rangle = \boldsymbol{w}_j^{\mathsf{T}}\cdot\mathbf{Q}_T(\boldsymbol{y},\boldsymbol{y}')\cdot\boldsymbol{w}_{j'} \tag{29}$$

By the independence of the Gaussian vectors $\boldsymbol{w},\boldsymbol{w}'$:

$$\begin{aligned}\mathbb{E}\big[\langle\boldsymbol{w}*\boldsymbol{y}|\boldsymbol{w}'*\boldsymbol{y}'\rangle\big] &= \mathbb{E}\big[\boldsymbol{w}_{j'}^{\mathsf{T}}\big]\cdot\mathbf{Q}_T(\boldsymbol{y},\boldsymbol{y}')\cdot\mathbb{E}\big[\boldsymbol{w}_j\big] \\ &= \mu^2\boldsymbol{\delta}^{\mathsf{T}}\cdot\mathbf{Q}_T(\boldsymbol{y},\boldsymbol{y}')\cdot\boldsymbol{\delta} \\ &= \mu^2\mathbf{Q}_T(\boldsymbol{y},\boldsymbol{y}')[0,0]. \end{aligned} \tag{30}$$

Recalling that $\mathbf{Q}_T(\boldsymbol{y},\boldsymbol{y}')[0,0]=\langle\boldsymbol{y}|\boldsymbol{y}'\rangle$ concludes the proof. ■

*Proof of Proposition III.3.* By Lemma VI.1:

$$\begin{aligned}&\mathbb{E}\big[\big|\langle\boldsymbol{w}*\boldsymbol{y}|\boldsymbol{w}'*\boldsymbol{y}'\rangle\big|^2\big] = \mathbb{E}\big[\langle\boldsymbol{w}*\boldsymbol{y}|\boldsymbol{w}'*\boldsymbol{y}'\rangle\overline{\langle\boldsymbol{w}*\boldsymbol{y}|\boldsymbol{w}'*\boldsymbol{y}'\rangle}\big] \\ &= \sum_{t,t',u,u'}\mathbf{R}_{(\boldsymbol{y}',\boldsymbol{y})}[t'-t]\overline{\mathbf{R}_{(\boldsymbol{y}',\boldsymbol{y})}[u'-u]}\mathbb{E}\big[\boldsymbol{w}[t]\boldsymbol{w}'[t']\boldsymbol{w}[u]\boldsymbol{w}'[u']\big], \end{aligned} \tag{31}$$

where the indices $t,t',u,u'$ range between zero and $(T-1)$. We distinguish six cases in the sum above:

1) if $t=t'=u=u'=0$, $\mathbb{E}\big[\boldsymbol{w}[t]^2\big]=\mathbb{E}\big[\boldsymbol{w}'[t']^2\big]=\mu^2+\sigma^2$. This term is weighted by $|\mathbf{R}_{(\boldsymbol{y}',\boldsymbol{y})}[0]|^2$.
2) if $t=t'=u=u'\neq 0$, $\mathbb{E}\big[\boldsymbol{w}[t]^2\big]=\mathbb{E}\big[\boldsymbol{w}'[t']^2\big]=\sigma^2$. There are $(T-1)$ such terms, weighted by $|\mathbf{R}_{(\boldsymbol{y}',\boldsymbol{y})}[0]|^2$.
3) if $0=t=u\neq t'=u'$, $\mathbb{E}\big[\boldsymbol{w}[t]^2\big]=\mu^2+\sigma^2$ and $\mathbb{E}\big[\boldsymbol{w}'[t]^2\big]=\sigma^2$. These terms are weighted by $|\mathbf{R}_{(\boldsymbol{y}',\boldsymbol{y})}[t']|^2$ for $0<t'<T$.
4) if $t=u\neq t'=u'=0$, $\mathbb{E}\big[\boldsymbol{w}[t]^2\big]=\sigma^2$ and $\mathbb{E}\big[\boldsymbol{w}'[t]^2\big]=\mu^2+\sigma^2$. These terms are weighted by $|\mathbf{R}_{(\boldsymbol{y}',\boldsymbol{y})}[-t]|^2$ for $0<t<T$.
5) if $0\neq t=u\neq t'=u'\neq 0$, $\mathbb{E}\big[\boldsymbol{w}[t]^2\big]=\sigma^2$ and $\mathbb{E}\big[\boldsymbol{w}'[t]^2\big]=\sigma^2$. These terms are weighted by $|\mathbf{R}_{(\boldsymbol{y}',\boldsymbol{y})}[t'-t]|^2$ for $0<t\neq t'<T$.
6) otherwise, the term equals zero.

We obtain: $\mathbb{E}\big[\big|\langle\boldsymbol{w}*\boldsymbol{y}|\boldsymbol{w}'*\boldsymbol{y}'\rangle\big|^2\big] = \alpha\mu^4+\beta\mu^2\sigma^2+\gamma\sigma^4$ with

$$\alpha = |\mathbf{R}_{(\boldsymbol{y}',\boldsymbol{y})}[0]|^2 = |\langle\boldsymbol{y}|\boldsymbol{y}'\rangle|^2, \tag{32}$$

$$\begin{aligned}\beta &= 2|\mathbf{R}_{(\boldsymbol{y}',\boldsymbol{y})}[0]|^2 + \sum_{\tau=1}^{T-1}\Big(|\mathbf{R}_{(\boldsymbol{y}',\boldsymbol{y})}[\tau]|^2 + |\mathbf{R}_{(\boldsymbol{y}',\boldsymbol{y})}[-\tau]|^2\Big) \\ &= 2\sum_{\tau=0}^{T-1}|\mathbf{R}_{(\boldsymbol{y}',\boldsymbol{y})}[\tau]|^2, \end{aligned} \tag{33}$$

