Efficient and Robust Spike Ensemble Coding of Signals

Anonymous Author(s)
Affiliation
Address
email

Abstract

Sensory stimuli in animals are encoded into spike trains by neurons. We present a signal processing framework that deterministically encodes continuous-time signals into spike trains and addresses the question of reconstruction bounds. The framework encodes a signal through spike trains generated by an ensemble of neurons using a convolve-then-threshold mechanism with various convolution kernels. A closed-form solution to the inverse problem, from spike trains to signal reconstruction, is derived in the Hilbert space of shifted kernel functions, ensuring sparse representation of a generalized Finite Rate of Innovation (FRI) class of signals. Additionally, inspired by real-time processing in biological systems, an efficient iterative version of the optimal reconstruction is formulated that considers only a finite window of past spikes, ensuring robustness of the technique to ill-conditioned encoding; convergence guarantees of the windowed reconstruction to the optimal solution are then provided. Experiments on a large audio dataset demonstrate excellent reconstruction accuracy at spike rates as low as one-fifth of the Nyquist rate, while showing clear competitive advantage in comparison to state-of-the-art sparse coding techniques in the low spike rate regime.

1 Introduction

In most animals, sensory stimuli are communicated to the brain via ensembles of discrete, spatio-temporally compact electrical events generated by neurons, known as action potentials or spikes Rieke et al. [1999]. The conversion of continuous-time stimuli to spike trains occurs at an early stage of sensory processing, such as in the retinal ganglion cells in the visual pathway or spiral ganglion cells in the auditory pathway. Nature likely resorts to spike-based encoding due to several advantages: sparsity of representation [Olshausen and Field, 1996] and energy efficiency [Laughlin and Sejnowski, 2003], noise robustness [London and Häusser, 2005], high temporal precision [Buzsáki, 2006] and facilitation of downstream computation [Földiák, 1990, Graham and Field, 2007]. Based on how neural spike responses are represented, coding models can be broadly divided into two categories [Authors, 2023]: 1) rate coding, where spike train responses are converted into an average rate, and 2) temporal coding, where the precise timing of spikes convey information about the stimuli. In the rate coding literature, spike responses to stimuli are converted to an average instantaneous rate $r(t)$, and stimulus reconstruction is typically formulated probabilistically by choosing the stimulus s that maximizes the likelihood $P(s|r)$. Rate coding is criticized for losing temporal precision, especially since studies have shown that neurons can exhibit sub-millisecond precision [Mainen and Sejnowski, 1995]. Temporal coding, on the other hand, although emphasizes the precise timing of individual spikes, in this literature stimulus reconstruction is often formulated probabilistically (e.g. Bayesian Inference [Pillow et al., 2008]) or through a linear transformation of spike-responses (e.g. reverse correlation [Rieke et al., 1997]). Probabilistic approaches to reconstruction complicate the development of a deterministic signal processing framework from spike trains, while simple linear transformations may be too restrictive for representing a generalized class of signals. Recent advanced temporal coding schemes, such as those by Sophie et al. [Brendel et al., 2017], leverage recurrent

Submitted to 38th Conference on Neural Information Processing Systems (NeurIPS 2024). Do not distribute.

networks and have shown near perfect reconstruction for certain signals. However, these techniques involve complex training procedures and do not provide reconstruction guarantees for a generalized class of signals. This paper is therefore motivated to build a signal processing framework that deterministically encodes continuous-time signals into biologically feasible spike trains, addressing questions of representable signal classes and reconstruction bounds. While a preliminary version of this framework was introduced in [Chattopadhyay and Banerjee, 2023] for continuous-time signal coding and reconstruction, this work extends that framework by analyzing the stability of solutions and presenting more comprehensive experimental results. Specifically, the paper extends the analysis from section 5 onwards.

2 Coding

For encoding, we make the following assumptions: (1) We consider the set of input signals $\mathcal{F}$ to be the class of all *finite-support, bounded functions* (formally, $\mathcal{F} = \{X(t) | t \in [0, \tau], |X(t)| \leq b\}$, for some arbitrary but fixed $\tau, b \in \mathbb{R}^+$) that satisfy a finite rate of innovation bound [Vetterli et al., 2002]. Naturally, $X(t) \in L^2$, i.e., square integrable. (2) We assume an ensemble of m spiking neurons $\Phi = \{\Phi^j | j \in Z^+, 1 \leq j \leq m\}$, each characterized by a continuous kernel function $\Phi^j(t)$, where $\forall j, \Phi^j(t) \in C[0, \tau], \tau \in \mathbb{R}^+$. Also we assume that each kernel Φ^j is normalized, i.e., $||\Phi^j||_2 = 1, \forall j$. (3) Finally, we assume that Φ^j has a time varying threshold $T^j(t)$. The ensemble of kernels Φ encodes a given input signal $X(t)$ into a sequence of spikes $\{(t_i, \Phi^{j_i})\}$, where the ith spike is produced by the $j_i th$ kernel Φ^{j_i} at time t_i if and only if: $\int X(t)\Phi^{j_i}(t_i - t)dt = T^{j_i}(t_i)$.

In our implementation a threshold function is assumed in which the time varying threshold $T^j(t)$ of the jth kernel remains constant at C until that kernel produces a spike, at which time an *after-hyperpolarization potential (ahp)* increments the threshold by a value M. This increment then returns to zero linearly within a refractory period δ. Formally,

$$T^j(t) = C + \sum_{t_p^j \in [t-\delta, t]} M(1 - \frac{t - t_p^j}{\delta}) \quad (C, M, \delta \in \mathbb{R}^+) \tag{1}$$

where the sum is taken over all spike times t_p^j in the interval $[t - \delta, t]$ at which the kernel Φ^j generated a spike. We would like to note that the paper adheres to a consistent notation, which is clarified in the appendix. Various sections of the appendix include proofs to theorems as well.

Corollary 0.1. *Let Φ^j be a function in $C[0, \tau]$, where $\tau \in \mathbb{R}^+$ and $||\Phi^j||_2 = 1$. Let $X(t) \in \mathcal{F} = \{f(t) \mid t \in [0, \tau'], |f(t)| \leq b\}$, where $b, \tau' \in \mathbb{R}^+$, be the input to our model. Then: (a) The convolution $C^j(t)$ between $X(t)$ and $\Phi^j(t)$, defined by $C^j(t) = \int X(t')\Phi^j(t - t')dt'$ for $t \in [0, \tau + \tau']$, is a bounded continuous function. Specifically, one can show that $|C^j(t)| < b\sqrt{\tau}$ for all $t \in [0, \tau + \tau']$. (b) Suppose the parameter M in the Eq. 1 is chosen such that $M > 2b\sqrt{\tau}$. Then the interspike interval between any two spikes produced by the given neuron Φ^j is greater than $\frac{\delta}{2}$.*

Corollary 0.2. *Suppose the assumptions of the Corollary 0.1 hold true. Then, the spike rate generated by our framework for any input signal $X(t)$ is bounded. Consequently, the maximum number of preceding spikes that can overlap with any given spike is bounded above by a constant value.*

Corollary 0.3. *Let $X(t)$ be an input signal and Φ^j be a kernel. Suppose the convolution between $X(t)$ and $\Phi^j(t)$ at time t_p is denoted by $C^j(t_p) = \langle X(t), \Phi^j(t_p - t) \rangle > 0$, and let the absolute refractory period be δ as modeled in Eq. 1. If the baseline threshold C is set such that $0 < C \leq C^j(t_p)$, then the kernel Φ^j must produce a spike in the interval $[t_p - \delta, t_p]$, according to the threshold model defined in Eq. 1.*

Assumption 1. *$||\mathcal{P}_{\mathcal{S}(\{\phi_1, ..., \phi_{n-1}\})}(\phi_n)|| \leq \beta < 1$, for some $\beta \in \mathbb{R}$, $\forall n \in \{1, ..., N\}$, where N is the total number of spikes produced by the system. In words, the norm of the projection of every spike onto the span of all previous spikes is bounded from above by some constant strictly less than 1.*

Justification: See the appendix for justification of the assumption. This assumption supports the subsequent Theorem 1.

3 Decoding

The objective of the decoding module is to reconstruct the original signal from the encoded spike trains. Considering the prospect of the invertibility of the coding scheme, we seek a signal that

satisfies the same set of constraints as the original signal when generating all spikes apropos the set of kernels in ensemble Φ. Recognizing that such a signal might not be unique, we choose the reconstructed signal as the one with minimum $L2$-norm. Formally, the reconstruction X^* of the input signal X is formulated to be the solution to:

$$X^* = \underset{\tilde{X}}{\operatorname{argmin}} ||\tilde{X}||_2^2 \quad \text{s.t.} \int \tilde{X}(\tau)\Phi^{j_i}(t_i - \tau)d\tau = T^{j_i}(t_i); 1 \leq i \leq N \tag{2}$$

where $\{(t_i, \Phi^{j_i}) | i \in \{1, ..., N\}\}$ is the set of all spikes generated by the encoder. The choice of $L2$ minimization is in congruence with the dictum of energy efficiency in biological systems. The assumption is that, of all signals, the one with the minimum energy that is consistent with the spike trains is desirable. Also, an $L2$ minimization in the objective of (2) reduces the convex optimization problem to a solvable linear system of equations as described below.

4 Signal Class for Perfect Reconstruction

We observe that in general the encoding of $L^2[0, T]$ signals into spike trains is not an injective map; the same set of spikes can be generated by different signals so as to result in the same convolved values at the spike times. Naturally, with a finite and fixed ensemble of kernels Φ, one cannot achieve perfect reconstruction for all $L^2[0, T]$ signals. Assuming, additionally, a finite rate of innovation, as $\mathcal{F}$ was previously defined changes the story. We now restrict ourselves to a subset $\mathcal{G}$ of $\mathcal{F}$ defined as $\mathcal{G} = \{X | X \in \mathcal{F}, X = \sum_{p=1}^{N} \alpha_p \Phi^{j_p}(t_p - t), j_p \in \{1, ..., m\}, \alpha_p \in \mathbb{R}, t_p \in \mathbb{R}^+, N \in Z^+\}$ and address the question of reconstruction accuracy. Essentially $\mathcal{G}$ consists of all linear combinations of arbitrarily shifted inverted kernel functions. N is bounded above by the total number of spikes that the ensemble Φ can generate over $[0, T]$. For the class $\mathcal{G}$ the *perfect reconstruction theorem* is presented below. The theorem is proved with the help of two lemmas.

Theorem 1. *(Perfect Reconstruction Theorem) Let $X \in \mathcal{G}$ be an input signal. The solution X^* to the reconstruction problem Eq. (2) can be written as: $X^* = \sum_{i=1}^{N} \alpha_i \Phi^{j_i}(t_i - t), \alpha_i \in \mathbb{R}$. If the set of spikes $\{t_i, \Phi^{j_i}\}_{i=1}^{N}$ produced is* linearly independent *(as guaranteed by Assumption 1), the coefficients can be uniquely solved from a system of linear equations of the form $P\alpha = T$ (where P is an $N \times N$ gram matrix with $[P]_{ik} = \langle \Phi^{j_i}(t - t_i), \Phi^{j_k}(t - t_k) \rangle$ and T is an N-dim vector of thresholds $T^{j_i}(t_i)$). And therefore, for appropriately chosen time-varying thresholds of the kernels, the reconstruction X^*, is accurate with respect to the $L2$ metric, i.e., $||X^* - X||_2 = 0$.*

5 Approximate Reconstruction

Theorem 1 stipulates the conditions under which perfect reconstruction is feasible in the purview of our framework. Specifically the theorem shows the ideal conditions—when the input signal lies in the span of shifted kernel functions and the spikes are generated at certain desired locations—where perfect reconstruction is attainable. However, under realistic scenarios such conditions may not be feasible and hence the need for quantification of reconstruction error as the system deviates from the ideal conditions. For example, even though corollary 0.2 shows that a spike can be produced arbitrarily close to the desired location by setting the *ahp* parameters C and the δ of Eq. 1 at reasonably low values, it begs the question to what extent the reconstruction suffers due to small deviations in spike times. Likewise, the input signal may not perfectly fit in the signal class $\mathcal{G}$, i.e. the input may not be exactly representable by the kernel functions due to the presence of internal or external noise. Under such non-ideal scenarios how much the reconstruction suffers is addressed in the following theorem.

Theorem 2. *(Approximate Reconstruction Theorem). Let the input signal X be represented as $X = \sum_{i=1}^{N} \alpha_i f^{p_i}(t_i - t)$, where $\alpha_i \in \mathbb{R}$ and $f^{p_i}(t)$ are bounded functions on finite support. Assume that there is at least one kernel function Φ^{j_i} in the ensemble for which $||f^{p_i}(t) - \Phi^{j_i}(t)||_2 < \delta$ for all $i \in \{1, ..., N\}$. Additionally, assume each kernel Φ^{j_i} produces a spike within a γ interval of t_i, for some δ,$\gamma \in \mathbb{R}^+$, for all i. Further, assume the functions f^{p_i} satisfy a frame bound type of condition: $\sum_{k \neq i} \langle f_{p_i}(t - t_i), f_{p_k}(t - t_k) \rangle \leq \eta \; \forall \, i \in \{1, ..., N\}$, and that the kernel functions are Lipschitz continuous. Under such conditions, the L^2 error in the reconstruction X^* of the input X has bounded SNR.*

6 Stability of the solution and Windowed Iterative Reconstruction:

Theorem 2 shows that our technique manages reconstruction error under non-ideal conditions with suitable parameter choices in equation 1. This may increase spike rates, worsening the condition number of the P matrix and causing instability. The *ahp* mitigates this by maintaining spike separation, but the condition number worsens exponentially with more spikes (see Theorem 4 in appendix). To address this, we propose an approximate reconstruction scheme using a smaller window of past spikes, ensuring robustness. We then identify conditions for the convergence of this solution in assumption 2 and present the Windowing Theorem.

Windowed iterative reconstruction: Theorem 1 establishes that the reconstruction X^* is the projection onto the span of all spikes, i.e., $X^* = \mathcal{P}_{\mathcal{S}(\{\phi_1,...\phi_N\})}(X)$. This observation enables us to formulate the reconstruction iteratively by updating an existing reconstruction on a set of n spikes $\{\phi_i\}_1^n$ with each new incoming spike ϕ_{n+1}, instead of solving the $P\alpha = T$ equation for the full set of spikes as shown in lemma 1. The iterative update of the reconstruction then follows from:

$$\mathcal{P}_{\mathcal{S}(\bigcup_{i=1}^{n+1}\{\phi_i\})}(X) = \mathcal{P}_{\mathcal{S}(\bigcup_{i=1}^{n}\{\phi_i\})}(X) + \langle X, \phi_{n+1}^{\perp}\rangle \frac{\phi_{n+1}^{\perp}}{||\phi_{n+1}^{\perp}||^2} \tag{3}$$

where $\phi_{n+1}^{\perp}$ is the orthogonal complement of the additional $n+1$-th spike with respect to the span of all previous spikes, i.e. $\phi_{n+1}^{\perp} = \phi_{n+1} - \mathcal{P}_{\mathcal{S}\{\phi_1,...,\phi_n\}}(\phi_{n+1})$. The above iterative scheme Eq. 3, motivates us to formulate a windowed iterative reconstruction. Here, when a new spike ϕ_{n+1} appears, instead of calculating its orthogonal complement with respect to the span of all previous spike, the orthogonal projection of ϕ_{n+1} is computed with respect to the span of w previous spikes, where w is chosen as a fixed window size. Mathematically, for the $(n+1)$-th incoming spike, we define the *windowed iterative reconstruction*, $X_{n+1,w}^*$, for an input signal X with window size w iteratively as:

$$X_{n+1,w}^* = X_{n,w}^* + \langle X, \phi_{n+1,w}^{\perp}\rangle \frac{\phi_{n+1,w}^{\perp}}{||\phi_{n+1,w}^{\perp}||^2} \tag{4}$$

where $\phi_{n+1,w}^{\perp}$ is defined as:

$$\phi_{n+1,w}^{\perp} = \phi_{n+1} - \mathcal{P}_{\mathcal{S}(\bigcup_{n-w+1}^{n}\{\phi_i\})}(\phi_{n+1})$$

The idea is that $\phi_{n+1,w}^{\perp}$ closely approximates $\phi_{n+1}^{\perp}$ for reasonably large window size w, allowing us to formulate an iterative reconstruction based only on a finite window of w spikes rather than inverting a large P-matrix of size $N \times N$ as formulated in lemma 1. Eq. 4 involves computing $\phi_{n+1,w}^{\perp}$ for each new spike, derived by inverting the $w \times w$ gram matrix corresponding to the previous w spikes. Since w is chosen as a finite constant independent of N, it speeds up the decoding process and holds the condition number of the solution in check as per Theorem 4. The following Theorem 3 establishes how the window-based solution formulated in Eq. 4 converges to the optimal solution of Eq. 1 under an assumption feasible in the context of spikes produced by our framework via biological kernels. The assumption, an extension of assumption 1, is stated as follows:

Assumption 2. $||\mathcal{P}_{\mathcal{S}(\{\phi_1,...,\phi_N\}\setminus\{\phi_n\})}(\phi_n)|| \leq \beta < 1$, *for some* $\beta \in \mathbb{R}, \forall n \in \{1,...,N\}$. *In words, the norm of the projection of each spike onto the span of the remaining spikes is bounded above by a constant strictly less than 1.*

Justification: Appendix A.5 provides a detailed explanation for extending Assumption 1 and demonstrates why it is valid in the context of biological encoding.

