SUPPLEMENTARY: ROBUST ONLINE RECONSTRUCTION OF CONTINUOUS TIME SIGNALS FROM A LEAN SPIKE TRAIN ENSEMBLE CODE

Anik Chattopadhyay, Arunava Banerjee, *Member, IEEE,*

I. 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, 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 $\frac{\delta}{2}$.

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:

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

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-

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

(Using Cauchy-Schwarz inequality and $||\Phi^j||_2 = 1$)

(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 1, $|C^j(t_k) - C^j(t_0)| = |\sum_{i=1}^k \Delta_i| < b\sqrt{\tau}, \forall k \in \{1, ..., L\}$. 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 Eq. 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

A preliminary version has appeared in [1].

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,

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

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

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

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,

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

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. □

II. Detailed Proof of the Perfect Reconstruction Theorem

The following section provides a detailed proof of the Perfect Reconstruction Theorem as presented in the main paper. While the main paper 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 paper but not explicitly stated there. These additions provide a more comprehensive understanding and are included here to complement the content of the main paper.

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 paper. 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 paper, 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. Then for appropriately chosen time-varying thresholds of the kernels, the reconstruction, X^*, resulting from the proposed coding-decoding framework 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[2] 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 paper, 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 paper. 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 $X_1 = \sum_{i=1}^{M}a_ie_i$ and $X_2 = \sum_{i=1}^{M}b_ie_i$, 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:

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

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

$= T^{j_i}(t_i) - T^{j_i}(t_i) = 0$ (by Eq. (2) & (3) of main paper)

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

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

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: $X = \sum_{p=1}^N \alpha_p \Phi^{j_p}(t_p - t)$ $(\alpha_p \in \mathbb{R}, t_p \in \mathbb{R}^+, N \in Z^+)$. 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: $||X - X^*||_2 \leq ||X - X||_2 = 0$. □

III. 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:

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

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

$$X_{hyp} = \sum_{i=1}^N a_i \Phi^{j_i}(t_i' - t), a_i \in R$$
$$\text{s.t. } \langle X_{hyp}, \Phi^{j_i}(t_i' - t) \rangle = T_i, \forall i \tag{3}$$

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:

$$||X - X_{hyp}|| \geq ||X - X^*|| \tag{4}$$

$||X - X_{hyp}||_2^2 = \langle X - X_{hyp}, X - X_{hyp} \rangle$

$= \langle X - X_{hyp}, X \rangle - \langle X - X_{hyp}, X_{hyp} \rangle$

$= \langle X - X_{hyp}, X \rangle - \Sigma_{i=1}^N a_i \langle X - X_{hyp}, \Phi^{j_i}(t - t_i') \rangle$

$= ||X||_2^2 - \langle X, X_{hyp} \rangle$

(Since by construction $\langle X_{hyp}, \Phi^{j_i}(t - t_i') \rangle = T_i \ \forall i \in \{1...N\}$)

$= \Sigma_{i=1}^N \Sigma_{k=1}^N \alpha_i \alpha_k \langle f_i(t - t_i), f_k(t - t_k) \rangle$
$\quad - \Sigma_{i=1}^N \Sigma_{k=1}^N \alpha_i a_k \langle f_i(t - t_i), \Phi^{j_k}(t - t_k')) \rangle$

$$= \alpha^T F \alpha - \alpha^T F_K a \tag{5}$$

(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 Eq. (5) we get,

$$||X - X_{hyp}||_2^2 = \alpha^T F\alpha - \alpha^T F_K P^{-1} F_K^T\alpha \quad (6)$$

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

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

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

$$= (P)_{ik} - (\mathcal{E}_K)_{ik} \quad (7)$$

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

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

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

$$= (\mathcal{E})_{ik} - (\mathcal{E}_K)_{ik} - (\mathcal{E}_K)_{ki} + (P)_{ik} \quad (8)$$

Combining Eq. (6), (7) and (8) 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, $\alpha^T\mathcal{E}_K P^{-1}\mathcal{E}_K^T\alpha > 0$) (9)

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.)