$$\begin{aligned}\gamma &= T|\mathbf{R}_{(\boldsymbol{y}',\boldsymbol{y})}[0]|^2 + 2\sum_{\tau=1}^{T-1}|\mathbf{R}_{(\boldsymbol{y}',\boldsymbol{y})}[\tau]|^2 + \sum_{0<t\neq t'<T}|\mathbf{R}_{(\boldsymbol{y}',\boldsymbol{y})}[t'-t]|^2 \\ &= \sum_{t=0}^{T-1}\sum_{t'=0}^{T-1}|\mathbf{R}_{(\boldsymbol{y}',\boldsymbol{y})}[t'-t]|^2 = \sum_{\tau=-T}^{T}\big(T-|\tau|\big)|\mathbf{R}_{(\boldsymbol{y}',\boldsymbol{y})}[\tau]|^2. \end{aligned} \tag{34}$$

Proposition III.2 yields:

$$\begin{aligned}\mathbb{V}\big[\langle\boldsymbol{w}*\boldsymbol{y}|\boldsymbol{w}'*\boldsymbol{y}'\rangle\big] &= \mathbb{E}\big[\big|\langle\boldsymbol{w}*\boldsymbol{y}|\boldsymbol{w}'*\boldsymbol{y}'\rangle\big|^2\big] - \big|\mathbb{E}\big[\langle\boldsymbol{w}*\boldsymbol{y}|\boldsymbol{w}'*\boldsymbol{y}'\rangle\big]\big|^2 \\ &= \big(\alpha-|\langle\boldsymbol{y}|\boldsymbol{y}'\rangle|^2\big)\mu^4 + \beta\mu^2\sigma^2 + \gamma\sigma^4. \end{aligned} \tag{35}$$

After having replaced $\alpha$, $\beta$, and $\gamma$ by their definitions, we cancel the term $\big(\alpha-|\langle\boldsymbol{y}|\boldsymbol{y}'\rangle|^2\big)$, which completes the proof. ■

*Proof of Proposition III.4.* By Definition III.1, by independence of $\boldsymbol{w}_j$ and $\boldsymbol{w}_{j'}$, and by property of the Hermitian dot product in $\mathbb{C}^N$:

$$\begin{aligned}&\mathbb{E}\big[|(\boldsymbol{w}_j*\boldsymbol{\psi}_j)\wedge(\boldsymbol{w}_{j'}*\boldsymbol{\psi}_{j'})|^2\big] \\ &= \mathbb{E}\big[\|\boldsymbol{w}_j*\boldsymbol{\psi}_j\|_2^2\|\boldsymbol{w}_{j'}*\boldsymbol{\psi}_{j'}\|_2^2 - |\langle\boldsymbol{w}_j*\boldsymbol{\psi}_j|\boldsymbol{w}_{j'}*\boldsymbol{\psi}_{j'}\rangle|^2\big] \\ &= \mathbb{E}\big[\|\boldsymbol{w}_j*\boldsymbol{\psi}_j\|^2\big]\mathbb{E}\big[\|\boldsymbol{w}_{j'}*\boldsymbol{\psi}_{j'}\|^2\big] - \mathbb{E}\big[|\langle\boldsymbol{w}_j*\boldsymbol{\psi}_j|\boldsymbol{w}_{j'}*\boldsymbol{\psi}_{j'}\rangle|^2\big]. \end{aligned} \tag{36}$$

Applying Propositions II.2, III.2, and III.3 concludes the proof. ■

*Proof of Theorem III.5.* By Propositions II.2 and III.4. ■

## REFERENCES


[1] G. Richard, P. Chouteau, and B. Torres, "A fully differentiable model for unsupervised singing voice separation," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2024, pp. 946–950.

[2] N. Shlezinger, J. Whang, Y. C. Eldar, and A. G. Dimakis, "Model-based deep learning," *Proceedings of the IEEE*, vol. 111, no. 5, pp. 465–499, 2023.

[3] G. Richard, V. Lostanlen, Y.-H. Yang, and M. Müller, "Model-based deep learning for music information research: Leveraging diverse knowledge sources to enhance explainability, controllability, and resource efficiency," *IEEE Signal Processing Magazine*, vol. 41, no. 6, pp. 51–59, 2025.

[4] D. Haider, F. Perfler, V. Lostanlen, M. Ehler, and P. Balazs, "Hold me tight: Stable encoder-decoder design for speech enhancement," in *Proceedings of the International Speech Communication Association Conference (INTERSPEECH)*. ISCA, 2024.

[5] D. Haider, V. Lostanlen, M. Ehler, and P. Balazs, "Instabilities in convnets for raw audio," *IEEE Signal Processing Letters*, vol. 31, pp. 1084–1088, 2024.

[6] Y. Bengio, I. Goodfellow, and A. Courville, *Deep learning*. MIT Press, 2017.

[7] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2016, pp. 770–778.

[8] P. Balazs, D. Haider, V. Lostanlen, and F. Perfler, "Trainable signal encoders that are robust against noise," in *Proceedings of the International Congress and Exposition on Noise Control Engineering (INTER-NOISE)*, vol. 270, no. 10, 2024, pp. 1836–1844.

[9] V. Lostanlen, D. Haider, H. Han, M. Lagrange, P. Balazs, and M. Ehler, "Fitting auditory filterbanks with multiresolution neural networks," in *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*. IEEE, 2023.