Theorem 3 (Windowing Theorem)**.** *For an input signal X with bounded L_2 norm, suppose our framework produces a set of $n+1$ successive spikes $S = \{\phi_1, ..., \phi_{n+1}\}$, sorted by their time of occurrence and satisfying Assumption 2. The error in the iterative reconstruction of X with respect to the last spike ϕ_{n+1} due to windowing, as formulated in Eq. 4, is bounded. Specifically,*

$$\begin{aligned} &\forall \epsilon > 0, \exists w_0 > 0 \text{ s.t. } ||\mathcal{P}_{\phi_{n+1,w}^{\perp}}(X) - \mathcal{P}_{\phi_{n+1}^{\perp}}(X)|| < \epsilon, \\ &\forall w \geq w_0 \text{ and } w \leq n \end{aligned} \tag{5}$$

where w_0 is independent of n for arbitrarily large $n \in \mathbb{N}$.

7 Experiments on Real Signals

The proposed framework was tested on audio signals. **Dataset:** We chose the Freesound Dataset Kaggle 2018, an audio dataset of natural sounds referred in [Fonseca et al., 2018], containing 18,873 audio files. All audio samples in this dataset are provided as uncompressed PCM 16bit, 44.1kHz, mono audio files. **Set of Kernels:** We chose gammatone filters ($at^{n-1}e^{-2\pi bt}\cos(2\pi ft+\phi)$) in our experiments since they are widely used as a reasonable model of cochlear filters in auditory systems [Patterson et al., 1988]. **Results:** The proposed framework was tested extensively against the mentioned dataset. Comprehensive results with 600 randomly selected audio snippets using 50 kernels are shown in Figure 1(a). In the experiment, kernels were normalized, and parameters for the time-varying threshold function (1) were selected through a systematic grid search on a smaller dataset of 20 randomly chosen snippets. In each trial, an audio snippet of length $\approx 2.5s$ was processed with fixed parameter values, except for the refractory period, which was gradually decreased leading to improvement in reconstructions at higher spike rates. The refractory period varied from approximately $250ms$ to $5ms$. After converting an audio snippet into a sequence of spikes, reconstruction was performed iteratively using a fixed-sized window of spikes as described in Section 6. The *ahp* period was systematically varied, but each trial on an audio snippet was run with a single value of the *ahp* period. Correspondingly, the window size was set to be inversely proportional to the *ahp* period, varying from 5k to 15k as per Theorem 3. The results in Figure 1(a) show that increasing the spike rate by tuning the refractory period allows near-perfect reconstruction, aligning with our theoretical analysis. Some variability in reconstruction accuracy across signals can be attributed to dataset idiosyncrasies, e.g. certain audio samples could be noisy or ill-represented in the kernels. However, the overall trend shows promise, with an average of $\approx 20dB$ at $1/5$th Nyquist Rate. This in conjunction with the fact that signals are represented in this scheme only via set of spike times and kernel indexes (thresholds can be inferred) shows potential for an extremely efficient coding mechanism. Since the generation of spikes requires scanning through convolutions in a single pass, encoding is highly efficient. However, decoding is slightly more time-consuming because it involves solving the linear system $P\alpha = T$ to derive the coefficients. But then reconstruction is performed iteratively on a finite window, as described in Section 6. Considering $O(w^3)$ as the time complexity for inverting a $w \times w$ matrix, the overall time complexity of decoding is $O(Nw^3)$, where N is the length of the signal and w is the chosen window size. Thus the overall process still remains linear, making it a suitable choice for lengthy continuous-time signals.

Comparison With Convolutional Orthogonal Matching Pursuit: Our proposed framework is comparable to Convolutional Sparse Coding (CSC) techniques [Garcia-Cardona and Wohlberg, 2018]. In CSC, the objective is to efficiently represent signals using a small number of basis functions convolved with sparse coefficients. Finding sparse code for signals, in general is an NP-Hard problem [Davis and Avellaneda, 1997], and several heuristic-based approaches are used to address this challenge. In [Chattopadhyay and Banerjee, 2023], the proposed framework was compared against one such heuristic-based implementation of CSC, specifically Convolutional Orthogonal Matching Pursuit (COMP), using a small dataset of about 20 audio snippets. COMP employs a greedy technique to iteratively find dictionary atoms and is relatively slow due to the orthogonalization it performs to the atoms in each step. Most of the current leading CSC algorithms [Wohlberg, 2014, et. al., 2015, Wohlberg, 2016] use L_1 regularization as a relaxation of L_0 regularization in their sparse reconstruction objectives, making them amenable to efficient optimization methods such as the Alternating Direction Method of Multipliers (ADMM). In this paper, we extensively compare our technique with the efficient CBPDN algorithm implemented within the state-of-art SPORCO python library [Wohlberg, 2017]. In this experiment, we used 200 randomly chosen audio snippets from the same dataset, utilizing 10 gammatone kernels. The snippet lengths varied from 0.5s to 3.5s to compare processing times as a function of snippet length. Figure 1(b) shows the reconstruction accuracy comparison between our framework and CBPDN. CBPDN, which is based on L_1 optimization, achieved reconstructions at different sparsity levels by varying the regularization parameter λ, while our framework achieved different spike rates by varying the *ahp* period. The results in Figure 1(b) reveal that our framework outperforms CBPDN, particularly in the low spike rate regime, achieving better SNR values on average at consistently lower spike rates. In the high spike rate regime, CBPDN occasionally outperforms our framework. This occurs because our framework is not well suited for high spike rates, where spikes begin to overlap significantly, leading to poorer bounds as discussed in our theoretical analysis. Additionally, the limited datapoints where CBPDN outperforms (around 30 dB) make these specific results unreliable.

Overall, our framework shows promise as a superior alternative to CSC techniques in specific scenarios, particularly for low spike rate representation of natural audio signals with biological kernels, as demonstrated in our experiments.

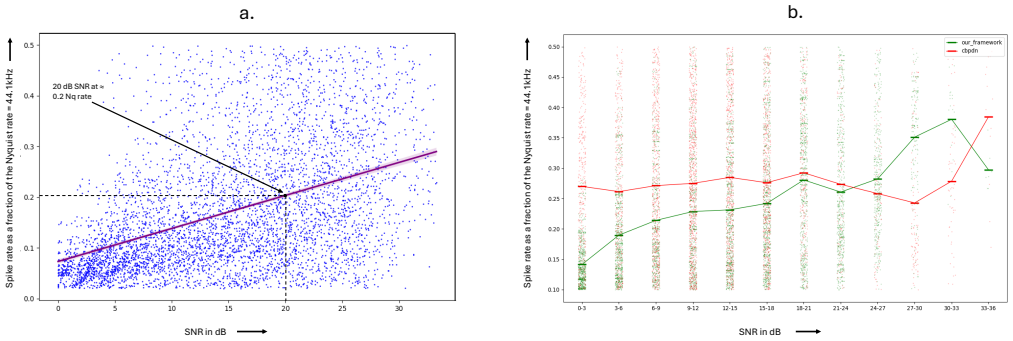

Figure 1: shows a sample reconstruction in an experiment with 10 kernels after training. **(a)** Comprehensive results of experiments on 600 audio snippets with 50 kernels. Scatter plot of reconstructions where each dot represents a single reconstruction performed on one of the 600 sound snippet for a particular setting of the *ahp* parameters as described in Section 7. The plot shows the SNR value of the reconstructions (x-axis) against corresponding spike-rate of the ensemble (y-axis). The trend line in purple is generated using seaborn regression fit, and the black dot on the line highlights the point on the trend line at 20 dB SNR which has an average spike-rate of approximately one-fifth the Nyquist rate. **(b)** Experimental results of comparison between our framework and cbpdn in terms of overall reconstruction accuracy and spike rate. Strip plot comparing spike rates of reconstruction between our framework and cbpdn. Like (a), the x-axis bins SNR values, and the strips of dots (red dots for cbpdn and green dots for our framework) in each bin show the distribution of reconstructions for that bin. The y-axis denotes the spike-rate of the ensemble as a fraction of the Nyquist rate for each reconstruction. The two lines through the strip plot connect the corresponding average spike rates of each bin for our framework (green) and cbpdn (red). As is evident, in the low spike rate regime the green line lies well below the red line, indicating a much lower spike rate achieved by our framework. Although briefly in the higher SNR regime, cbpdn outperforms our framework, the very few reconstruction data points available for cbpdn, as evidenced in (a), makes it challenging to draw reliable conclusions.

8 Conclusion

The experimental results establish the efficiency and robustness of the proposed spike-based encoding framework, which clearly outperforms state-of-the-art CSC techniques in the low spike rate regime. Notably, this high-fidelity coding and reconstruction is achieved through a simplified abstraction of a complex biological sensory processing system, which typically involves numerous neurons across multiple layers with diverse goals—such as feature extraction, decision making, classification and more—rather than being limited to reconstruction. For context, the human auditory system's cochlear nerve contains about 50,000 spiral ganglion cells (analogous to 50,000 kernels). The fact that our framework, with a single layer of roughly 100 neurons using a simple convolve-and-threshold model, achieves such high-quality reconstruction underscores the potential of fundamental biological signal processing principles. Our framework differs fundamentally from the Nyquist-Shannon theory, primarily in its mode of representation and coding. Instead of sampling the value of a function at uniform or non-uniform pre-specified points, our coding scheme identifies the non-uniform points where the function takes specific convolved values. This efficient coding scheme, combined with our proposed window-based fast processing of continuous-time signals, shows great promise for achieving significant compression in real-time signal communication.

References

Various Authors. A survey of encoding techniques for signal processing in spiking neural networks. *Neural Processing Letters*, 47(2):321–345, 2023.

Wieland Brendel, Ralph Bourdoukan, Pietro Vertechi, Christian K. Machens, and Sophie Deneve. Learning to represent signals spike by spike. *PLoS Computational Biology*, 13(10):e1005799, 2017.

György Buzsáki. *Rhythms of the Brain*. Oxford University Press, Oxford, UK, 2006. ISBN 9780195301069.

Anik Chattopadhyay and Arunava Banerjee. Beyond rate coding: Signal coding and reconstruction using lean spike trains. In *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1–5, 2023.

C.M. da Fonseca and J. Petronilho. Explicit inverses of some tridiagonal matrices. *Linear Algebra and its Applications*, 325(1-3):7–21, 2001. doi: 10.1016/S0024-3795(00)00271-5.

S. Davis, G.and Mallat and M. Avellaneda. Adaptive greedy approximations. *Constructive Approximation*, 13 (1):57–98, 1997.

Heide et. al. Fast and flexible convolutional sparse coding. In *2015 IEEE Conf. on CVPR*, pages 5135–5143, 2015.

P. Földiák. Forming sparse representations by local anti-hebbian learning. *Biological Cybernetics*, 64(2): 165–170, 1990.

Eduardo Fonseca, Manoj Plakal, Frederic Font, Daniel P. W. Ellis, Xavier Favory, Jordi Pons, and Xavier Serra. General-purpose tagging of freesound audio with audioset labels: Task description, dataset, and baseline, 2018.

Cristina Garcia-Cardona and Brendt Wohlberg. Convolutional dictionary learning: A comparative review and new algorithms. *IEEE Transactions on Comp. Imaging*, 4(3):366–381, 2018.

Daniel Graham and David Field. Sparse coding in the neocortex. *Evolution of Nervous Systems*, 3:181–187, 2007.

Simon B. Laughlin and Terrence J. Sejnowski. Communication in neuronal networks. *Science*, 301(5641): 1870–1874, 2003.

Michael S. Lewicki. Efficient coding of natural sounds. *Nature Neuroscience*, 5:356–363, 2002.

Michael London and Michael Häusser. Dendritic computation. *Annual Review of Neuroscience*, 28:503–532, 2005.

Zachary F. Mainen and Terrence J. Sejnowski. Reliability of spike timing in neocortical neurons. *Science*, 268: 1503–1506, 1995.

Bruno A. Olshausen and David J. Field. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381(6583):607–609, 1996.

R. Patterson, Ian Nimmo-Smith, J. Holdsworth, and P. Rice. An efficient auditory filterbank based on the gammatone function. 01 1988.

Jonathan W. Pillow, Jonathon Shlens, Liam Paninski, Alexander Sher, Alan M. Litke, E. J. Chichilnisky, and Eero P. Simoncelli. Spatio-temporal correlations and visual signaling in a complete neuronal population. *Nature*, 454:995–999, 2008.

Fred Rieke, David Warland, Rob de Ruyter van Steveninck, and William Bialek. *Spikes: Exploring the Neural Code*. MIT Press, Cambridge, MA, 1997. ISBN 9780262181783.

Fred Rieke, Davd Warland, Rob de Ruyter van Steveninck, and William Bialek. *Spikes: Exploring the Neural Code*. MIT Press, Cambridge, MA, USA, 1999. ISBN 0262181746.

Bernhard Schölkopf, Ralf Herbrich, and Alex J Smola. A generalized representer theorem. In *International Conference on Computational Learning Theory*, pages 416–426. Springer, 2001.

M. Vetterli, P. Marziliano, and T. Blu. Sampling signals with finite rate of innovation. *IEEE Transactions on Signal Processing*, 50(6):1417–1428, 2002.

Brendt Wohlberg. Efficient convolutional sparse coding. In *2014 IEEE Int. Conf. on Acoustics, Speech and Signal Proc. (ICASSP)*, pages 7173–7177, 2014.

Brendt Wohlberg. Boundary handling for convolutional sparse representations. In *2016 IEEE ICIP*, pages 1833–1837, 2016.

Brendt Wohlberg. SPORCO: A Python package for standard and convolutional sparse representations. In *Proceedings of the 15th Python in Sci. Conf.*, July 2017.

A Appendix / supplemental material

A.1 Notations followed in the paper:

- t_i: Time of occurrence of the ith spike.
- T_i: The threshold value.
- $X(t)$ or X: The input signal. For ease of notation we drop the time t as function argument and simply indicate the input as X instead of $X(t)$.
- X^*: The reconstructed signal.
- m: The number of kernels that comprise our framework.
- N: The total number of spikes produced by the system.
- (ϕ^{j_i}, t_i): Tuple denotes the ith spike produced by kernel ϕ^{j_i} at time t_i.
- $\Phi^{j_i}(t_i - t)$ or ϕ_i: The kernel function producing the ith spike, inverted and shifted to the time of the spike's occurrence t_i. For brevity's sake, this function is alternatively termed as the ith spike instead of the tuple notation (ϕ^{j_i}, t_i) and is denoted via the shorthand ϕ_i whenever appropriate. Also, the mathematical definition of the term "spike" must not be confused with the real physical object representing the elicitation of a neuron's action potential. These distinctions should be clear from the context.
- $\mathcal{S}(V)$: Subspace spanned by V in a Hilbert space $\mathcal{H}, V \subseteq \mathcal{H}$.
- $\mathcal{P}_v(u)$: In a Hilbert space $\mathcal{H}$, the projection of u on a vector v for $u, v \in \mathcal{H}$.
- $\mathcal{P}_{\mathcal{S}(V)}(u)$: In a Hilbert space $\mathcal{H}$ the projection of u on the subspace $\mathcal{S}(V)$ for $u \in \mathcal{H}, V \subseteq \mathcal{H}$. Note that this notation is similar to the above notation of $\mathcal{P}_v(u)$ except here the projection is taken w.r.t. a subspace $\mathcal{S}(V)$ instead of a single vector v. This should be clear from the context.

A.2 Proof of Corollary 0.1

Corollary 0.1: Let Φ^j be a function in $C[0,\tau], \tau \in \mathbb{R}^+$ and $||\Phi^j||_2 = 1$. Let $X(t) \in \mathcal{F} = \{f(t) | t \in [0,\tau'], |f(t)| \leq b\}$, where $b, \tau' \in \mathbb{R}^+$, be the input to our model. Then: (a) The convolution $C^j(t)$ between $X(t)$ and $\Phi^j(t)$ defined by $C^j(t) = \int X(t')\Phi^j(t-t')dt'$, for $t \in [0, \tau + \tau']$, is a bounded and continuous function. Specifically, one can show that $|C^j(t)| < b\sqrt{\tau}$ for all $t \in [0, \tau + \tau']$. (b) Suppose the parameter M in the threshold equation 1 is chosen such that $M > 2b\sqrt{\tau}$. Then the interspike interval between any two spikes produced by the given neuron Φ^j is greater than $\frac{\delta}{2}$.