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

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

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

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

$$\Rightarrow (\mathcal{E})_{ik} \leq x_{ik}.(\delta + C\gamma)^2 \quad (10)$$

(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 Eq. (10)) (11)

(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:

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

$\quad \geq 1 - \eta$ (12)

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

Combining the results from Eq. (9), (11) and (12) we get:

$||X(t)-X_{hyp}(t)||^2/||X(t)||^2 \leq \alpha^T\mathcal{E}\alpha/\alpha^T F\alpha$

$\leq \Lambda_{max}(\mathcal{E})/\Lambda_{min}(F)$

$$\leq (\delta + C\gamma)^2(x_{max}+1)/(1-\eta) \quad (13)$$

Finally using Eq. (4) we conclude,

$||X(t)-X^*(t)||^2/||X(t)||^2 \leq ||X(t)-X_{hyp}(t)||^2/||X(t)||^2$

$\quad \leq (\delta + C\gamma)^2(x_{max}+1)/(1-\eta)$

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

IV. Analysis of Overlapping Kernels in Figure 2

This section contains an analysis of the overlapping sine kernels scenario depicted in Figure 2. The assumption 1 from the main paper 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 2 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 2, where each spike is generated by a normalized sine wave kernel. In this setup, the projection of any spike ϕ_n onto the span of all other spikes in the sequence satisfies $||\mathcal{P}_{\mathcal{S}(\bigcup_1^N\{\phi_i\}\setminus\{\phi_n\})}(\phi_n)|| \rightarrow 1$ for $n \rightarrow \infty$ and $N - n \rightarrow \infty$.

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 paper. 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 paper, 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.

$$P_1\alpha = \begin{bmatrix} 1 & -\frac{1}{2} & 0 & \cdots & 0 \\ -\frac{1}{2} & 1 & -\frac{1}{2} & \cdots & 0 \\ 0 & -\frac{1}{2} & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & -\frac{1}{2} \\ 0 & 0 & 0 & -\frac{1}{2} & 1 \end{bmatrix} \begin{bmatrix} \alpha_{n-1} \\ \vdots \\ \alpha_1 \end{bmatrix} = \begin{bmatrix} -\frac{1}{2} \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{14}$$

In the above Eq. (14), 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*[3]. 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. (14) 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

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

$$\Rightarrow ||\sum_{i=1}^{n-1}\alpha_i\phi_i||^2 \to \frac{1}{2} \text{ as } n \to \infty \tag{15}$$

Likewise using symmetrical arguments we can show that:

$$||\sum_{i=n+1}^{N}\alpha_i\phi_i||^2 = \frac{N-n-1}{N-n}\cdot\frac{1}{2} \to \frac{1}{2} \text{ as } N-n \to \infty \tag{16}$$

Therefore combining Eq. (15) and (16) we get:

$$||\mathcal{P}_{\mathcal{S}(\bigcup_1^N\{\phi_i\}\setminus\{\phi_n\})}(\phi_n)||^2$$

$$= ||\sum_{i=1}^{n-1}\alpha_i\phi_i||^2 + ||\sum_{i=n+1}^{N}\alpha_i\phi_i||^2 \to 1 \text{ for} n, N-n \to \infty \ \square$$

V. Proof of Lemma 3

Lemma 3. *Let $S = \{\phi_i\}_{i=1}^N$ denote the set of spikes generated by our framework, satisfying Assumption 2, i.e., $\forall n \in \{1, ..., N\}$, $||\mathcal{P}_{\mathcal{S}(\bigcup_{i=1}^N\{\phi_i\}\setminus\{\phi_n\})}(\phi_n)|| \le \beta$, 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 $||\mathcal{P}_{\mathcal{S}(S\setminus V)}(v)|| \le \beta_d$ where β_d is a real constant that depends on β and d. Specifically, we can show that $\beta_d^2 \le (1 + \frac{1-\beta^2}{d^2\beta^2})^{-1} < 1$.*

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:
$||\phi_{v_i}^{\parallel}|| \le \beta$ and $||\hat{\phi}_{v_i}||^2 \ge (1-\beta^2)$.
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 $m \in \{1, ..., d\}$.

$$||Y||^2 = ||\Sigma_{i=1}^d\alpha_i\mathcal{P}_{\mathcal{S}(S\setminus V)}(\phi_{v_i})||^2$$

$$\le (\Sigma_{i=1}^d|\alpha_i|||\mathcal{P}_{\mathcal{S}(S\setminus V)}(\phi_{v_i})||)^2 \quad \text{(Using Triangle inequality)}$$

$$\leq \beta^2(\Sigma_{i=1}^{d}|\alpha_i|)^2 \leq d^2\beta^2\alpha_m^2 \quad (17)$$

(Since $\forall n, ||\mathcal{P}_{\mathcal{S}(\bigcup_1^N\{\phi_i\}\setminus\{\phi_n\})}(\phi_n)|| \leq \beta$)