Proof: (a) For showing continuity, we observe that $\Phi^j \in C[0,\tau]$ is uniformly continuous. Therefore, for $t, h \in \mathbb{R}$ we have:

$$|C^j(t) - C^j(t+h)| =$$
$$|\int \Phi^j(t-t')X(t')dt' - \Phi^j(t+h-t')X(t')dt'|$$

$$\leq (\sup_{t\in\mathbb{R}} |\Phi^j(t) - \Phi^j(t+h)|) \int |X(t')|dt'$$
$$\leq ||\Phi^j(t) - \Phi^j(t+h)||_\infty ||X||_1 \to 0 \text{ as } h \to 0$$

where the convergence follows from the fact that $\Phi^j(t)$ is uniformly continuous and $||X||_1$ is a bounded quantity as the input is a bounded function on compact support. A bound on $|C^j(t)|$ can be shown as follows-

$$
\begin{aligned}
&|C^j(t)| = |\int X(\tau)(\Phi^j(t-\tau))d\tau| \leq \int |X(\tau)||(\Phi^j(t-\tau))|d\tau \\
&< b\int |(\Phi^j(t-\tau)|d\tau \ \ (\text{since}|X(t)| < b) \\
&= b\int_0^\tau 1.|(\Phi^j(\tau)|d\tau = b||\Phi^j||_2\sqrt{\int_0^\tau 1 d\tau} \leq b\sqrt{\tau} \qquad (6) \\
&(\text{Using Cauchy-Schwarz inequality and } ||\Phi^j||_2 = 1)
\end{aligned}
$$

(b) Assume that $t_0, t_1, ..., t_k, t_{k+1}, ...t_L$ be a sequence of times at which kernel Φ^j produced a spike. We want to establish the above corollary by induction on this sequence. Let us denote the interspike intervals by δ_i, i.e. $\delta_i = t_i - t_{i-1}$ and the corresponding differences in the spiking thresholds by Δ_i, i.e. $\Delta_i = T^j(t_i) - T^j(t_{i-1}) = C^j(t_i) - C^j(t_{i-1}), \forall i \in \{1, ..., L\}$. Clearly, $\sum_{i=1}^k \Delta_i \geq 0, \forall k \in \{1, ..., L\}$, as any deviation from this condition would result in the convolution value falling below the baseline threshold C, consequently rendering the system incapable of producing a spike. Also, since the convolution $C^j(t)$ is a continuous function with $C^j(0) = 0$ (by definition) convolution value for the first spike equals the baseline threshold, i.e. $C^j(t_0) = C > 0$. Then, based on equation 6, $|C^j(t_k) - C^j(t_0)| = |\sum_{i=1}^k \Delta_i| < b\sqrt{\tau}, \forall k \in \{1, ..., L\}$. The threshold equation 1 formulates that for each spike at time t_i, due to the *ahp* effect the threshold value is immediately incremented by M and then within time δ the *ahp* effect linearly drops to zero. Let A_k^i denote the drop in the *ahp* due to the spike at time t_i in the interval δ_k. Note that $A_k^i = 0$ whenever $i \geq k$ and $\Sigma_k A_k^i = M$, i.e. the total *ahp* drop for each spike across all intervals is M. The proof follows by establishing the following two invariant by induction on k.

$$
\begin{aligned}
A_k^{k-1} &= M - \Sigma_{i=1}^k \Delta_i, \\
\delta_k &< \frac{\delta}{2} \quad \forall k \in \{1, ..., L\}
\end{aligned}
$$

The base case for $k = 1$, is clearly true. Because, $C^j(t_0) = T^j(t_0) = C$ and $C^j(t_1) = T^j(t_1) = C + M(1 - \frac{\delta_1}{\delta} I_{(\delta_1 < \delta)})$ (by equation 1). Here $I_{(\delta_1<\delta)} = 1$ if $\delta_1 < \delta$ and 0 otherwise. When $\delta_1 < \delta$, we get:

$$
\begin{aligned}
C^j(t_1) &= C + M(1 - \frac{\delta_1}{\delta}) \\
&= C^j(t_0) + \Delta_1 \\
&\Rightarrow \delta_1 = \frac{M - \Delta_1}{M}\delta > \frac{M - b\sqrt{\tau}}{M}\delta > \frac{\delta}{2}.
\end{aligned}
$$

And the *ahp* drop due to spike at t_0 in the interval δ_1 is: $A_1^0 = M - (C^j(t_1) - C^j(t_0)) = M - \Delta_1$. For the other case when $\delta_1 \geq \delta$, trivially $\delta \geq \frac{\delta}{2}$. Also in that case $C^j(t_1) = C$ since there is no spike in the previous δ interval of t_1 and therefore $\Delta_1 = 0$. But the *ahp* drop of the spike at t_0 in interval δ_1 is: $A_1^0 = M$ (since total drop is M) $= M - \Delta_1$. This establishes the invariants for the base case, $k = 1$.

Now for the induction step we assume that the invariants hold for all $k \in \{1, \ldots, n\}$, for some $n < L$, and we show that the invariants are true for $k = n + 1$. Specifically, we assume that $\delta_k > \frac{\delta}{2}$ and $A_k^{k-1} = M - \Sigma_{i=1}^k \Delta_i$, for all $k \leq n$. By assumption, $C^j(t_{n+1}) = C^j(t_n) + \Delta_{n+1} = T^j(t_{n+1})$. But $T^j(t_{n+1})$, the threshold at time t_{n+1}, is the sum of $T^j(t_n)$ and the *ahp* effect of the spike at time t_n and the *ahp* drop due to the spike at time t_{n-1}. Note that for $T^j(t_{n+1})$ we don't need to consider the *ahp* effects of spikes prior to t_{n-1} since the spikes prior to t_{n-1} are outside the δ interval of t_{n+1}

by induction assumption and therefore have zero *ahp* effect. Mathematically,

$$
\begin{aligned}
T^j(t_{n+1}) &= T^j(t_n) + M\left(1 - \frac{\delta_{n+1}}{\delta}\mathbf{I}_{(\delta_{n+1}<\delta)}\right) - A_{n+1}^{n-1} \\
\Rightarrow C^j(t_{n+1}) &= C^j(t_n) + M\left(1 - \frac{\delta_{n+1}}{\delta}\mathbf{I}_{(\delta_{n+1}<\delta)}\right) - A_{n+1}^{n-1} \\
&= C^j(t_n) + \Delta_{n+1} \\
\Rightarrow \Delta_{n+1} &= M\left(1 - \frac{\delta_{n+1}}{\delta}\mathbf{I}_{(\delta_{n+1}<\delta)}\right) - A_{n+1}^{n-1}.
\end{aligned}
$$

Therefore for the case when $\delta_{n+1} < \delta$ we get:

$$
\begin{aligned}
\Delta_{n+1} &= M\left(1 - \frac{\delta_{n+1}}{\delta}\right) - A_{n+1}^{n-1} \\
\Rightarrow M\frac{\delta_{n+1}}{\delta} &= M - \Delta_{n+1} - A_{n+1}^{n-1} \\
&\geq M - \Delta_{n+1} - \sum_{i=1}^{n}\Delta_i \quad (\text{since } A_{n+1}^{n-1} + A_n^{n-1} \leq M). \\
\Rightarrow M\frac{\delta_{n+1}}{\delta} &\geq M - \sum_{i=1}^{n+1}\Delta_i > M - b\sqrt{\tau} \Rightarrow \delta_{n+1} > \frac{\delta}{2}.
\end{aligned}
$$

Since $\delta_n + \delta_{n+1} > \delta$, the total drop in *ahp* during δ_n and δ_{n+1} due to spike at t_{n-1} is M, i.e. $A_{n+1}^{n-1} + A_n^{n-1} = M$. But,

$$
\begin{aligned}
A_{n+1}^{n-1} + A_{n+1}^{n} &= M - \Delta_{n+1} \\
\Rightarrow A_{n+1}^{n} &= M - \Delta_{n+1} - (M - A_n^{n-1}) = M - \sum_{i=1}^{n+1}\Delta_i,
\end{aligned}
$$

establishing the invariant. For the case $\delta_{n+1} \geq \delta$, trivially we get $\delta_{n+1} > \frac{\delta}{2}$. Since the *ahp* effect drops to zero for every spike within time δ , the threshold at time t_{n+1} must come down to the baseline value C, i.e. $T^j(t_{n+1}) = C = C^j(t_{n+1}) \Rightarrow \sum_{i=1}^{n+1}\Delta_i = 0$. Also, since $\delta_{n+1} \geq \delta$, the *ahp* drop due to the spike at t_n in interval δ_{n+1} is M, i.e. $A_{n+1}^{n} = M = M - \sum_{i=1}^{n+1}\Delta_i$, establishing the invariant for this case and hence, completing the proof.

A.3 Proof of Corollary 0.2

Corollary 0.2: Suppose the assumptions of the Corollary 0.1 hold true. Then, the spike rate generated by our framework for any input signal $X(t)$ is bounded. Consequently, the maximum number of preceding spikes that can overlap with any given spike is bounded above by a constant value.

Proof: Based on Corollary 0.1, the total spike rate is bounded above by $\frac{2m}{\delta}$, where m is the number of kernels employed by our model. Since each kernel $\Phi^j \in C[0, \tau]$ is compact support, the maximum number of preceding spikes any given spike can overlap with is bounded above by $\tau\frac{2m}{\delta}$, where τ is the maximum length of support for any kernel Φ^j.

A.4 Proof of Corollary 0.2

Corollary 0.3: Let $X(t)$ be an input signal and Φ^j be a kernel. Suppose the convolution between $X(t)$ and $\Phi^j(t)$ at time t_p is denoted by $C^j(t_p) = \langle X(t), \Phi^j(t_p - t)\rangle > 0$, and let the absolute refractory period be δ as modeled in Eq. 1. If the baseline threshold C is set such that $0 < C \leq C^j(t_p)$, then the kernel Φ^j must produce a spike in the interval $[t_p - \delta, t_p]$, according to the threshold model defined in Eq. 1.

Proof: We prove by contradiction, utilizing the continuity of the convolution function, $C^j(t)$. **Case 1:** No spike is produced by the kernel Φ^j before or at t_p. By definition $C^j(0) = 0$. However, since

$0 < C \leq C^j(t_p)$, the continuous function $C^j(t)$ must intersect the baseline threshold C between $t = 0$ and $t = t_p$ by the intermediate value theorem. This contradiction implies that the kernel Φ^j must produce a spike prior to or at time t_p. **Case 2:** Assuming spikes occurred before or at time t_p by the kernel Φ^j, let t_l be the time of last spike produced by Φ^j before or at t_p. Suppose $t_l < t_p - \delta$ to ensure no spike in the interval $[t_p - \delta, t_p]$. Then, by Eq. 1, the threshold of kernel Φ^j at t_p is $T^j(t_p) = C$. However, $C^j(t_p) = \langle X(t), \Phi^j(t_p - t)\rangle \geq C$. Since $C^j(t_l) = T^j(t_l)$ and $C^j(t)$ is continuous, and considering Eq. 1, where the *ahp* raises by a high value M at t_l and then linearly decreases to C before t_p, the intermediate value theorem implies that $C^j(t)$ must cross the threshold between t_l and t_p. However, since t_l was the last spike of Φ^j before t_p, this contradicts our assumption. Hence, there must be a spike in $[t_p - \delta, t_p]$.

A.5 Justification of Assumption 2

Corollary 0.1 suggests that for appropriately chosen parameters to the threshold Eq. 1, the spikes produced by the same kernel are sufficiently disjoint in time. Therefore, each new spike ϕ_n comes with a component that is disjoint in time with respect to the spikes produced by the same kernel. Since the biological kernels of our framework are causal in nature, due to the disjoint component in time a new spike maintains an orthogonal component with respect to all the previous spikes by the same kernel. For spikes produced by different kernels, we observe that different biological kernels correspond to different frequency responses (e.g., it has been observed that the responses of the auditory nerves can be well approximated by a bank of linear gammatone filters [Lewicki, 2002, Patterson et al., 1988]). Since there are only finitely many kernels in our framework, this leads to the fact that a new spike is poorly represented by the previous spikes produced by other kernels. Overall, a new spike Φ_n, for appropriately chosen *ahp* parameters in Eq. 1 will not be fully represented by previous spikes either due to disjoincy in time or frequency. Hence, the overall set of spikes grows as a *linearly independent* set. The technical need for this assumption will become clear in later sections.

A.6 Detailed Proof of the Perfect Reconstruction Theorem

The following section provides a detailed proof of the Perfect Reconstruction Theorem as presented in the main text. While the main text includes the theorem and its basic proof, this supplementary section offers a full proof of Lemma 2 and includes a formal claim that is referenced in the main text but not explicitly stated there. These additions provide a more comprehensive understanding and are included here to complement the content of the main text.

The theorem assumes that the input signal belongs to a more restrictive class of signals, $\mathcal{G}$, which is a subset of the class $\mathcal{F}$. The class of input signals was initially modeled by $\mathcal{F}$ in the Coding Section of the main text. The class $\mathcal{G}$ is defined as:

$\mathcal{G} = \{X | X \in \mathcal{F}, X = \sum_{p=1}^{N} \alpha_p \Phi^{j_p}(t_p - t), j_p \in \{1, ..., m\}, \alpha_p \in \mathbb{R}, t_p \in \mathbb{R}^+, N \in Z^+\}$. The theorem is restated below exactly as it appears in the main text, followed by its proof using two lemmas, as described in the original text.

Theorem 1. (Perfect Reconstruction Theorem) Let $X \in \mathcal{G}$ be an input signal. The solution X^* to the reconstruction problem Eq. (2) can be written as: $X^* = \sum_{i=1}^{N} \alpha_i \Phi^{j_i}(t_i - t), \alpha_i \in \mathbb{R}$. If the set of spikes $\{t_i, \Phi^{j_i}\}_{i=1}^{N}$ produced is *linearly independent* the coefficients can be uniquely solved from a system of linear equations of the form $P\alpha = T$ (where P is an $N \times N$ gram matrix with $[P]_{ik} = \langle \Phi^{j_i}(t - t_i), \Phi^{j_k}(t - t_k)\rangle$ and T is an N-dim vector of thresholds $T^{j_i}(t_i)$). And therefore, for appropriately chosen time-varying thresholds of the kernels, the reconstruction X^*, is accurate with respect to the $L2$ metric, i.e., $||X^* - X||_2 = 0$.

Lemma 1. *The solution X^* to the reconstruction problem Eq. (2) can be written as: $X^* = \sum_{i=1}^{N} \alpha_i \Phi^{j_i}(t_i - t)$ where the coefficients $\alpha_i \in \mathbb{R}$ can be uniquely solved from a system of linear equations if the set of spikes $\{\phi_i = \Phi^{j_i}(t_i - t)\}_{i=1}^{N}$ produced is* linearly independent.

Proof: An argument similar to that of the Representer Theorem [Schölkopf et al., 2001] on (2) directly results in: $X^* = \sum_{i=1}^{N} \alpha_i \Phi^{j_i}(t_i - t)$ where the α_i's are real valued coefficients. This holds true because any component of X^* orthogonal to the span of the $\Phi^{j_i}(t_i - t)$'s does not contribute to the convolution (inner product) constraints. In essence, X^* is an orthogonal projection of X

on the span of the spikes $\{\phi_i = \Phi^{j_i}(t_i - t) | i \in \{1, 2, ..., N\}\}$. Therefore, the coefficients can be derived by solving the linear system: $P\alpha = T$ where P is the $N \times N$ Gram matrix of the spikes, i.e., $[P]_{ik} = \langle \Phi^{j_i}(t_i - t), \Phi^{j_k}(t_k - t) \rangle$, and $T = \langle T^{j_1}(t_1), \ldots, T^{j_N}(t_N) \rangle^T$. Furthermore, the system has a unique solution if the Gram Matrix P is invertible. And the Gram Matrix P would be invertible if the set of spikes $\{\phi_i\}_{i=1}^N$ is linearly independent, which in turn follows from the assumption 1. We claim that even when the P-matrix is non-invertible, a unique reconstruction X^* can still be obtained following Eq. (2) in the main text, which can be calculated using the pseudo-inverse of P. □

Next, we prove the claim we just made about the uniqueness of existence of X^* as per Eq. (2). before proceeding to the next lemma and the subsequent proof of the theorem.

Claim: Let X be an input signal to our framework, generating a set of N spikes, $\{\phi_i = \Phi^{j_i}(t_i - t)\}_{i=1}^N$. Let X_1 and X_2 be two possible reconstructions of X from these N spikes, obtained by solving the optimization problem in Eq. (2) of the main text. Then $X_1 = X_2$.