Again, $||Z||^2 = ||\alpha_m\hat{\phi}_{v_m} + \Sigma_{i\in\{1,...,d\}\setminus\{m\}}\alpha_i\hat{\phi}_{v_i}||^2 \quad (18)$

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 :

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

(since every $\hat{\phi}_{v_i}$ 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, $||\hat{\phi}_{v_m}^{\perp}||^2 = 1 - ||\mathcal{P}_{\mathcal{S}((S\setminus V)\cup(\cup_i\{\hat{\phi}_{v_i}\}\setminus\{\hat{\phi}_{v_m}\}))}(\phi_{v_m})||^2 \geq 1 - ||\mathcal{P}_{\mathcal{S}(S\setminus\{\phi_{v_m}\})}(\phi_{v_m})||^2 \geq 1 - \beta^2$ (by assumption) (19)

Combining (18) and (19) we get:

$$||Z||^2 = ||\alpha_m\hat{\phi}_{v_m} + \Sigma_{i\in\{1,...,d\}\setminus\{m\}}\alpha_i\hat{\phi}_{v_i}||^2$$
$$\geq \alpha_m^2||\hat{\phi}_{v_m}^{\perp}||^2 \geq \alpha_m^2(1-\beta^2) \quad (20)$$

From (17) and (20) 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. □

VI. Complete Proof of Windowing Theorem

The following section presents a detailed proof of the *Windowing Theorem* as discussed in the main paper. While the main text includes the theorem and its basic proof, this supplementary section provides a comprehensive proof of *Claim 4.2* and *Claim 4.3*, as well as a complete derivation for the final part of the theorem's proof. These additions are intended to enhance understanding and complement the content of the main paper. The theorem is restated below exactly as it appears in the main paper, followed by its proof with the help of Lemma 4, as referenced in the original text.

Theorem 4 (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. 12, is bounded. Specifically,

$$\forall\epsilon > 0, \exists w_0 > 0 \text{ s.t. } ||\mathcal{P}_{\phi_{n+1,w}^{\perp}}(X) - \mathcal{P}_{\phi_{n+1}^{\perp}}(X)|| < \epsilon,$$
$$\forall w \geq w_0 \text{ and } w \leq n \quad (21)$$

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 21 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 3. 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 $\phi_{n+1,w}^{\perp}$ and $\phi_{n+1}^{\perp}$ decreases steadily as w increases, based on the assumptions stated in the theorem. This ensures that $\phi_{n+1,w}^{\perp}$ becomes a good approximation of $\phi_{n+1}^{\perp}$. The proof of the theorem then follows by establishing a bound on $\|\mathcal{P}_{\phi_{n+1,w}^{\perp}}(X) - \mathcal{P}_{\phi_{n+1}^{\perp}}(X)\|$ for a given choice of w for any bounded input X. The lemma is provided below:

Lemma 4. *Under the conditions of Theorem 4, for any $\delta > 0$, there exists $w_0 \in \mathbb{N}^+$ such that $\|\phi_{n+1,w}^{\perp} - \phi_{n+1}^{\perp}\| < \delta\ \forall w \geq w_0, w \leq n$, 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 $\phi_{n+1}^{\perp}$ due to windowing, i.e., $\|\phi_{n+1,w}^{\perp} - \phi_{n+1}^{\perp}\|$, decreases faster than a geometric sequence as more of these partitions are included within the window. This convergence is illustrated schematically in Figure 3 of the main paper.

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 main paper). 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:

$$\begin{aligned}
\tilde{\phi}_i &= \phi_i - \mathcal{P}_{\mathcal{S}(\bigcup_{j=i+1}^{n}\{\phi_j\})}(\phi_i) \quad (p_k \leq i \leq p_{k-1}-1) \\
&= \phi_i - \mathcal{P}_{\mathcal{S}((\bigcup_{j=p_{k-1}}^{n}\{\phi_j\}) \bigcup (\bigcup_{j=i+1}^{p_{k-1}-1}\{\tilde{\phi}_j\}))}(\phi_i) \\
&= \phi_i - \mathcal{P}_{\sum_{j=1}^{k-1} V_j}(\phi_i) - \sum_{j=i+1}^{p_{k-1}-1} \mathcal{P}_{\tilde{\phi}_j}(\phi_i) \qquad (22) \\
&\quad (\text{As } \sum_{j=1}^{k-1} V_j \text{ and all } \tilde{\phi}_i\text{'s are mutually orthogonal} \\
&\quad \text{for } p_k \leq i \leq p_{k-1}-1)
\end{aligned}$$