Proof: The uniqueness of the reconstruction of X, as formulated in Eq. (2) follows from the fact that the reconstruction is essentially the projection of X onto the span of the spikes $\{\phi_i = \Phi^{j_i}(t_i - t)\}_{i=1}^N$. We now provide a formal proof. Let S be the subspace of L^2-functions spanned by $\{\phi_i = \Phi^{j_i}(t_i - t)\}_{i=1}^N$ with the standard inner product. Since each $\Phi^{j_i}(t_i - t)$ is assumed to be in L^2, S is a subspace of the larger space of all L^2-functions. Clearly, S is a Hilbert space with $dim(S) \leq N$. Therefore, there exists an orthonormal basis $\{e_1, ..., e_M\}$ for S, where $M \leq N$. Assume for contradiction that $X_1 \neq X_2$. Then there exist coefficients $\{a_i\}$ and $\{b_i\}$ such that $X_1 = \sum_{i=1}^M a_i e_i$ and $X_2 = \sum_{i=1}^M b_i e_i$, where not all a_i are equal to the corresponding b_i. Hence, there exists some k such that $a_k \neq b_k$, which implies:

$$\langle X_1, e_k \rangle = a_k \neq b_k = \langle X_2, e_k \rangle$$

However, since e_k is in the span of $\{\phi_i = \Phi^{j_i}(t_i - t)\}_{i=1}^N$, there exists $\{c_1, ..., c_N\}$ such that $e_k = \sum_{i=1}^N c_i \Phi^{j_i}(t_i - t)$. Therefore:

$$\langle X_1, e_k \rangle = \sum_{i=1}^N c_i \langle X_1, \Phi^{j_i}(t_i - t) \rangle = \sum_{i=1}^N c_i T^{j_i}(t_i)$$

Since both X_1 and X_2 are solutions to the optimization problem in Eq. (2), it follows that:

$$\langle X_1, e_k \rangle = \sum_{i=1}^N c_i T^{j_i}(t_i) = \langle X_2, e_k \rangle$$

This leads to a contradiction to the assumption that $a_k \neq b_k$. Thus, $X_1 = X_2$, completing the proof. □

Lemma 2. *Let X^* be the reconstruction of an input signal X by our framework with $\{\phi_i = \Phi^{j_i}(t_i - t)\}_{i=1}^N$ being the set of generated spikes. Then, for any arbitrary signal $\tilde{X}$ within the span of the spikes given by $\tilde{X} = \sum_{i=1}^N a_i \phi_i,\ a_i \in \mathbb{R}$, the following holds: $||X - X^*|| \leq ||X - \tilde{X}||$.*

Proof:

$$||X(t) - \tilde{X}(t)|| = || \underbrace{X(t) - X^*(t)}_{\text{A}} + \underbrace{X^*(t) - \tilde{X}(t)}_{\text{B}} ||$$

$$\langle A, \Phi^{j_i}(t_i - t) \rangle = \langle X(t), \Phi^{j_i}(t_i - t) \rangle - \langle X^*(t), \Phi^{j_i}(t_i - t) \rangle, \quad \forall i \in \{1, 2, .., N\}$$

$$= T^{j_i}(t_i) - T^{j_i}(t_i) = 0 \text{ (by Eq (2) \& (2) of main text)}$$

$$\langle A, B \rangle = \langle A, \sum_{i=1}^N (\alpha_i - a_i)\phi_i \rangle \text{ (By Lemma1 } X^*(t) = \sum_{i=1}^N \alpha_i \phi_i)$$

$$= \sum_{i=1}^N (\alpha_i - a_i) \langle A, \phi_i \rangle = 0 \Rightarrow A \perp B$$

Therefore,

$$||X(t) - \tilde{X}(t)||^2 = ||A + B||^2 = ||A||^2 + ||B||^2 \ (A \perp B)$$
$$\geq ||A||^2 = ||X(t) - X^*(t)||^2$$
$$\Rightarrow ||X(t) - \tilde{X}(t)|| \geq ||X(t) - X^*(t)|| \qquad \square$$

Proof of the Theorem 1: The proof of the theorem follows directly from Lemma 2. Since the input signal $X \in \mathcal{G}$, let X be given by: $X = \sum_{p=1}^{N} \alpha_p \Phi^{j_p}(t_p - t) \quad (\alpha_p \in \mathbb{R}, t_p \in \mathbb{R}^+, N \in Z^+)$. Assume that the time varying thresholds of the kernels in our kernel ensemble Φ are set in such a manner that the following conditions are satisfied: $\langle X, \Phi^{j_p}(t_p - t)\rangle = T^{j_p}(t_p) \quad \forall p \in \{1, ..., N\}$ i.e., each of the kernels Φ^{j_p} at the very least produces a spike at time t_p against X (regardless of other spikes at other times). Clearly then X lies in the span of the set of spikes generated by the framework. Applying Lemma 2 it follows that: $||X - X^*||_2 \leq ||X - X||_2 = 0$. $\square$

A.7 Proof of Condition Number Theorem

Theorem 4 (Condition Number Theorem)**.** *Let $\{P_k\}$ denote the set of all Gram matrices corresponding to any set of k successive spikes $\{\phi_1, \ldots, \phi_k\}$, i.e., $P_k[i,j] = \langle \phi_i, \phi_j \rangle$, where each spike satisfies the Assumption 1: $||\mathcal{P}_{S(\{\phi_1,\ldots,\phi_{i-1}\})}(\phi_i)|| \leq \beta < 1, \ \forall i \in \{2, \ldots, k\}$. If C_k denotes the least upper bound on the condition number of the class of matrices $\{P_k\}$, then* $(1-\beta^2)^{-k+1} \leq C_k \leq (1 + (k-1)\beta)(\frac{1-\beta^2}{2})^{-k+1}$

Proof: The condition number of P_k is defined as $\frac{\Lambda_{max}(P_k)}{\Lambda_{min}(P_k)}$, where $\Lambda_{max}(P_k)$ and $\Lambda_{min}(P_k)$ denote the maximum and minimum eigen values of P_k. First we find the infimum L_k on Λ_{min} over all $\{P_k\}$, the class of all gram-matrices of k successive spikes. We find the infimum inductively on k. By definition,

$$\Lambda_{min}(P_k) = \min_{e,||e||=1} ||\Sigma_{i=1}^{k} e_i \phi_i||^2$$

(where $e = [e_1, ..., e_k]^T$, a k-vector)

$$= \min_{e,||e||=1} [e_k^2 + 2e_k \langle \phi_k, \Sigma_{i=1}^{k-1} e_i \phi_i \rangle + ||\Sigma_{i=1}^{k-1} e_i \phi_i||^2]$$

$$\geq \min_{e,||e||=1} [e_k^2 - 2|e_k|\beta\sqrt{1-e_k^2}||\Sigma_{i=1}^{k-1} \frac{e_i}{\sqrt{1-e_k^2}} \phi_i|| +$$
$$||\Sigma_{i=1}^{k-1} \frac{e_i}{\sqrt{1-e_k^2}} \phi_i||^2] \quad (\text{since, } ||\mathcal{P}_{S\{\phi_1,\ldots,\phi_{i-1}\}}(\phi_i)|| \leq \beta)$$

$$\geq \min_{e,||e||=1} [e_k^2 - 2|e_k|\beta z\sqrt{1-e_k^2} + (1-e_k^2)z^2] \tag{7}$$

$$(\text{denoting, } z = ||\Sigma_{i=1}^{k-1} \frac{e_i}{\sqrt{1-e_k^2}} \phi_i||)$$

Now we set $|e_k| = \cos\theta, |e_k| \leq 1$ to obtain:

$$\Lambda_{min}(P_k) \geq \min_{e,||e||=1} [\cos^2\theta - 2\beta z \cos\theta \sin\theta + z^2 \sin^2\theta]$$

$$\geq \min_{e,||e||=1} [\frac{1+z^2}{2} + \frac{1-z^2}{2}\cos 2\theta - \beta z \sin 2\theta]$$

$$\geq \min_{e,||e||=1} [\underbrace{\frac{1+z^2}{2} - \sqrt{(\frac{1-z^2}{2})^2 + \beta^2 z^2}}_{g(z^2)}] \tag{8}$$

But, $z^2 = ||\Sigma_{i=1}^{k-1} \frac{e_i}{\sqrt{1-e_k^2}} \phi_i||^2 \geq L_{k-1}$

(Since, $\Sigma_{i=1}^{k-1} \frac{e_i^2}{1-e_k^2} = 1$ we get this inductively)

Since for $|\beta| < 1$ the expression $g(z^2)$ in 8 is a monotonic in z^2, and L_k is the infimum of $\Lambda_{min}(P_k)$, we may write,

$$L_k \geq [\frac{1+L_{k-1}}{2} - \sqrt{(\frac{1-L_{k-1}}{2})^2 + \beta^2 L_{k-1}}] \tag{9}$$

But L_k is a lower bound of Λ_{min}, and all the inequalities above are tight. Specifically, Eq. 7 & 8 show how given P_{k-1}, a Gram matrix of $k-1$ successive spikes with $\Lambda_{min} = L_{k-1}$, one can choose ϕ_k and e to result in a matrix P_k, so that $\Lambda_{min}(P_k)$ achieves the lower bound of 9. Therefore, the inequality of 9 can be turned into an equality.

$$L_k = [\frac{1+L_{k-1}}{2} - \sqrt{(\frac{1-L_{k-1}}{2})^2 + \beta^2 L_{k-1}}]$$

$$L_k = [\frac{(1-\beta^2)L_{k-1}}{\frac{1+L_{k-1}}{2} + \sqrt{(\frac{1-L_{k-1}}{2})^2 + \beta^2 L_{k-1}}}] \tag{10}$$

Since, $|\beta| \geq 0$, setting $|\beta| = 0$ in denominator of 10,

$$L_k \leq (1-\beta^2)L_{k-1} \tag{11}$$

Again, $L_1 = 1$ and using induction we can get $L_k \leq 1$.
Therefore, setting $L_{k-1} = 1, \beta = 1$ in denominator of 10,

$$L_k \geq \frac{(1-\beta^2)L_{k-1}}{2} \tag{12}$$

$$\Rightarrow \frac{(1-\beta^2)L_{k-1}}{2} \leq L_k \leq (1-\beta^2)L_{k-1} \qquad \text{(Using 12 \& 11)}$$

$$\Rightarrow (\frac{1-\beta^2}{2})^{k-1} \leq L_k \leq (1-\beta^2)^{k-1} \tag{13}$$

Eq. 13 establishes a bound on the infimum of Λ_{min}. To complete the proof and establish a bound on C_k we need to show a bound on the supremum of $\Lambda_{max}(P_k)$, call it U_k.
A bound on U_k can be shown as follows:

$$\Lambda_{max}(P_k) \leq \sup_i (P_k[i,i] + \sum_{i \neq j} |P_k[i,j]|)$$

(using Gershgorin Circle Theorem)

$$= \sup_i \{\langle \phi_i, \phi_i \rangle + \sum_{i \neq j} |\langle \phi_i, \phi_j \rangle|\} \leq (1 + (k-1)\beta)$$

(Since $||\mathcal{P}_{S\{\phi_1,\ldots,\phi_{i-1}\}}(\phi_i)|| \leq \beta \Rightarrow |\langle \phi_i, \phi_j \rangle| \leq \beta$)

$$\Rightarrow 1 \leq U_k \leq (1 + (k-1)\beta) \tag{14}$$

($1 \leq U_k$ is trivial because $\Lambda_{max} = 1$ for $P_k = I$)

Combining 13 & 14 we get:

$$(1-\beta^2)^{-k+1} \leq C_k \leq (1 + (k-1)\beta)(\frac{1-\beta^2}{2})^{-k+1} \qquad \square$$

The above theorem provides a tight upper bound on the condition number of the P-matrix and clearly shows how the condition number can degrade even if the spikes are sufficiently disjoint in time, i.e. $\beta \approx 0$.

A.8 Proof of Approximate Reconstruction Theorem

Theorem 2. (Approximate Reconstruction Theorem)
Let the input signal X be represented as $X = \sum_{i=1}^{N} \alpha_i f^{p_i}(t_i - t)$, where $\alpha_i \in \mathbb{R}$ and $f^{p_i}(t)$ are

bounded functions on finite support that constitute the input signal. Assume that there is at least one kernel function Φ^{j_i} in the ensemble for which $||f^{p_i}(t) - \Phi^{j_i}(t)||_2 < \delta$ for all $i \in \{1, ..., N\}$. Additionally, assume that each of these kernels Φ^{j_i} produces a spike within a γ interval of t_i, for some δ and $\gamma \in \mathbb{R}^+$, for all i. Also, assume that the functions f^{p_i} satisfy a frame bound type of condition: $\sum_{k \neq i} \langle f_{p_i}(t - t_i), f_{p_k}(t - t_k) \rangle \leq \eta \; \forall \, i \in \{1, ..., N\}$, and that the kernel functions are Lipschitz continuous. Under such conditions, the L^2 error in the reconstruction X^* of the input X has bounded SNR. Specifically one can show that $\frac{\|X(t) - X^*(t)\|^2}{\|X(t)\|^2} \leq \frac{(\delta + C\gamma)^2 (x_{\max} + 1)}{1 - \eta}$, where $\eta < 1$, C is a Lipschitz constant, and $x_{max} \in [0, N-1]$ is a constant depending on the overlap of the components in the input representation.

Proof: By hypothesis each kernel Φ^{j_i} produces a spike at time $t_i' \; \forall i \in \{1, ..., N\}$. Let us call these spikes as *fitting spikes*. But the coding model might generate some other spikes against X too. Other than the set of *fitting spikes* $\{(t_i', \Phi^{j_i}) | i \in \{1, ..., N\}\}$, let $\{(\tilde{t}_k, \Phi^{\tilde{j}_k}) | k \in \{1, ..., M\}\}$ denote those extra set of spikes that the coding model produces for input X against the bag of kernels Φ and call these extra spikes as *spurious spikes*. Here, M is the number of spurious spikes. By Lemma1, the reconstruction of X, denoted X^*, can be represented as below:

$$X^* = \sum_{i=1}^{N} \alpha_i \Phi^{j_i}(t_i' - t) + \sum_{k=1}^{M} \tilde{\alpha_k} \Phi^{\tilde{j_k}}(\tilde{t_k} - t)$$

where α_i and $\tilde{\alpha_k}$ are real coefficients whose values can be formulated again from Lemma1. Let T_i be the thresholds at which kernel Φ^{j_i} produced the spike at time t_i' as given in the hypothesis. Hence for generation of the *fitting spikes* the following condition must be satisfied:

$$\langle X, \Phi^{j_i}(t_i' - t) \rangle = T_i \;\; \forall i \in \{1, 2, ..., N\} \tag{15}$$

Consider a hypothetical signal X_{hyp} defined by the equations below:

$$X_{hyp} = \sum_{i=1}^{N} a_i \Phi^{j_i}(t_i' - t), a_i \in R$$
$$\text{s.t. } \langle X_{hyp}, \Phi^{j_i}(t_i' - t) \rangle = T_i, \forall i \tag{16}$$

Clearly this hypothetical signal X_{hyp} can be deemed as if it is the reconstructed signal where we are only considering the *fitting spikes* and ignoring all *spurious spikes*. Since, X_{hyp} lies in the span of the shifted kernels used in reconstruction of X using Lemma 3 we may now write:

$$||X - X_{hyp}|| \geq ||X - X^*|| \tag{17}$$
$$||X - X_{hyp}||_2^2 = \langle X - X_{hyp}, X - X_{hyp} \rangle$$
$$= \langle X - X_{hyp}, X \rangle - \langle X - X_{hyp}, X_{hyp} \rangle$$
$$= \langle X - X_{hyp}, X \rangle - \Sigma_{i=1}^{N} a_i \langle X - X_{hyp}, \Phi^{j_i}(t - t_i') \rangle$$
$$= ||X||_2^2 - \langle X, X_{hyp} \rangle$$
(Since by construction $\langle X_{hyp}, \Phi^{j_i}(t - t_i') \rangle = T_i \; \forall i \in \{1 ... N\}$)
$$= \Sigma_{i=1}^{N} \Sigma_{k=1}^{N} \alpha_i \alpha_k \langle f_i(t - t_i), f_k(t - t_k) \rangle$$
$$- \Sigma_{i=1}^{N} \Sigma_{k=1}^{N} \alpha_i a_k \langle f_i(t - t_i), \Phi^{j_k}(t - t_k')) \rangle$$
$$= \alpha^T F \alpha - \alpha^T F_K a \tag{18}$$
(denoting $a = [a_1, a_2, ..., a_N]^T, \alpha = [\alpha_1, \alpha_2, ..., \alpha_N]^T$,
$F = [F_{ik}]_{N \times N}$, an $N \times N$ matrix, where $F_{ik} =$
$\langle f_i(t - t_i), f_k(t - t_k) \rangle$ and $F_K = [(F_K)_{ik}]_{NXN}$ where
$(F_K)_{ik} = \langle f_i(t - t_i), \Phi^{j_k}(t - t_k') \rangle$)
But using Lemma1 a can be written as:
$a = P^{-1} T$, $P = [P_{ik}]_{NXN}, P_{ik} = \langle \Phi^{j_i}(t - t_i'), \Phi^{j_k}(t - t_k') \rangle$
And, $T = [T_i]_{N \times 1}$ where $T_i = \langle X(t), \Phi^{j_i}(t - t_i') \rangle$
$= \Sigma_{k=1}^{N} \alpha_k \langle f_k(t - t_k), \Phi^{j_i}(t - t_i') \rangle = F_K^T \alpha$
$\Rightarrow a = P^{-1} F_K^T \alpha$