Using the formulation in 22, 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:

$$\begin{aligned}
&y = x - \mathcal{P}_{\sum_{j=1}^{k-1} V_j}(x), \text{ for some } x = \sum_{i=p_k}^{p_{k-1}-1} \alpha_i \phi_i, \quad \alpha_i \in \mathbb{R} \\
&\Rightarrow y = \sum_{i=p_k}^{p_{k-1}-1} \alpha_i \phi_i - \mathcal{P}_{\sum_{j=1}^{k-1} V_j}(\sum_{i=p_k}^{p_{k-1}-1} \alpha_i \phi_i) \\
&= \sum_{i=p_k}^{p_{k-1}-1} \alpha_i (\phi_i - \mathcal{P}_{\sum_{j=1}^{k-1} V_j}(\phi_i)) \text{ (by linearity of projection)} \\
&\Rightarrow y = \sum_{i=p_k}^{p_{k-1}-1} \alpha_i (\tilde{\phi}_i + \sum_{j=i+1}^{p_{k-1}-1} \mathcal{P}_{\tilde{\phi}_j}(\phi_i)) \text{ (using (22))} \qquad (23)
\end{aligned}$$

Equation (23) 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 22, 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 (22) to observe the following:

$$\tilde{\phi}_{n-1} = (\phi_{n-1} - \mathcal{P}_{\sum_{j=1}^{k-1} V_j}(\phi_{n-1})) - \sum_{j=n}^{p_{k-1}-1} \mathcal{P}_{\tilde{\phi}_j}(\phi_{n-1}) \qquad (24)$$

Analyzing the expression for $\tilde{\phi}_{n-1}$ on the right-hand side of equation (24), 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:

$$\begin{aligned}
U_k &= \sum_{i=k}^{m} \tilde{V_i} = \mathcal{S}(\bigcup_{j=1}^{p_{k-1}-1}\{\tilde{\phi}_j\}) \qquad (25) \\
&= \{x - \mathcal{P}_{\sum_{j=1}^{k-1} V_j}(x) | x \in \sum_{i=k}^{m} V_i\} \qquad (26)
\end{aligned}$$

where the equality between 25 and 26 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:

$$\begin{aligned}
e_k &= ||\phi^{\perp}_{n+1,v_k} - \phi^{\perp}_{n+1}|| \\
&= ||\mathcal{P}_{\sum_{i=1}^{k} V_i}(\phi_{n+1}) - \mathcal{P}_{\sum_{i=1}^{m} V_i}(\phi_{n+1})|| \\
&= ||\mathcal{P}_{\sum_{i=1}^{k} V_i}(\phi_{n+1}) - \mathcal{P}_{\sum_{i=1}^{k} V_i + U_{k+1}}(\phi_{n+1})|| \text{ (by def. of } U_k) \\
&= ||\mathcal{P}_{\sum_{i=1}^{k} V_i}(\phi_{n+1}) - (\mathcal{P}_{\sum_{i=1}^{k} V_i}(\phi_{n+1}) + \mathcal{P}_{U_{k+1}}(\phi_{n+1}))|| \\
&\qquad (\text{Since by construction } U_{k+1} \perp \Sigma_{i=1}^{k} V_i) \\
&\Rightarrow e_k = ||\mathcal{P}_{U_{k+1}}(\phi_{n+1}))|| \qquad (27)
\end{aligned}$$

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

$$e_k = ||\mathcal{P}_{U_{k+1}}(\phi_{n+1}))|| \leq |\beta| < 1, \ \forall k \in \{1,...,m\} \qquad (28)$$

Now Assume that:

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

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:

$$\alpha_k = \frac{\langle \phi_{n+1}, \tilde{\psi}_k \rangle}{||\tilde{\psi}_k||^2} \tag{29}$$

For $k > 0$ from 29 and 27 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)

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

$$\Rightarrow e_k = \frac{|\langle \phi_{n+1}, \mathcal{P}_{\tilde{V}_k}(\psi_k) \rangle|}{||\tilde{\psi}_k||} \; (k > 0, \psi_k \in \sum_{i=k+1}^{m} V_i \perp \sum_{j=1}^{k-1} V_j) \tag{30}$$

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 $U_{k+1} \perp \tilde{V}_k$)

$$\Rightarrow e_{k-1}^2 = e_k^2 + ||\mathcal{P}_{\tilde{V}_k}(\phi_{n+1}))||^2 \tag{31}$$

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

$$\Rightarrow ||\mathcal{P}_{\tilde{V}_k}(\phi_{n+1}))|| = ||\beta_k\tilde{\theta}_k|| = \frac{|\langle \phi_{n+1}, \tilde{\theta}_k \rangle|}{||\tilde{\theta}_k||} \tag{32}$$

From 31 & 32 we get, $e_{k-1}^2 = e_k^2 + \frac{|\langle \phi_{n+1}, \tilde{\theta}_k \rangle|^2}{||\tilde{\theta}_k||^2}$ (33)

Again, for $k > 0$ we further analyze e_k from 30 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,

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

$$\Rightarrow e_k = \frac{|\langle \phi_{n+1}, \tilde{\theta}_k \rangle|}{||\tilde{\theta}_k||} \frac{|\langle \psi_k, \frac{\tilde{\theta}_k}{||\tilde{\theta}_k||} \rangle|}{||\tilde{\psi}_k||} \tag{34}$$

Combining 33 and 34 we obtain:

$$e_{k-1}^2 = e_k^2 + e_k^2 \frac{||\tilde{\psi}_k||^2}{|\langle \psi_k, \frac{\tilde{\theta}_k}{||\tilde{\theta}_k||} \rangle|^2} \tag{35}$$

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

$$\Rightarrow e_{k-1}^2 \geq e_k^2(1 + ||\tilde{\psi}_k||^2) \tag{36}$$

Eq. 36 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 $\tilde{\eta}_k \in \tilde{V}_k, ||\tilde{\eta}_k|| = 1$

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

$$\Rightarrow \mathcal{P}_{\psi_k}(\tilde{\eta}_k) = \langle \psi_k, \tilde{\eta}_k \rangle \psi_k, \text{ (Since, } ||\psi_k|| = 1) \tag{37}$$

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,

$$||\eta_k||^2 = 1 + ||\mathcal{P}_{\sum_{i=1}^{k-1} V_i}(\eta_k)||^2 \tag{38}$$

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

$$= \mathcal{P}_{\psi_k}(\eta_k) \quad \text{(Since } \psi_k \perp \sum_{i=1}^{k-1} V_i) \tag{39}$$

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 4, 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.

$$||\mathcal{P}_{\mathcal{S}(\{\psi_k\}) + \sum_{i=1}^{k-1} V_i}(\frac{\eta_k}{||\eta_k||})||^2 \leq \beta_d^2 \quad (|\beta_d| < 1)$$

$$\Rightarrow \frac{||\mathcal{P}_{\psi_k}(\eta_k)||^2 + ||\mathcal{P}_{\sum_{i=1}^{k-1} V_i}(\eta_k)||^2}{||\eta_k||^2} \leq \beta_d^2$$

$$\Rightarrow \frac{||\mathcal{P}_{\psi_k}(\eta_k)||^2 + ||\mathcal{P}_{\sum_{i=1}^{k-1} V_i}(\eta_k)||^2}{1 + ||\mathcal{P}_{\sum_{i=1}^{k-1} V_i}(\eta_k)||^2} \leq \beta_d^2 \quad \text{(using 38)}$$

$$\Rightarrow ||\mathcal{P}_{\psi_k}(\eta_k)||^2 \leq \beta_d^2(1 + ||\mathcal{P}_{\psi_k}(\eta_k)||^2) - ||\mathcal{P}_{\psi_k}(\eta_k)||^2$$

$$\Rightarrow ||\mathcal{P}_{\psi_k}(\eta_k)||^2 \leq \beta_d^2 - (1 - \beta_d^2)||\mathcal{P}_{\psi_k}(\eta_k)||^2$$

$$\Rightarrow ||\mathcal{P}_{\psi_k}(\eta_k)||^2 \leq \beta_d^2 \quad \text{(Since} |\beta_d| < 1)$$

$$\Rightarrow ||\mathcal{P}_{\psi_k}(\tilde{\eta}_k)||^2 = ||\mathcal{P}_{\psi_k}(\eta_k)||^2 \leq \beta_d^2$$

$$\Rightarrow ||\mathcal{P}_{\tilde{\eta}_k}(\psi_k)|| = ||\mathcal{P}_{\psi_k}(\tilde{\eta}_k)|| \leq \beta_d^2$$