Plugging this expression of a in equations (18) we get,

$$||X - X_{hyp}||_2^2 = \alpha^T F\alpha - \alpha^T F_K P^{-1} F_K^T \alpha \tag{19}$$

But,$(F_K)_{ik} = \langle f_i(t-t_i), \Phi^{j_k}(t-t'_k)\rangle$

$= \langle \Phi^{j_i}(t-t'_i), \Phi^{j_k}(t-t'_k)\rangle$

$\quad - \langle \Phi^{j_i}(t-t'_i) - f_i(t-t_i), \Phi^{j_k}(t-t'_k)\rangle$

$= (P)_{ik} - (\mathcal{E}_K)_{ik}$ (20)

(denoting $\mathcal{E}_K = [(\mathcal{E}_K)_{ik}]_{N\times N}$,

where $(\mathcal{E}_K)_{ik} = \langle \Phi^{j_i}(t-t'_i) - f_i(t-t_i), \Phi^{j_k}(t-t'_k)\rangle)$

Also, $(F)_{ik} = \langle f_i(t-t_i), f_k(t-t_k)\rangle$

$= \langle f_i(t-t_i) - \Phi^{j_i}(t-t'_i) + \Phi^{j_i}(t-t'_i),$

$\quad f_k(t-t_k) - \Phi^{j_k}(t-t'_k) + \Phi^{j_k}(t-t'_k)\rangle$

$= (\mathcal{E})_{ik} - (\mathcal{E}_K)_{ik} - (\mathcal{E}_K)_{ki} + (P)_{ik}$ (21)

Combining (19), (20) and (21) we get,

$||X - X_{hyp}||_2^2 = \alpha^T F\alpha - \alpha^T F_K P^{-1} F_K^T \alpha$

$= \alpha^T \mathcal{E}\alpha - \alpha^T \mathcal{E}_K\alpha - \alpha^T \mathcal{E}_K^T\alpha + \alpha^T P\alpha$

$- \alpha^T P\alpha + \alpha^T \mathcal{E}_K\alpha + \alpha^T \mathcal{E}_K^T\alpha - \alpha^T \mathcal{E}_K P{-}1\mathcal{E}_K^T\alpha$

$= \alpha^T \mathcal{E}\alpha - \alpha^T \mathcal{E}_K P^{-1}\mathcal{E}_K^T\alpha \leq \alpha^T \mathcal{E}\alpha$

(Since, P is an SPD matrix, $\alpha^T \mathcal{E}_K P^{-1}\mathcal{E}_K^T\alpha > 0$) (22)

We seek for a bound for the above expression. For that we observe the following:

$|(\mathcal{E})_{ik}| = |\langle f_i(t-t_i) - \Phi^{j_i}(t-t_i^{'}), f_k(t-t_k) - \Phi^{j_k}(t-t_k^{'})\rangle|$

$= ||f_i(t-t_i) - \Phi^{j_i}(t-t_i^{'})||_2 ||f_k(t-t_k) - \Phi^{j_k}(t-t_k^{'})||_2 . x_{ik}$

(where $x_{ik} \in [0,1]$. We also note that x_{ik} is close to 0 when the overlaps in the supports of the two components and their corresponding fitting kernels are minimal.)

$\Rightarrow (\mathcal{E})_{ik} \leq (|||(f_i(t-t_i) - \Phi^{j_i}(t-t_i)||+$

$\quad ||\Phi^{j_i}(t-t_i) - \Phi^{j_i}(t-t_i^{'}))||).$

$\quad (||f_k(t-t_k) - \Phi^{j_k}(t-t_k)||$

$\quad + ||\Phi^{j_k}(t-t_k) - \Phi^{j_k}(t-t_k^{'})||).x_{ik}$

$\Rightarrow (\mathcal{E})_{ik} \leq x_{ik}.(\delta + C\gamma)^2$ (23)

(Assuming C is a Lipschitz constant that each kernel Φ^{j_i} satisfies, and by assumption we have $|t_i - t_i^{'}| < \delta$ for all i.)

Now, using Gershgorin circle theorem, the maximum eigen value of $\mathcal{E}$ can be obtained as follows:

$\Lambda_{max}(\mathcal{E}) \leq max_i((\mathcal{E})_{ii} + \Sigma_{k\neq i}|(\mathcal{E})_{ik}|)$

$\quad \leq (\delta + C\gamma)^2(x_{max} + 1)$ (Using (23)) (24)

(where $x_{max} \in [0, N-1]$ is a positive number that depends on the maximum overlap of the supports of the component signals and their fitting kernels.)

Similarly, the minimum eigen value of F is:

$\Lambda_{min}(F) = min_i((F)_{ii} - \Sigma_{i\neq k}|\langle f_{p_i}(t-t_i), f_{p_k}(t-t_k)\rangle|)$

$\quad \geq 1 - \eta$ (25)

(By assumption $\Sigma_{i\neq k}| < f_{p_i}(t-t_i), f_{p_k}(t-t_k) > | \leq \eta$)

Combining the results from (22), (24) and (25) we get:

$$
\begin{aligned}
&||X(t)-X_{hyp}(t)||^2/||X(t)||^2 \leq \alpha^T \mathcal{E}\alpha/\alpha^T F\alpha \\
&\leq \Lambda_{max}(\mathcal{E})/\Lambda_{min}(F) \\
&\leq (\delta + C\gamma)^2(x_{max}+1)/(1-\eta) \qquad (26)
\end{aligned}
$$

Finally using (17) we conclude,

$$
\begin{aligned}
||X(t)-X^*(t)||^2/||X(t)||^2 &\leq ||X(t)-X_{hyp}(t)||^2/||X(t)||^2 \\
&\leq (\delta + C\gamma)^2(x_{max}+1)/(1-\eta)
\end{aligned}
$$

where $\eta < 1$, and $x_{max} \in [0, N-1]$, a constant depending on the overlap of the components in the input representation. □

A.9 Justification of Assumption 2

Assumption 2 extends Assumption 1 to the future. For large N, it is possible to construct a spike sequence where each spike satisfies Assumption 1, yet the norm of the projection of an individual spike onto the set of remaining spikes (including both past and future spikes), i.e. $||\mathcal{P}_{\mathcal{S}(\bigcup_1^N\{\phi_i\}\setminus\{\phi_n\})}(\phi_n)|| \to 1$. An example illustrating this scenario is shown in Fig. 5, where each spike corresponds to one full cycle of a sine wave, and for any given spike ϕ_n the half wave of its tail (head) precisely overlaps with the half wave of the head (tail) of its previous (next) spike ϕ_{n-1} (ϕ_{n+1}). One can show that in such a scenario $||\mathcal{P}_{\mathcal{S}(\bigcup_1^N\{\phi_i\}\setminus\{\phi_n\})}(\phi_n)||$ converges to 1 for large values of n and N (see Appendix A.10 for details). This convergence occurs because within the compact support of a spike, the spike is fully represented by the overlapping components of the neighboring spikes. Conversely, if a spike is poorly represented within its compact support by the overlapping components of its neighboring spikes, it satisfies the condition of Assumption 2. Specifically, if within its compact support, a spike ϕ_n can produce a component $\hat{\phi_n}$, orthogonal to all overlapping parts of the neighboring spikes with $||\hat{\phi_n}|| > 0$, then ϕ_n satisfies the condition of Assumption 2, because $\hat{\phi_n}$ is orthogonal to the overlapping parts of the neighboring spikes and hence $\hat{\phi_n}$ is orthogonal to every spike other than ϕ_n. This observation enables us to assert Assumption 2 for spikes produced by our framework via the biological kernels (e.g. the gammatone kernels corresponding to the auditory processing [Patterson et al., 1988]) which inherently exhibit causal and fading memory properties, so that the overlapping parts of the neighboring spikes do not represent the given spike within its support. Figure 2 visually illustrates this property using the dot product of overlapping portions of gammatone kernels. We show that due to the *ahp*, overlapping spikes corresponding to the same gammatone kernel poorly represent each other as they temporally separate. Additionally, spikes corresponding to gammatones of different frequencies poorly represent each other, irrespective of temporal shifts. Thus, with appropriately chosen *ahp* parameters, the spikes produced by our framework's biological kernels are poorly represented within their supports by overlapping components of neighboring spikes, either due to temporal disjoincy or frequency mismatch. Since every spike overlaps with only finitely many other spikes (corollary 0.2), Assumption 2 holds.

A.10 Analysis of Overlapping Kernels in Figure 5

This section contains an analysis of the overlapping sine kernels scenario depicted in Figure 5. The assumption 1 from the main text states that the norm of the projection of each spike onto the span of all previous spikes is bounded from above by some constant strictly less than 1. In other words, each incoming spike has a component orthogonal to the span of all previous spikes, the norm of which is bounded below by some constant strictly greater than 0. However, this assumption does not necessarily imply that each spike is poorly represented in the span of the set of all other spikes, including both past and future spikes. Figure 5 provides a counterexample to this by constructing a sequence of spikes where, despite each spike having a bounded orthogonal component with respect to all previous spikes, one can show that a particular spike in this sequence can be almost perfectly represented by all others. Specifically, the projection of one spike onto the span of the others approaches a norm of 1 as the sequence length increases. In the figure, each spike is generated by a kernel function that consists of a single cycle of a sine wave. The spikes are arranged so that tail of one spike perfectly overlaps with the head of the previous spike. In this setup, each spike overlaps only with one previous spike and one subsequent spike.

Claim: Let $\phi_1, ..., \phi_N$ be the sequence of spikes arranged in time as shown in Figure 5, where each spike is generated by a normalized sine wave kernel. In this setup, the projection of any spike

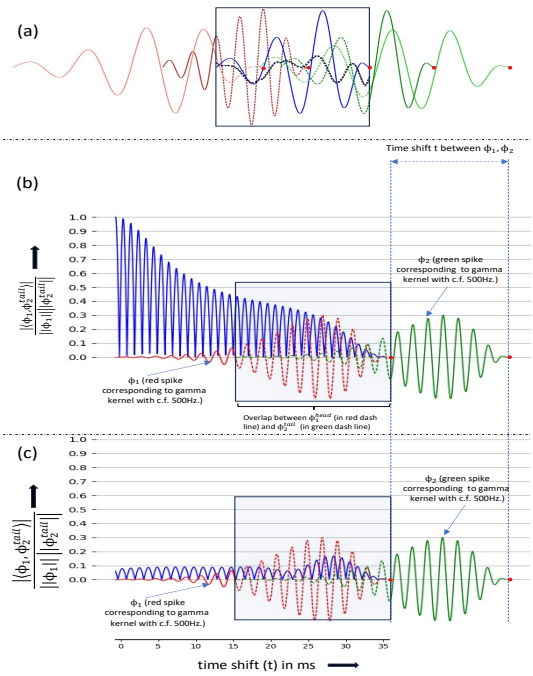

Figure 2: Illustration of why a spike produced by our framework using gammatone kernels is poorly represented by other overlapping spikes within its temporal support. **(a)** Shows a diagram of five spikes; the central spike (blue dashed line within the rectangle) overlaps with two past spikes (red) and two future spikes (green). Despite the overlaps, the distinct shapes of the gammatone kernels result in a poor representation of the central spike, leaving a component orthogonal to its neighbors (shown in black). **(b)** Details the interaction between two spikes, ϕ_1 and ϕ_2 (red and green curves), both using a gammatone kernel with a center frequency of 500 Hz. It examines how well ϕ_1 is represented by the overlapping tail of ϕ_2 (denoted ϕ_2^{tail}, green dashed line), focusing on representation within this overlapping support only. The graph in blue measures the dot product $\frac{|\langle \phi_1, \phi_2^{tail} \rangle|}{||\phi_1|| ||\phi_2^{tail}||}$ as a function of the time lag t between the two spikes, illustrating how the representation deteriorates rapidly as the lag increases, thus demonstrating the poor representational quality induced by the *ahp-effect*. **(c)** A similar plot of interaction between two spikes, ϕ_1 and ϕ_2 as in **(b)**, except here the center frequencies of the corresponding gammatone kernels are different (500 Hz for ϕ_1 and 400 Hz for ϕ_2). The blue graph represents the dot product $\frac{|\langle \phi_1, \phi_2^{tail} \rangle|}{||\phi_1|| ||\phi_2^{tail}||}$ as a function of time shift t, highlighting systematic poor representation due to frequency differences. Each spike's time of occurrence is marked by a red dot.

ϕ_n onto the span of all other spikes in the sequence satisfies $||\mathcal{P}_{\mathcal{S}(\bigcup_1^N \{\phi_i\} \setminus \{\phi_n\})}(\phi_n)|| \to 1$ for $n \to \infty$ and $N - n \to \infty$.

Proof: Each spike ϕ_i can be decomposed into two components: the positive half wave at the tail of the spike, denoted ϕ_i^t, and the negative half-wave at the head of the spike, denoted ϕ_i^h. By assumption, each spike is normalized, so the norm of each component is $\frac{1}{\sqrt{2}}$. Due to how the spikes are aligned, $\phi_i^h = -\phi_{i+1}^t$ for all $i \in [1, N-1]$, which means that $\langle \phi_i, \phi_j \rangle = -\frac{1}{2}$ for $|i-j| = 1$ and 0 for $|i-j| > 1$. To show that a spike ϕ_n in this sequence is almost perfectly represented by the others, we calculate the projection of ϕ_n onto the span of all the other spikes. This projection can be expressed as the sum of two components: one projection onto the span of the preceding spikes and one onto the span of the succeeding spikes. The coefficients of these projections can be determined by solving a system of linear equations involving the Gram matrix formed by the inner products of the spikes similar to the approach used Lemma 1 from the main text. The projection can be written as follows:

$$\mathcal{P}_{\mathcal{S}(\bigcup_1^N \{\phi_i\} \setminus \{\phi_n\})}(\phi_n) = \mathcal{P}_{\mathcal{S}(\bigcup_1^{n-1} \{\phi_i\})}(\phi_n) + \mathcal{P}_{\mathcal{S}(\bigcup_{n+1}^N \{\phi_i\})}(\phi_n)$$

(As $\mathcal{S}(\cup_{n+1}^N \{\phi_i\}) \perp \mathcal{S}(\cup_1^{n-1} \{\phi_i\})$ due to disjoint support)

$$\Rightarrow \mathcal{P}_{\mathcal{S}(\bigcup_1^N \{\phi_i\} \setminus \{\phi_n\})}(\phi_n) = \sum_{i=1}^{n-1} \alpha_i \phi_i + \sum_{i=n+1}^{N} \alpha_i \phi_i; \alpha_i \in \mathbb{R}$$

The values of α_i for $i \in [1, n-1]$, following Lemma 1 from the main text, can be obtained by solving a linear system of equations of the form $P_1\alpha = T$ where P_1 is an $(n-1) \times (n-1)$ matrix and $\alpha = [\alpha_{n-1}, \cdots, \alpha_1]^T$. Similarly, the values of α_i for $i \in [n+1, N]$ can be obtained by solving a linear system of equations of the form $P_2\alpha' = T$, where P_2 is an $(N-n) \times (N-n)$ matrix and $\alpha' = [\alpha_{n+1}, \cdots, \alpha_N]^T$. Specifically we have the following.

$$P_1\alpha = \begin{bmatrix} 1 & -\frac{1}{2} & 0 & \cdots & 0 \\ -\frac{1}{2} & 1 & -\frac{1}{2} & \cdots & 0 \\ 0 & -\frac{1}{2} & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & -\frac{1}{2} \\ 0 & 0 & 0 & -\frac{1}{2} & 1 \end{bmatrix} \begin{bmatrix} \alpha_{n-1} \\ \vdots \\ \alpha_1 \end{bmatrix} = \begin{bmatrix} -\frac{1}{2} \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{27}$$