(Since both $\tilde{\eta}_k$ and ψ_k are unit vectors)

Therefore, $||\tilde{\psi}_k||^2 = ||\psi_k - \mathcal{P}_{\tilde{V}_k}(\psi_k)||^2 = 1 - ||\mathcal{P}_{\tilde{\eta}_k}(\psi_k)||^2$

(Since by assumption $\mathcal{P}_{\tilde{V}_k}(\psi_k) = \beta_k \tilde{\eta}_k$ and $||\psi_k|| = 1$)

$\Rightarrow ||\tilde{\psi}_k||^2 \geq 1 - \beta_d^2 \qquad (|\beta_d| < 1)$ □

Finally, combining 36 with Claim 4.3, we obtain:

$$e_{k-1}^2 \geq e_k^2(1 + (1 - \beta_d^2)) \Rightarrow e_k^2 \leq \frac{e_{k-1}^2}{\gamma^2} \tag{40}$$

where $\gamma^2 = (1 + (1 - \beta_d^2))$ is a constant strictly greater than 1. Since $e_m = 0$, Eq. 40 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. 40 and the bound $e_1 < 1$ (Eq. 28), 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. □

Proof of Theorem 4: 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:

$$\tilde{X} = \mathcal{P}_{\mathcal{S}(\{\phi_{n+1,w}^{\perp}, \phi_{n+1}^{\perp}\})}(X)$$

$$X_u = \mathcal{P}_{\phi_{n+1}^{\perp}}(X)$$

$$X_v = \mathcal{P}_{\phi_{n+1,w}^{\perp}}(X)$$

So that $\tilde{X}, X_u$ and X_v lie in the same plane (see Fig. 1). 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. 1). 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:

$$\begin{aligned} p_w &= \phi_{n+1,w}^{\perp} - \phi_{n+1}^{\perp} \\ &= (\phi_{n+1} - \mathcal{P}_{\mathcal{S}(\{\phi_{n-w+1},\ldots,\phi_n\})}(\phi_{n+1})) \\ &\quad - (\phi_{n+1} - \mathcal{P}_{\mathcal{S}(\{\phi_1,\ldots,\phi_n\})}(\phi_{n+1})) \end{aligned} \tag{41}$$

(by definition) and thus:

$$\begin{aligned} p_w &= \mathcal{P}_{\mathcal{S}(\{\phi_{n-w+1},\ldots,\phi_n\})}(\phi_{n+1}) \\ &\quad - \mathcal{P}_{\mathcal{S}(\{\tilde{\phi}_1,\ldots,\tilde{\phi}_w\} \cup \{\phi_{n-w+1},\ldots,\phi_n\})}(\phi_{n+1}) \\ &= \mathcal{P}_{\mathcal{S}(\{\tilde{\phi}_1,\ldots,\tilde{\phi}_w\})}(\phi_{n+1}) \\ &\quad \text{(Since } \mathcal{S}(\{\tilde{\phi}_1,\ldots,\tilde{\phi}_w\}) \perp \mathcal{S}(\{\phi_{n-w+1},\ldots,\phi_n\}) \\ &= \mathcal{P}_{\mathcal{S}(\{\tilde{\phi}_1,\ldots,\tilde{\phi}_w\})}(\phi_{n+1} - \mathcal{P}_{\mathcal{S}(\{\phi_{n-w+1},\ldots,\phi_n\})}(\phi_{n+1})) \\ &= \mathcal{P}_{\mathcal{S}(\{\tilde{\phi}_1,\ldots,\tilde{\phi}_w\})}(\phi_{n+1,w}^{\perp}) \\ &\Rightarrow p_w = \phi_{n+1,w}^{\perp} - \phi_{n+1}^{\perp} = \mathcal{P}_{\mathcal{S}(\{\tilde{\phi}_1,\ldots,\tilde{\phi}_w\})}(\phi_{n+1,w}^{\perp}) \\ &\Rightarrow p_w \perp \phi_{n+1}^{\perp} \end{aligned}$$

(Since $\mathcal{P}_S(x) \perp (x - \mathcal{P}_S(x))$ for any $x \in \mathcal{H}, S \subseteq \mathcal{H}$)

Therefore, from Figure 1, we observe:

$$\sin^2(a - b) = \frac{||p_w||^2}{||\phi_{n+1,w}^{\perp}||^2} \tag{42}$$

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

$$\begin{aligned} &||\mathcal{P}_{\phi_{n+1,w}^{\perp}}(X) - \mathcal{P}_{\phi_{n+1}^{\perp}}(X)|| = ||X_u - X_v||^2 \\ &= ||X_u||^2 + ||X_v||^2 - 2||X_u||||X_v||cos(a - b) \\ &= ||\tilde{X}||^2[\cos^2 a + \cos^2 b - 2\cos a \cos b \cos(a - b)] \\ &\quad \text{(using the geometry of Figure 1)} \\ &= ||\tilde{X}||^2[\frac{1 + \cos 2a}{2} + \frac{1 + \cos 2b}{2} \\ &\qquad - (\cos(a - b) + \cos(a + b))\cos(a - b)] \\ &= ||\tilde{X}||^2[\frac{1}{2} - \frac{\cos 2(a - b)}{2}] = ||\tilde{X}||^2 \sin^2(a - b) \\ &= ||\tilde{X}||^2 \frac{||p_w||^2}{||\phi_{n+1,w}^{\perp}||^2} \quad \text{(from Eq. (42))} \\ &\leq \frac{||X||^2}{1 - \beta^2} ||\phi_{n+1,w}^{\perp} - \phi_{n+1}^{\perp}||^2 \end{aligned}$$

(Since $||\tilde{X}|| \leq ||X||$ and $||\phi_{n+1,w}^{\perp}|| \geq ||\phi_{n+1}^{\perp}|| \geq (1 - \beta^2)$ by assumption)

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 $||\mathcal{P}_{\phi_{n+1,w}^{\perp}}(X) - \mathcal{P}_{\phi_{n+1}^{\perp}}(X)|| \leq \frac{||X||}{\sqrt{1-\beta^2}} ||\phi_{n+1,w}^{\perp} - \phi_{n+1}^{\perp}|| < \epsilon, \forall \epsilon > 0, \forall w \geq w_0$ 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. □

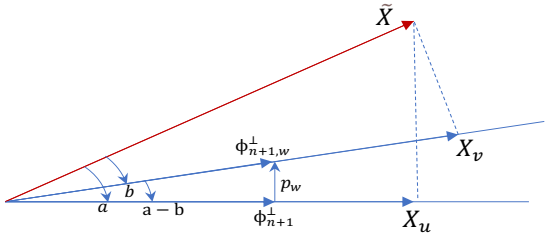


Fig. 1. Figure illustrating the vector projections of the input signal X onto vectors $\phi_{n+1}^{\perp}$ and $\phi_{n+1,w}^{\perp}$. The red vector represents $\tilde{X}$, the projection of X within the plane formed by $\phi_{n+1}^{\perp}$ and $\phi_{n+1,w}^{\perp}$. The vectors $\phi_{n+1}^{\perp}$ and $\phi_{n+1,w}^{\perp}$, 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 $\phi_{n+1,w}^{\perp}$ and $\phi_{n+1}^{\perp}$, 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 $\phi_{n+1,w}^{\perp}$ and $\phi_{n+1}^{\perp}$ are marked.

VII. Parameter Sensitivity in Experimental Results

The results of the comprehensive experiments presented in Section VII of the main paper depend on several parameter value choices. In this section, we succinctly describe the rationale behind these choices and present detailed sensitivity test results.

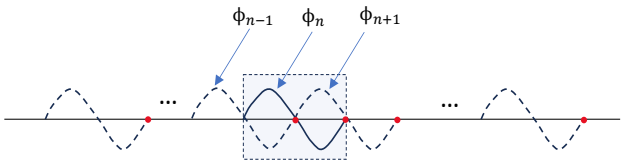


Fig. 2. An example scenario of infinite sequence of spikes in which every spike corresponding to one full sine wave and are arranged such that the spike ϕ_n's tail fully overlaps with a similar overlapping head of previous spike ϕ_{n-1} and likewise the head of ϕ_n fully overlaps with the tail of the next spike ϕ_{n+1}. In this infinite sequence of spikes for very large n, the spike ϕ_n get fully represented by the all other spikes.