In the above equation (27), the matrix P_1 is a symmetric tridiagonal matrix with unit diagonal elements and constant off-diagonal entries, the inverse of which can be calculated using *Chebyshev polynomials of the second kind* [da Fonseca and Petronilho, 2001]. Specifically, the elements of P_1^{-1} are given by:

$$(P_1^{-1})_{i,j} = (-1)^{i+j+1} \cdot 2 \cdot \frac{U_{i-1}(-1) \cdot U_{(n-1)-j}(-1)}{U_{n-1}(-1)},\ i \le j$$

where U_k denotes order-k Chebyshev polynomial of second kind. This simplifies to:

$$\Rightarrow (P_1^{-1})_{i,j} = \begin{cases} 2 \cdot \frac{i(n-j)}{n} & \text{for } i \le j \\ 2 \cdot \frac{j(n-i)}{n} & \text{for } i > j \ \text{ (using symmetry)} \end{cases}$$

Now, using this and Eq. (27) we get:

$$\begin{bmatrix} \alpha_{n-1} \\ \vdots \\ \alpha_1 \end{bmatrix} = P_1^{-1} \begin{bmatrix} -\frac{1}{2} \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} -\frac{1}{2}(P_1^{-1})_{1,1} \\ -\frac{1}{2}(P_1^{-1})_{2,1} \\ \vdots \\ -\frac{1}{2}(P_1^{-1})_{n-1,1} \end{bmatrix} = \begin{bmatrix} -\frac{n-1}{n} \\ -\frac{n-2}{n} \\ \vdots \\ -\frac{1}{n} \end{bmatrix}$$

$$\Rightarrow \sum_{i=1}^{n-1} \alpha_i\phi_i = -\sum_{i=1}^{n-1} \frac{i}{n}\phi_i = -\sum_{i=1}^{n-1} \frac{i}{n}(\phi_i^h + \phi_i^t)$$

$$= -\sum_{i=2}^{n-1} \frac{i}{n}(\phi_i^h - \phi_{i-1}^h) - \frac{1}{n}\phi_1$$

$$= -\frac{n-1}{n}\phi_{n-1}^h + \frac{1}{n}\sum_{i=1}^{n-2} \phi_{n-2}^h - \frac{1}{n}\phi_1^t$$

(since $\phi_i^h = -\phi_{i+1}^t$ for all $i \in [1, N-1]$ by construction)

Therefore,

$$||\sum_{i=1}^{n-1} \alpha_i\phi_i||^2 = (\frac{(n-1)^2}{n^2} + \frac{n-2}{n^2} + \frac{1}{n^2})||\phi_1^t||^2 = \frac{n-1}{n} \cdot \frac{1}{2}$$

(Since ϕ_i^ts are mutually orthogonal and

$$||\phi_{n-1}^h|| = ||\phi_{n-2}^h|| = \cdots = ||\phi_1^h|| = ||\phi_1^t|| = \frac{1}{\sqrt{2}})$$

$$\Rightarrow ||\sum_{i=1}^{n-1} \alpha_i\phi_i||^2 \to \frac{1}{2} \text{ as } n \to \infty \tag{28}$$

Likewise using symmetrical arguments we can show that:

$$||\sum_{i=n+1}^{N} \alpha_i\phi_i||^2 = \frac{N-n-1}{N-n} \cdot \frac{1}{2} \to \frac{1}{2} \text{ as } N-n \to \infty \tag{29}$$

Therefore combining (28) and (29) we get:

$$||\mathcal{P}_{\mathcal{S}(\bigcup_1^N\{\phi_i\}\setminus\{\phi_n\})}(\phi_n)||^2$$
$$= ||\sum_{i=1}^{n-1}\alpha_i\phi_i||^2 + ||\sum_{i=n+1}^{N}\alpha_i\phi_i||^2 \to 1 \text{ for} n, N-n\to\infty \ \square$$

A.11 Proof of Lemma 3

An important consequence of Assumption 2 is that the norm of the projection of every unit vector in a subspace of a finite partition of spikes onto the remaining subspace of spikes is strictly less than 1. This condition forms the basis for establishing the convergence of *windowed iterative reconstruction* to optimal reconstruction in our subsequent windowing theorem 3. We formally state the condition in Lemma 3 before presenting Windowing theorem 3.

Lemma 3. *Let $S = \{\phi_i\}_{i=1}^N$ denote the set of spikes generated by our framework, satisfying Assumption 2, i.e., $\forall n \in \{1,...,N\}$, $||\mathcal{P}_{\mathcal{S}(\bigcup_{i=1}^N\{\phi_i\}\setminus\{\phi_n\})}(\phi_n)|| \leq \beta$, where $\beta \in \mathbb{R}$ is a constant strictly less than 1. Consider a subset $V \subseteq S$ of a finite size d, $d < N$. Then, for every $v \in \mathcal{S}(V)$ with $||v|| = 1$, $\exists \beta_d < 1$, such that $||\mathcal{P}_{\mathcal{S}(S\setminus V)}(v)|| \leq \beta_d$ where β_d is a real constant that depends on β and d. Specifically, we can show that $\beta_d^2 \leq (1 + \frac{1-\beta^2}{d^2\beta^2})^{-1} < 1$.*

Proof: Let $S \supseteq V = \{\phi_{v_1}, ...\phi_{v_d}\}$ where $\phi_{v_1}, ..., \phi_{v_d}$ are d distinct spikes from S. Also, $\forall i \in \{1,...,d\}$ let us assume that: $\phi_{v_i}^{\parallel} = \mathcal{P}_{\mathcal{S}(S\setminus V)}(\phi_{v_i})$ and $\hat{\phi}_{v_i} = \phi_{v_i} - \phi_{v_i}^{\parallel}$, i.e. each ϕ_{v_i} is decomposed into two components, $\phi_{v_i}^{\parallel}$ - the component which is the projection of ϕ_{v_i} in the span of $S \setminus V$ and $\hat{\phi}_{v_i}$ - the component of ϕ_{v_i} which is the orthogonal complement of $\phi_{v_i}^{\parallel}$. By assumption we have the following hold true:

$||\phi_{v_i}^{\parallel}|| \leq \beta$ and $||\hat{\phi}_{v_i}||^2 \geq (1-\beta^2)$.

Also for any $v \in V, ||v|| = 1$ let us assume the following:

$$v = \Sigma_{i=1}^d \alpha_i\phi_{v_i} = \underbrace{\Sigma_{i=1}^d\alpha_i\phi_{v_i}^{\parallel}}_{Y} + \underbrace{\Sigma_{i=1}^d\alpha_i\hat{\phi}_{v_i}}_{Z}$$
$$\text{s.t. } ||v||^2 = ||\Sigma_{i=1}^d\alpha_i\phi_{v_i}^{\parallel}||^2 + ||\Sigma_{i=1}^d\alpha_i\hat{\phi}_{v_i}||^2 = 1$$

Here by definition $Y = \Sigma_{i=1}^d\alpha_i\phi_{v_i}^{\parallel} = \Sigma_{i=1}^d\alpha_i\mathcal{P}_{\mathcal{S}(S\setminus V)}(\phi_{v_i}) = \mathcal{P}_{\mathcal{S}(S\setminus V)}(v)$, i.e. Y is the projection of v in the span of $S \setminus V$ and $Z = \Sigma_{i=1}^d\alpha_i\hat{\phi}_{v_i}$ is the orthogonal complement with respect $S \setminus V$. And hence the objective of the Lemma is to establish an upper bound on $||Y||$. Assume that $|\alpha_m| = max(|\alpha_1|, ..., |\alpha_d|)$, for some $m \in \{1,...,d\}$.

$$||Y||^2 = ||\Sigma_{i=1}^d\alpha_i\mathcal{P}_{\mathcal{S}(S\setminus V)}(\phi_{v_i})||^2$$
$$\leq (\Sigma_{i=1}^d|\alpha_i|||\mathcal{P}_{\mathcal{S}(S\setminus V)}(\phi_{v_i})||)^2 \quad \text{(Using Triangle inequality)}$$
$$\leq \beta^2(\Sigma_{i=1}^d|\alpha_i|)^2 \leq d^2\beta^2\alpha_m^2 \tag{30}$$
$$\text{(Since } \forall n, ||\mathcal{P}_{\mathcal{S}(\bigcup_1^N\{\phi_i\}\setminus\{\phi_n\})}(\phi_n)|| \leq \beta)$$

Again, $||Z||^2 = ||\alpha_m\hat{\phi}_{v_m} + \Sigma_{i\in\{1,...,d\}\setminus\{m\}}\alpha_i\hat{\phi}_{v_i}||^2$ (31)

But, $\hat{\phi}_{v_m}$ can be decomposed into two components: $\hat{\phi}_{v_m}^{\parallel} = \mathcal{P}_{\mathcal{S}(\cup_i\{\hat{\phi}_{v_i}\}\setminus\{\hat{\phi}_{v_m}\})}(\hat{\phi}_{v_m})$ and$\hat{\phi}_{v_m}^{\perp} = \hat{\phi}_{v_m} - \hat{\phi}_{v_m}^{\parallel}$, the orthogonal complement of $\hat{\phi}_{v_m}^{\parallel}$. But $\hat{\phi}_{v_m}^{\perp}$ can be written as :

$$\hat{\phi}_{v_m}^{\perp} = \hat{\phi}_{v_m} - \hat{\phi}_{v_m}^{\parallel} = \phi_{v_m} - \phi_{v_m}^{\parallel} - \hat{\phi}_{v_m}^{\parallel}$$
$$= \phi_{v_m} - \mathcal{P}_{\mathcal{S}(S\setminus V)}(\phi_{v_m}) - \mathcal{P}_{\mathcal{S}(\cup_i\{\hat{\phi}_{v_i}\}\setminus\{\hat{\phi}_{v_m}\})}(\hat{\phi}_{v_m})$$
$$= \phi_{v_m} - \mathcal{P}_{\mathcal{S}((S\setminus V)\cup(\cup_i\{\hat{\phi}_{v_i}\}\setminus\{\hat{\phi}_{v_m}\}))}(\phi_{v_m})$$
$$\text{(since every } \hat{\phi}_{v_i} \text{ is orthogonal to } S\setminus V)$$

But $\mathcal{S}((S \setminus V) \cup (\cup_i \{\hat{\phi}_{v_i}\} \setminus \{\hat{\phi}_{v_m}\})) \subseteq \mathcal{S}(S \setminus \{\phi_{v_m}\})$ since every $\hat{\phi}_{v_i} = \phi_{v_i} - \mathcal{P}_{\mathcal{S}(S\setminus V)}(\phi_{v_i})$, except for $\hat{\phi}_{v_m}$, can be written as a linear combination of spikes in $S \setminus \{\phi_{v_m}\}$. Hence, $||\hat{\phi}^{\perp}_{v_m}||^2 = 1 - ||\mathcal{P}_{\mathcal{S}((S\setminus V)\cup(\cup_i\{\hat{\phi}_{v_i}\}\setminus\{\hat{\phi}_{v_m}\}))}(\phi_{v_m})||^2 \geq 1 - ||\mathcal{P}_{\mathcal{S}(S\setminus\{\phi_{v_m}\})}(\phi_{v_m})||^2 \geq 1 - \beta^2$ (by assumption)

(32)

Combining (31) and (32) we get:

$$||Z||^2 = ||\alpha_m \hat{\phi}_{v_m} + \Sigma_{i\in\{1,\ldots,d\}\setminus\{m\}} \alpha_i \hat{\phi}_{v_i}||^2$$
$$\geq \alpha_m^2 ||\hat{\phi}^{\perp}_{v_m}||^2 \geq \alpha_m^2 (1 - \beta^2) \tag{33}$$

From (30) and (33) we get: $\frac{||Y||^2}{d^2\beta^2} \leq \alpha_m^2 \leq \frac{||Z||^2}{1-\beta^2}$. But,

$$1 = ||Y||^2 + ||Z||^2 \leq ||Z||^2 + d^2\beta^2\alpha_m^2 \leq \frac{d^2\beta^2}{1-\beta^2}||Z||^2 +$$

$$||Z||^2 \Rightarrow ||Z||^2 \geq \frac{1-\beta^2}{1+(d^2-1)\beta^2} \Rightarrow \beta_d^2 = ||\mathcal{P}_{\mathcal{S}(S\setminus V)}(v)||^2$$

$$= 1 - ||Z||^2 \leq 1 - \frac{1-\beta^2}{1+(d^2-1)\beta^2} \Rightarrow \beta_d^2 \leq (1 + \frac{1-\beta^2}{d^2\beta^2})^{-1}$$

Here β_d is a quantity strictly less than 1 when $\beta < 1$ and d is a finite positive integer. □

A.12 Complete Proof of Windowing Theorem

The following section presents a detailed proof of the *Windowing Theorem* as discussed in the main text.

Theorem 3 (Windowing Theorem). For an input signal X with bounded L_2 norm, suppose our framework produces a set of $n+1$ successive spikes $S = \{\phi_1, ..., \phi_{n+1}\}$, sorted by their time of occurrence and satisfying Assumption 2. The error in the iterative reconstruction of X with respect to the last spike ϕ_{n+1} due to windowing, as formulated in Eq. 4, is bounded. Specifically,

$$\forall \epsilon > 0, \exists w_0 > 0 \text{ s.t. } ||\mathcal{P}_{\phi^{\perp}_{n+1,w}}(X) - \mathcal{P}_{\phi^{\perp}_{n+1}}(X)|| < \epsilon,$$
$$\forall w \geq w_0 \text{ and } w \leq n \tag{34}$$

where w_0 is independent of n for arbitrarily large $n \in \mathbb{N}$.

Import and Proof Idea: The Theorem implies that the error from windowing can be made arbitrarily small by choosing a sufficiently large window size w, independent of n, when n is arbitrarily large. At first glance, one might think that the condition in Equation 34 is trivially satisfied by choosing $w = n$, i.e., a window inclusive of all spikes. However, the key aspect of the theorem is that w should be independent of n when n is arbitrarily large. This allows us to use the same window size regardless of the number of previous spikes, even for large signals producing many spikes, thereby maintaining the condition number of the overall solution as per Theorem 4. Our proof demonstrates this by showing that the reconstruction error converges geometrically as a function of the window size, depending only on the spike rate which in turn depends on the *ahp* parameters in Equation 1, but not on n. The proof hinges on a central lemma showing that the L_2 norm of the difference between $\phi^{\perp}_{n+1,w}$ and $\phi^{\perp}_{n+1}$ decreases steadily as w increases, based on the assumptions stated in the theorem. This ensures that $\phi^{\perp}_{n+1,w}$ becomes a good approximation of $\phi^{\perp}_{n+1}$. The proof of the theorem then follows by establishing a bound on $\|\mathcal{P}_{\phi^{\perp}_{n+1,w}}(X) - \mathcal{P}_{\phi^{\perp}_{n+1}}(X)\|$ for a given choice of w for any bounded input X. The lemma is provided below:

Lemma 4. *Under the conditions of Theorem 3, for any $\delta > 0$, there exists $w_0 \in \mathbb{N}^+$ such that $\|\phi^{\perp}_{n+1,w} - \phi^{\perp}_{n+1}\| < \delta \ \forall w \geq w_0, w \leq n$, where the choice of w_0 is independent of n for arbitrarily large $n \in \mathbb{N}^+$.*

Proof Idea: The proof leverages Corollary 0.2, which states that the maximum number of spikes overlapping in time is bounded by a constant $d \in \mathbb{N}^+$, dependent on the *ahp* parameters. Using this corollary, we partition the set of all spikes in time into a chain of subsets, where each subset overlaps only with its neighboring subsets and is disjoint from all others. Each subset contains at most d spikes. The proof then demonstrates that the error in approximating $\phi^{\perp}_{n+1}$ due to windowing, i.e., $\|\phi^{\perp}_{n+1,w} - \phi^{\perp}_{n+1}\|$, decreases faster than a geometric sequence as more of these partitions are included within the window. This convergence is illustrated schematically in Figure 3.

Proof of Lemma 4: Let the set of spikes $S = \{\phi_1, ..., \phi_n\}$ be partitioned into a chain of subsets of spikes $v_1, ..., v_m$ in descending order of time, defined recursively. The first subset v_1 consists of all previous spikes overlapping with the support of ϕ_{n+1}. Recursively, v_{i+1} is the set of spikes overlapping with the support of any spike in v_i for all $i \leq m$. This process continues until the first spike ϕ_1 is included in the final subset v_m, where $m \leq n$. An example of this partitioning is illustrated in Figure 3. The individual spikes in each partition are indexed as follows:

$$v_i = \{\phi_{p_i}, ..., \phi_{p_{i-1}-1}\}, \forall i \leq m$$
$$\text{where } 1 = p_m < ... < p_1 < p_0 = n + 1.$$

That is, the i^{th} partition v_i consists of spikes indexed from p_i to $p_{i-1} - 1$. Since the spikes $\phi_1, ..., \phi_n$ are sorted in order of their occurrence time, if both ϕ_{p_i} and $\phi_{p_{i-1}-1}$ are in partition v_i, then by construction, $\phi_j \in v_i$ for all $p_i \leq j \leq p_{i-1} - 1$.

Claim 4.1. *The number of spikes in every partition, v_i $\forall i \in \{1, ..., m\}$, is bounded by some constant $d \in \mathbb{N}^+$.*

Proof: This corollary follows from the observation that all spikes in a given partition v_{i+1} overlap in time with at least one spike from the preceding partition v_i, specifically the spike whose support extends furthest into the past, for all $i \in \{2, .., m\}$ (see fig 3). In the case of the partition v_1, each spike overlaps with ϕ_{n+1}. Then the proof follows from the corollary 0.2. □

Now, let $V_1, ..., V_m$ be subspaces spanned by the subsets of spikes $v_1, ..., v_m$ respectively. Before proceeding with the rest of the proof, we introduce the following notations.

$$\tilde{\phi}_i = \phi_i - \mathcal{P}_{\mathcal{S}(\bigcup_{j=i+1}^{n} \{\phi_j\})}(\phi_i) \quad \forall i \in \{1, ..., n\}$$

where $\tilde{\phi}_i$ denotes the orthogonal complement of the spike ϕ_i with respect to all the future spikes up to ϕ_n. We also denote,

$$\tilde{V_k} = \mathcal{S}(\bigcup_{j=p_k}^{p_{k-1}-1} \{\tilde{\phi}_j\}) \quad \forall k \in \{1, ..., m\}$$

where $\tilde{V_k}$ is the subspace spanned by spikes in the partition v_k, with each spike orthogonalized with respect to all future spikes.

Claim 4.2. *For the subspaces defined on the partitions as above, the following holds:*

$$\tilde{V_k} = \{x - \mathcal{P}_{\sum_{j=1}^{k-1} V_j}(x) | x \in V_k\} = \mathcal{S}(\bigcup_{j=p_k}^{p_{k-1}-1} \{\tilde{\phi}_j\}).$$

Proof: Denote $A = \{x - \mathcal{P}_{\sum_{j=1}^{k-1} V_j}(x) | x \in V_k\}$ and $B = \mathcal{S}(\bigcup_{j=p_k}^{p_{k-1}-1} \{\tilde{\phi}_j\})$. We need to show that $A = B$. Before showing that, we rewrite $\tilde{\phi}_i$ for $p_k \leq i \leq p_{k-1} - 1$ as follows:

$$\begin{aligned}
\tilde{\phi}_i &= \phi_i - \mathcal{P}_{\mathcal{S}(\bigcup_{j=i+1}^{n} \{\phi_j\})}(\phi_i) \quad (p_k \leq i \leq p_{k-1} - 1) \\
&= \phi_i - \mathcal{P}_{\mathcal{S}((\bigcup_{j=p_{k-1}}^{n} \{\phi_j\}) \bigcup (\bigcup_{j=i+1}^{p_{k-1}-1} \{\tilde{\phi}_j\}))}(\phi_i) \\
&= \phi_i - \mathcal{P}_{\sum_{j=1}^{k-1} V_j}(\phi_i) - \sum_{j=i+1}^{p_{k-1}-1} \mathcal{P}_{\tilde{\phi}_j}(\phi_i) \qquad (35) \\
&(\text{As } \sum_{j=1}^{k-1} V_j \text{ and all } \tilde{\phi}_i\text{'s are mutually orthogonal}
\end{aligned}$$

for $p_k \leq i \leq p_{k-1} - 1$)

Using the formulation in 35, we will show $A = B$. First, we show that $A \subseteq B$. Let there be any $y \in A$. Then by assumption, we can write y as follows:

$$y = x - \mathcal{P}_{\sum_{j=1}^{k-1} V_j}(x), \text{for some } x = \sum_{i=p_k}^{p_{k-1}-1} \alpha_i \phi_i, \quad \alpha_i \in \mathbb{R}$$

$$\Rightarrow y = \sum_{i=p_k}^{p_{k-1}-1} \alpha_i \phi_i - \mathcal{P}_{\sum_{j=1}^{k-1} V_j}(\sum_{i=p_k}^{p_{k-1}-1} \alpha_i \phi_i)$$

$$= \sum_{i=p_k}^{p_{k-1}-1} \alpha_i (\phi_i - \mathcal{P}_{\sum_{j=1}^{k-1} V_j}(\phi_i)) \text{ (by linearity of projection)}$$

$$\Rightarrow y = \sum_{i=p_k}^{p_{k-1}-1} \alpha_i (\tilde{\phi}_i + \sum_{j=i+1}^{p_{k-1}-1} \mathcal{P}_{\tilde{\phi_j}}(\phi_i)) \text{ (using (35))} \tag{36}$$

Equation (36) shows that y is written as a linear combination of $\tilde{\phi}_i$ for $p_k \leq i \leq p_{k-1} - 1$, and hence $y \in B$. This proves that $A \subseteq B$.

To prove $B \subseteq A$, we observe that any $z \in B$ can be written as a linear combination of $\tilde{\phi}_i$ for $p_k \leq i \leq p_{k-1} - 1$. So it suffices to show that $\tilde{\phi}_i \in A$ for $p_k \leq i \leq p_{k-1} - 1$. We do that via induction on the set $\{\tilde{\phi}_{p_{k-1}-1}, ..., \tilde{\phi}_{p_k}\}$, starting with $\tilde{\phi}_{p_{k-1}-1}$ as the base case. The base case is trivially true, i.e., $\tilde{\phi}_{p_{k-1}-1} \in A$, because by equation 35, we immediately obtain $\tilde{\phi}_{p_{k-1}-1} = \phi_{p_{k-1}-1} - \mathcal{P}_{\sum_{j=1}^{k-1} V_j}(\phi_{p_{k-1}-1})$.

For the induction step, assume that $\tilde{\phi}_i \in A$ for all $i$such that $i \in [n, p_{k-1} - 1]$ for some$n \in \mathbb{N}^+$ and $p_k < n \leq p_{k-1} - 1$. We need to show that $\tilde{\phi}_{n-1} \in A$. For that, we again use equation (35) to observe the following:

$$\tilde{\phi}_{n-1} = (\phi_{n-1} - \mathcal{P}_{\sum_{j=1}^{k-1} V_j}(\phi_{n-1})) - \sum_{j=n}^{p_{k-1}-1} \mathcal{P}_{\tilde{\phi_j}}(\phi_{n-1}) \tag{37}$$

Analyzing the expression for $\tilde{\phi}_{n-1}$ on the right-hand side of equation (37), we observe that the expression $(\phi_{n-1} - \mathcal{P}_{\sum_{j=1}^{k-1} V_j}(\phi_{n-1})) \in A$ by construction, and the expression $\sum_{j=n}^{p_{k-1}-1} \mathcal{P}_{\tilde{\phi_j}}(\phi_{n-1})$, a linear combination of $\tilde{\phi}_i$, $n \leq i \leq p_{k-1} - 1$, is also in A because by induction hypothesis, each $\tilde{\phi}_i$, for $n \leq i \leq p_{k-1} - 1$ is in A. Thus, we obtain $\tilde{\phi}_{n-1} \in A$. This proves that $B \subseteq A$. Therefore, $A = B$. □

Following the claim, we define the subspace U_k as below:

$$U_k = \sum_{i=k}^{m} \tilde{V}_i = \mathcal{S}(\bigcup_{j=1}^{p_{k-1}-1} \{\tilde{\phi}_j\}) \tag{38}$$

$$= \{x - \mathcal{P}_{\sum_{j=1}^{k-1} V_j}(x) | x \in \sum_{i=k}^{m} V_i\} \tag{39}$$

where the equality between 38 and 39 follows from the Claim 4.2. Lastly, define $\phi^{\perp}_{n+1,v_k}$ as:

$$\phi^{\perp}_{n+1,v_k} = \phi_{n+1} - \mathcal{P}_{\sum_{i=1}^{k} V_i}(\phi_{n+1})$$

i.e. $\phi^{\perp}_{n+1,v_k}$ is the orthogonal complement of ϕ_{n+1} with respect to window of spikes up to partition v_k. Now we proceed to quantify the norm of the difference of $\phi^{\perp}_{n+1}$ and $\phi^{\perp}_{n+1,v_k}$. For that denote e_k as follows:

$$e_k = ||\phi^{\perp}_{n+1,v_k} - \phi^{\perp}_{n+1}||$$
$$= ||\mathcal{P}_{\sum_{i=1}^{k} V_i}(\phi_{n+1}) - \mathcal{P}_{\sum_{i=1}^{m} V_i}(\phi_{n+1})||$$

$= ||\mathcal{P}_{\sum_{i=1}^{k} V_i}(\phi_{n+1}) - \mathcal{P}_{\sum_{i=1}^{k} V_i + U_{k+1}}(\phi_{n+1})||$ (by def. of U_k)

$= ||\mathcal{P}_{\sum_{i=1}^{k} V_i}(\phi_{n+1}) - (\mathcal{P}_{\sum_{i=1}^{k} V_i}(\phi_{n+1}) + \mathcal{P}_{U_{k+1}}(\phi_{n+1}))||$

(Since by construction $U_{k+1} \perp \Sigma_{i=1}^{k} V_i$)

$\Rightarrow e_k = ||\mathcal{P}_{U_{k+1}}(\phi_{n+1}))||$ (40)

Note that by definition $e_m = 0$ and by Assumption 2 we get,

$e_k = ||\mathcal{P}_{U_{k+1}}(\phi_{n+1}))|| \leq |\beta| < 1, \ \forall k \in \{1, ..., m\}$ (41)

Now Assume that:

$\mathcal{P}_{U_{k+1}}(\phi_{n+1})) = \alpha_k \tilde{\psi}_k$, where $\alpha_k \in \mathcal{R}, \tilde{\psi}_k \in U_{k+1}$

Then it follows from the definition of U_{k+1} and Claim 4.2

that $\tilde{\psi}_k$ is of the form:

$$\tilde{\psi}_k = \psi_k - \mathcal{P}_{\sum_{j=1}^{k} V_i}(\psi_k) \text{ for some } \psi_k \in \sum_{i=k+1}^{m} V_i$$

W.L.O.G. assume $||\psi_k|| = 1$ by appropriately choosing α_k.

Since $\alpha_k \tilde{\psi}_k$ is a projection of ϕ_{n+1} we obtain:

$$\alpha_k = \frac{\langle \phi_{n+1}, \tilde{\psi}_k \rangle}{||\tilde{\psi}_k||^2} \tag{42}$$

For $k > 0$ from 42 and 40 we obtain,

$$e_k = ||\alpha_k \tilde{\psi}_k|| = |\alpha_k| ||\tilde{\psi}_k|| = \frac{|\langle \phi_{n+1}, \tilde{\psi}_k \rangle|}{||\tilde{\psi}_k||}$$

$$= \frac{|\langle \phi_{n+1}, \psi_k - \mathcal{P}_{\sum_{j=1}^{k} V_j}(\psi_k) \rangle|}{||\tilde{\psi}_k||}$$

$$= \frac{|\langle \phi_{n+1}, \mathcal{P}_{\sum_{j=1}^{k} V_j}(\psi_k) \rangle|}{||\tilde{\psi}_k||} = \frac{|\langle \phi_{n+1}, \mathcal{P}_{\sum_{j=1}^{k-1} V_j + \tilde{V}_k}(\psi_k) \rangle|}{||\tilde{\psi}_k||}$$

(for $k > 0, \psi_k \in \sum_{i=k+1}^{m} V_i \perp \phi_{n+1}$ due to disjoint support)

$$= \frac{|\langle \phi_{n+1}, \mathcal{P}_{\sum_{j=1}^{k-1} V_j}(\psi_k) + \mathcal{P}_{\tilde{V}_k}(\psi_k) \rangle|}{||\tilde{\psi}_k||} \text{ (by def. } \tilde{V}_k \perp \sum_{j=1}^{k-1} V_j)$$

$$\Rightarrow e_k = \frac{|\langle \phi_{n+1}, \mathcal{P}_{\tilde{V}_k}(\psi_k) \rangle|}{||\tilde{\psi}_k||} \ (k > 0, \psi_k \in \sum_{i=k+1}^{m} V_i \perp \sum_{j=1}^{k-1} V_j) \tag{43}$$

Likewise, $e_{k-1} = ||\mathcal{P}_{U_k}(\phi_{n+1}))|| = ||\mathcal{P}_{U_{k+1} + \tilde{V}_k}(\phi_{n+1}))||$

$\Rightarrow e_{k-1}^2 = ||\mathcal{P}_{U_{k+1}}(\phi_{n+1}))||^2 + ||\mathcal{P}_{\tilde{V}_k}(\phi_{n+1}))||^2$

(Since by construction $U_{k+1} \perp \tilde{V}_k$)

$\Rightarrow e_{k-1}^2 = e_k^2 + ||\mathcal{P}_{\tilde{V}_k}(\phi_{n+1}))||^2$ (44)

Assume that $\mathcal{P}_{\tilde{V}_k}(\phi_{n+1}) = \beta_k \tilde{\theta}_k$, where $\tilde{\theta}_k \in \tilde{V}_k, \beta_k \in \mathcal{R}$

$$\Rightarrow ||\mathcal{P}_{\tilde{V}_k}(\phi_{n+1}))|| = ||\beta_k \tilde{\theta}_k|| = \frac{|\langle \phi_{n+1}, \tilde{\theta}_k \rangle|}{||\tilde{\theta}_k||} \tag{45}$$

From 44 & 45 we get, $e_{k-1}^2 = e_k^2 + \frac{|\langle \phi_{n+1}, \tilde{\theta}_k \rangle|^2}{||\tilde{\theta}_k||^2}$ (46)

Again, for $k > 0$ we further analyze e_k from 43 to obtain:

$$e_k = \frac{|\langle \phi_{n+1}, \mathcal{P}_{\tilde{V}_k}(\psi_k)\rangle|}{||\tilde{\psi}_k||}$$
$$= \frac{|\langle \phi_{n+1}, \mathcal{P}_{\mathcal{S}(\{\tilde{\theta}_k\})}(\psi_k) + \mathcal{P}_{\tilde{V}_k/\mathcal{S}(\{\tilde{\theta}_k\})}(\psi_k)\rangle|}{||\tilde{\psi}_k||}$$

The last line above is essentially written by breaking ψ_k into two mutually orthogonal subspaces: $\mathcal{S}(\{\tilde{\theta}_k\})$, subspace of $\tilde{V_k}$ spanned by $\tilde{\theta}_k$, and $\tilde{V_k}/\mathcal{S}(\{\tilde{\theta}_k\}$, the subspace of $\tilde{V_k}$ orthogonal to the subspace $\mathcal{S}(\{\tilde{\theta}_k\})$. Also, observe that $\phi_{n+1} \perp \tilde{V_k}/\mathcal{S}(\{\tilde{\theta}_k\})$ since $\mathcal{P}_{\tilde{V}_k}(\phi_{n+1})) = \beta_k\tilde{\theta}_k$. Therefore,

$$e_k = \frac{|\langle \phi_{n+1}, \mathcal{P}_{\mathcal{S}(\{\tilde{\theta}_k\})}(\psi_k)|}{||\tilde{\psi}_k||} \quad = \frac{|\langle \phi_{n+1}, \frac{\langle \psi_k, \tilde{\theta}_k\rangle \tilde{\theta}_k}{||\tilde{\theta}_k||^2}\rangle|}{||\tilde{\psi}_k||}$$

$$\Rightarrow e_k = \frac{|\langle \phi_{n+1}, \tilde{\theta}_k\rangle|}{||\tilde{\theta}_k||} \frac{|\langle \psi_k, \frac{\tilde{\theta}_k}{||\tilde{\theta}_k||}\rangle|}{||\tilde{\psi}_k||} \tag{47}$$

Combining 46 and 47 we obtain:

$$e_{k-1}^2 = e_k^2 + e_k^2 \frac{||\tilde{\psi}_k||^2}{|\langle \psi_k, \frac{\tilde{\theta}_k}{||\tilde{\theta}_k||}\rangle|^2} \tag{48}$$

Since both ψ_k and $\frac{\tilde{\theta}_k}{||\tilde{\theta}_k||}$ are unit norm $|\langle \psi_k, \frac{\tilde{\theta}_k}{||\tilde{\theta}_k||}\rangle| \leq 1$.

$$\Rightarrow e_{k-1}^2 \geq e_k^2(1 + ||\tilde{\psi}_k||^2) \tag{49}$$

Eq. 49 demonstrates that the sequence $\{e_k\}$ converges geometrically to 0 when $(1 + ||\tilde{\psi}_k||^2) > 1$, i.e. $||\tilde{\psi}_k|| > 0$. To complete the proof of Lemma 4, we need to establish a positive lower bound on $||\tilde{\psi}_k||$, ensuring the convergence of the sequence e_k. The following Corollary provides the necessary lower bound on $||\tilde{\psi}_k||$.

Claim 4.3. *Following Claim 4.1, if the number of spikes in each partition v_i is bounded by d, then $||\tilde{\psi}_k||^2 \geq 1 - \beta_d^2$, where $\beta_d > 0$ is as defined in Lemma 3.*

Proof: $\tilde{\psi}_k = \psi_k - \mathcal{P}_{\tilde{V}_k}(\psi_k), ||\psi_k|| = 1.$

Let, $\mathcal{P}_{\tilde{V}_k}(\psi_k) = \beta_k\tilde{\eta}_k$ for some $\tilde{\eta}_k \in \tilde{V}_k, ||\tilde{\eta}_k|| = 1$

$\Rightarrow \beta_k = \langle \psi_k, \tilde{\eta}_k\rangle$

$$\Rightarrow \mathcal{P}_{\psi_k}(\tilde{\eta}_k) = \langle \psi_k, \tilde{\eta}_k\rangle \psi_k, \quad \text{(Since, } ||\psi_k|| = 1) \tag{50}$$

But $\tilde{\eta}_k \in \tilde{V}_k \implies \tilde{\eta}_k = \eta_k - \mathcal{P}_{\sum_{i=1}^{k-1} V_i}(\eta_k)$

for some $\eta_k \in V_k$. Also,

$$||\eta_k||^2 = 1 + ||\mathcal{P}_{\sum_{i=1}^{k-1} V_i}(\eta_k)||^2 \tag{51}$$

We observe, $\mathcal{P}_{\psi_k}(\tilde{\eta}_k) = \mathcal{P}_{\psi_k}(\eta_k - \mathcal{P}_{\sum_{i=1}^{k-1} V_i}(\eta_k))$

$$= \mathcal{P}_{\psi_k}(\eta_k) \quad \text{(Since } \psi_k \perp \sum_{i=1}^{k-1} V_i) \tag{52}$$

Now, consider the vector $\frac{\eta_k}{||\eta_k||} \in V_k$. Here $\frac{\eta_k}{||\eta_k||}$ is an unit vector in V_k, the span of a finite partition of spikes v_k, the size of which is bounded from above by some constant d according to Claim 4.1. Therefore, based on the assumption stated in the theorem 3, we can infer that the norm of any projection of $\frac{\eta_k}{||\eta_k||}$ on any subspace spanned by a set of spikes other than those in v_k, would bounded

from above by some constant $\beta_d > 1$. Hence, we can write the following.

$$
\begin{aligned}
&||\mathcal{P}_{\mathcal{S}(\{\psi_k\})+\sum_{i=1}^{k-1} V_i}(\frac{\eta_k}{||\eta_k||})||^2 \leq \beta_d^2 \quad (|\beta_d| < 1) \\
&\Rightarrow \frac{||\mathcal{P}_{\psi_k}(\eta_k)||^2 + ||\mathcal{P}_{\sum_{i=1}^{k-1} V_i}(\eta_k)||^2}{||\eta_k||^2} \leq \beta_d^2 \\
&\Rightarrow \frac{||\mathcal{P}_{\psi_k}(\eta_k)||^2 + ||\mathcal{P}_{\sum_{i=1}^{k-1} V_i}(\eta_k)||^2}{1 + ||\mathcal{P}_{\sum_{i=1}^{k-1} V_i}(\eta_k)||^2} \leq \beta_d^2 \quad \text{(using 51)} \\
&\Rightarrow ||\mathcal{P}_{\psi_k}(\eta_k)||^2 \leq \beta_d^2(1 + ||\mathcal{P}_{\psi_k}(\eta_k)||^2) - ||\mathcal{P}_{\psi_k}(\eta_k)||^2 \\
&\Rightarrow ||\mathcal{P}_{\psi_k}(\eta_k)||^2 \leq \beta_d^2 - (1-\beta_d^2)||\mathcal{P}_{\psi_k}(\eta_k)||^2 \\
&\Rightarrow ||\mathcal{P}_{\psi_k}(\eta_k)||^2 \leq \beta_d^2 \quad (\text{Since} |\beta_d| < 1) \\
&\Rightarrow ||\mathcal{P}_{\psi_k}(\tilde{\eta}_k)||^2 = ||\mathcal{P}_{\psi_k}(\eta_k)||^2 \leq \beta_d^2 \\
&\Rightarrow ||\mathcal{P}_{\tilde{\eta}_k}(\psi_k)|| = ||\mathcal{P}_{\psi_k}(\tilde{\eta}_k)|| \leq \beta_d^2 \\
&\text{(Since both } \tilde{\eta}_k \text{ and } \psi_k \text{ are unit vectors)} \\
&\text{Therefore, } ||\tilde{\psi_k}||^2 = ||\psi_k - \mathcal{P}_{\tilde{V}_k}(\psi_k)||^2 = 1 - ||\mathcal{P}_{\tilde{\eta}_k}(\psi_k)||^2 \\
&\quad \text{(Since by assumption } \mathcal{P}_{\tilde{V}_k}(\psi_k) = \beta_k \tilde{\eta}_k \text{ and } ||\psi_k|| = 1) \\
&\Rightarrow ||\tilde{\psi_k}||^2 \geq 1 - \beta_d^2 \qquad (|\beta_d| < 1) \qquad \square
\end{aligned}
$$

Finally, combining 49 with Claim 4.3, we obtain:

$$
e_{k-1}^2 \geq e_k^2(1 + (1-\beta_d^2)) \Rightarrow e_k^2 \leq \frac{e_{k-1}^2}{\gamma^2} \tag{53}
$$

where $\gamma^2 = (1 + (1-\beta_d^2))$ is a constant strictly greater than 1. Since $e_m = 0$, Eq. 53 shows that the sequence $\{e_k\}_{k=1}^m$ converges to 0 faster than geometrically. Thus, $\forall \delta > 0, \exists k_0 \in \mathbb{N}^+$ such that $e_k < \delta$, for all $k \geq k_0$ and $m \geq k$. Given the geometric drop in Eq. 53 and the bound $e_1 < 1$ (Eq. 41), for arbitrarily large m (hence n) the choice of k_0 is independent of m and depends only on β_d, which is determined by the *ahp* parameters in Eq. 1). Since the number of spikes in each partition is bounded by d (Claim 4.1), choosing a window size $w_0 = k_0 * d$ ensures $\forall \delta > 0, \exists w_0 \in \mathbb{N}^+$ such that $||\phi_{n+1,w}^{\perp} - \phi_{n+1}^{\perp}|| < \delta$ for all $w \geq w_0$ and $w \leq n$. For arbitrarily large n, the choice of w_0 is independent of n. $\square$

Proof of Theorem 3: Having established a bound on the norm of the difference between $\phi_{n+1}^{\perp}$ and $\phi_{n+1,w}^{\perp}$, we now need to bound the norm of the difference between the projections of the input signal X with respect to these two vectors. Specifically, we seek to bound $||\mathcal{P}_{\phi_{n+1,w}^{\perp}}(X) - \mathcal{P}_{\phi_{n+1}^{\perp}}(X)||$ based on the window size. We use the following notations:

$$
\begin{aligned}
\tilde{X} &= \mathcal{P}_{\mathcal{S}(\{\phi_{n+1,w}^{\perp}, \phi_{n+1}^{\perp}\})}(X) \\
X_u &= \mathcal{P}_{\phi_{n+1}^{\perp}}(X) \\
X_v &= \mathcal{P}_{\phi_{n+1,w}^{\perp}}(X)
\end{aligned}
$$

So that $\tilde{X}$, X_u and X_v lie in the same plane (see Fig. 4). Assume the angle between $\tilde{X}$ and $\phi_{n+1}^{\perp}\})$ is a, and the angle between $\tilde{X}$ and $\phi_{n+1,w}^{\perp}\})$ is b. Hence, the angle between $\phi_{n+1,w}^{\perp}$ and $\phi_{n+1}^{\perp}$ is $a - b$ (see Fig. 4). Note that the input X may not necessarily lie in the same plane as X_u and X_v. Hence, for our calculation, we consider the projection $\tilde{X}$ of X onto this plane. We also denote:

$$
\begin{aligned}
p_w &= \phi_{n+1,w}^{\perp} - \phi_{n+1}^{\perp} \\
&= (\phi_{n+1} - \mathcal{P}_{\mathcal{S}(\{\phi_{n-w+1},\ldots,\phi_n\})}(\phi_{n+1})) \\
&\qquad - (\phi_{n+1} - \mathcal{P}_{\mathcal{S}(\{\phi_1,\ldots,\phi_n\})}(\phi_{n+1})) \\
&\text{(by definition) and thus:} \\
p_w &= \mathcal{P}_{\mathcal{S}(\{\phi_{n-w+1},\ldots,\phi_n\})}(\phi_{n+1})
\end{aligned} \tag{54}
$$

$$
\begin{aligned}
&\quad - \mathcal{P}_{\mathcal{S}(\{\tilde{\phi}_1,\ldots,\tilde{\phi}_w\}\cup\{\phi_{n-w+1},\ldots,\phi_n\})}(\phi_{n+1}) \\
&= \mathcal{P}_{\mathcal{S}(\{\tilde{\phi}_1,\ldots,\tilde{\phi}_w\})}(\phi_{n+1}) \\
&\qquad\qquad \text{(Since } \mathcal{S}(\{\tilde{\phi}_1,\ldots,\tilde{\phi}_w\}) \perp \mathcal{S}(\{\phi_{n-w+1},\ldots,\phi_n\}) \\
&= \mathcal{P}_{\mathcal{S}(\{\tilde{\phi}_1,\ldots,\tilde{\phi}_w\})}(\phi_{n+1} - \mathcal{P}_{\mathcal{S}(\{\phi_{n-w+1},\ldots,\phi_n\})}(\phi_{n+1})) \\
&= \mathcal{P}_{\mathcal{S}(\{\tilde{\phi}_1,\ldots,\tilde{\phi}_w\})}(\phi^{\perp}_{n+1,w}) \\
&\Rightarrow p_w = \phi^{\perp}_{n+1,w} - \phi^{\perp}_{n+1} = \mathcal{P}_{\mathcal{S}(\{\tilde{\phi}_1,\ldots,\tilde{\phi}_w\})}(\phi^{\perp}_{n+1,w}) \\
&\Rightarrow p_w \perp \phi^{\perp}_{n+1} \\
&\text{(Since } \mathcal{P}_S(x) \perp (x - \mathcal{P}_S(x)) \text{ for any } x \in \mathcal{H}, S \subseteq \mathcal{H}) \\
&\text{Therefore, from Figure 4, we observe:} \\
&\sin^2(a-b) = \frac{||p_w||^2}{||\phi^{\perp}_{n+1,w}||^2} \qquad (55)
\end{aligned}
$$

Now, we can quantify the norm of the difference in two the projections of X with respect to $\phi^{\perp}_{n+1}$ and $\phi^{\perp}_{n+1,w}$ as follows:

$$
\begin{aligned}
&||\mathcal{P}_{\phi^{\perp}_{n+1,w}}(X) - \mathcal{P}_{\phi^{\perp}_{n+1}}(X)|| = ||X_u - X_v||^2 \\
&= ||X_u||^2 + ||X_v||^2 - 2||X_u||||X_v||cos(a-b) \\
&= ||\tilde{X}||^2[\cos^2 a + \cos^2 b - 2\cos a \cos b \cos(a-b)] \\
&\qquad \text{(using the geometry of Figure 4)} \\
&= ||\tilde{X}||^2[\frac{1+\cos 2a}{2} + \frac{1+\cos 2b}{2} \\
&\qquad\qquad - (\cos(a-b) + \cos(a+b))\cos(a-b)] \\
&= ||\tilde{X}||^2[\frac{1}{2} - \frac{\cos 2(a-b)}{2}] = ||\tilde{X}||^2 \sin^2(a-b) \\
&= ||\tilde{X}||^2 \frac{||p_w||^2}{||\phi^{\perp}_{n+1,w}||^2} \quad \text{(from Eq. (55))} \\
&\leq \frac{||X||^2}{1-\beta^2}||\phi^{\perp}_{n+1,w} - \phi^{\perp}_{n+1}||^2 \\
&\text{(Since } ||\tilde{X}|| \leq ||X|| \text{ and } ||\phi^{\perp}_{n+1,w}|| \geq ||\phi^{\perp}_{n+1}|| \geq (1-\beta^2) \\
&\text{by assumption)}
\end{aligned}
$$

Finally, since the input signal has a bounded L_2 norm (i.e. $||X||$ is bounded), setting $\delta = \epsilon\frac{\sqrt{1-\beta^2}}{||X||}$ in Lemma 4 gives a window size w_0 such that $||\mathcal{P}_{\phi^{\perp}_{n+1,w}}(X) - \mathcal{P}_{\phi^{\perp}_{n+1}}(X)|| \leq \frac{||X||}{\sqrt{1-\beta^2}}||\phi^{\perp}_{n+1,w} - \phi^{\perp}_{n+1}|| < \epsilon, \forall \epsilon > 0, \forall w \geq w_0$ and $w \leq n$. The choice of w_0 is independent of n for arbitrarily large $n \in \mathbb{N}$. This completes the proof of the theorem. □

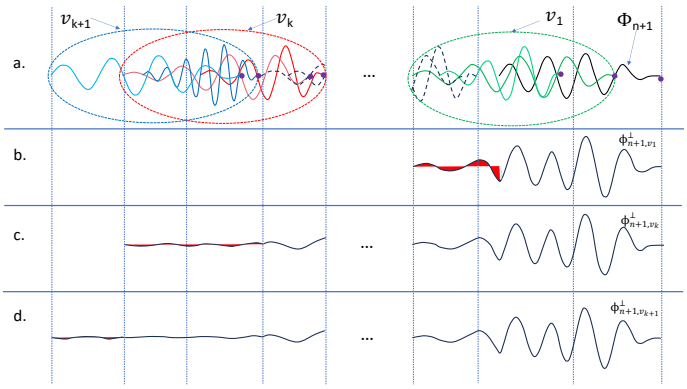

Figure 3: Illustration of Lemma 4. The diagram shows the convergence of the windowed orthogonal complement $\phi^{\perp}_{n+1,w}$ of spike ϕ_{n+1} to $\phi^{\perp}_{n+1}$ by orthogonalizing ϕ_{n+1} across the partitions of spikes. (a) Displays all spikes up to ϕ_{n+1} (black), with partitions circled: v_1 (green), v_k (red), and v_{k+1} (blue). Spikes contained in each partition are shaded accordingly, with the time of each spike marked by a purple dot. (b), (c), and (d) show orthogonal complements $\phi^{\perp}_{n+1,v_1}$, $\phi^{\perp}_{n+1,v_k}$, and $\phi^{\perp}_{n+1,v_{k+1}}$ respectively. The support of $\phi^{\perp}_{n+1,w}$ extends as more partitions are included, with the extending tail for each additional partition highlighted in red. This tail's diminishing energy as more partitions are added illustrates Lemma 4.

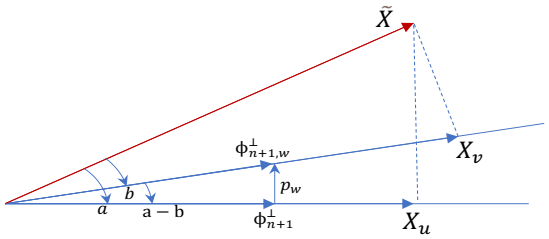

Figure 4: Figure illustrating the vector projections of the input signal X onto vectors $\phi^{\perp}_{n+1}$ and $\phi^{\perp}_{n+1,w}$. The red vector represents $\tilde{X}$, the projection of X within the plane formed by $\phi^{\perp}_{n+1}$ and $\phi^{\perp}_{n+1,w}$. The vectors $\phi^{\perp}_{n+1}$ and $\phi^{\perp}_{n+1,w}$, as well as the projections X_u and X_v of $\tilde{X}$ onto them, are indicated in blue. The vector p_w, representing the difference between $\phi^{\perp}_{n+1,w}$ and $\phi^{\perp}_{n+1}$, is also shown in blue. The angles a between $\tilde{X}$ and X_u, b between $\tilde{X}$ and X_v, and $a - b$ between $\phi^{\perp}_{n+1,w}$ and $\phi^{\perp}_{n+1}$ are marked.

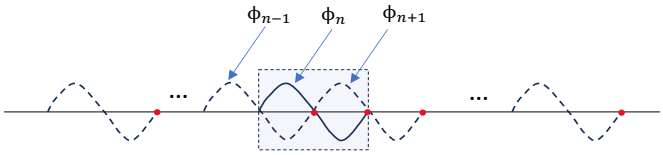

Figure 5: The scenario illustrating the need for Assumption 2. See text for details. For derivation of $||\mathcal{P}_{\mathcal{S}(\bigcup_1^N\{\phi_i\}\setminus\{\phi_n\})}(\phi_n)|| \to 1$ see appendix A.10.