Grids Search for the Parameters of Threshold Eq. 1: The Threshold equation (Eq. 1) consists of three parameters, namely C (the baseline threshold), M (the *ahp* increment), and δ (the refractory period). In the comprehensive experiments described in Section VII, for each sound snippet, the first two parameters (C and M) were held fixed while the refractory period (δ) was systematically varied to obtain reconstructions at varying spike rates. To determine the initial values of C and M, we conducted a grid search on a small set of 20 short-length audio snippets (approximately 250 ms each), using the same set of kernels (about 50) as in the large experiments in Section VII. By selecting parameters on a small subset rather than the full dataset, we ensured that the final large-scale experiments on 600 snippets tested the generalization of these choices and avoided overfitting. The exact parameter values used in the grid search are tabulated in Table I. Detailed results of the grid search can be found in the GitHub repository[28] under the csv folder. The baseline threshold C was chosen within the range of 5×10^{-5} to 5×10^{-10}. Each value of C was paired with a corresponding value of M, which was represented as a multiple of C. Specifically, for each value x of C in the table and each value y shown for M, the *ahp* increment M was set as $M = x \times y$. The rationale for selecting M to be orders of magnitude larger than the baseline threshold follows from corollary 0.1, which suggests that for spikes to remain well separated, M needs to be larger than the range of the convolution $C(t)$ of the signal $X(t)$ with kernel ϕ^j. Additionally, the baseline threshold C should be lower than the range of $C(t)$ to ensure that the system produces spikes as dictated by corollary 0.3. From a numerical accuracy perspective, it is crucial that the values of M be order of magnitude higher than the rate of change of the convolution values. This ensures that the *ahp* increment M is high enough so that threshold adjustments are not immediately canceled out by fluctuations of $C(t)$. Consequently, in the grid search, M was tested in the range of 1 to 1000 times C. For each combination of C and M, the refractory period was varied over the values listed in Table I to ensure that the selected grid search parameters performed well across varying spike rates.

Results of our grid search experiment clearly reveal the sensitivity of our technique with respect to the threshold parameters discussed above. For example, Fig. 3 shows how the reconstruction accuracies change with varying baseline threshold (C). The optimal choice of the baseline threshold is subjective and expected to depend on the dynamic range of the convolution between the input signal and the kernels as confirmed by the results of the grid search experiments. Specifically, if the baseline threshold is too high compared to the dynamic range of convolution, the system will generate very few spikes and the reconstruction would be poor. On the other hand, if the baseline threshold is too low compared to the dynamic range of the convolution, the system will unnecessarily generate too many spikes. Based on the grid search results, as indicated in Fig. 3, we found the baseline threshold value of 5×10^{-6} to work best, which was then chosen as the initial baseline threshold for our large scale experiments. Also based on the grid search results, we picked the value for M to be 100 times C, i.e. 5×10^{-4}. In agreement with the theoretical insights from corollaries 0.1 and 0.3, one final modification applied in our experiments was allowing C to drop dynamically from the initial value based on the convolution values. Empirically, C was reduced by a factor of 10 until the condition $C < {max\_conv}/{100}$ was met, where max_conv represents the maximum convolution value of the snippet with any kernel in the system. Correspondingly, M was adjusted to always remain at $100 \times C$. This approach ensured that C and M were determined adaptively for each snippet. Once selected, C and M remained fixed for each snippet, and the only parameter that was varied systematically was the refractory period δ, which controlled reconstruction at different spike rates. The range of δ used in the large-scale experiments was the same as that in the grid search, as shown in Table I.

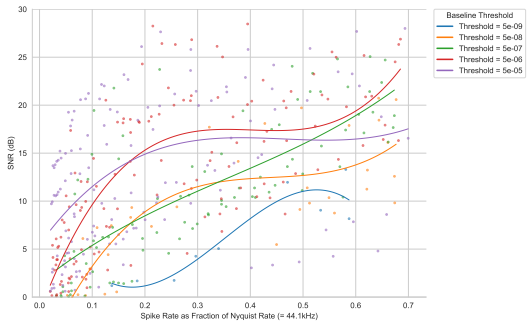


Fig. 3. Scatter plot of reconstructions of the grid search experiment (Section VII) for different choices of the baseline threshold (C). Here, for each fixed baseline threshold (C) the reconstructions were obtained at different spike rates by systematically varying the refractory period. For the displayed reconstructions, the value of *ahp*-high (M) was chosen to be $100 \times C$. Reconstructions corresponding to each baseline threshold have been color-coded accordingly. Based on the plot, while we see that reconstruction with varying baseline threshold follows similar trends, there is no clear correlation between the baseline threshold (C) and the average SNR; in fact accuracy depends heavily on the dynamic range of the signal to kernel convolution as explained in Section VII. For the given set of reconstructions, a baseline threshold value of 5×10^{-6} performed best and therefore was chosen as the initial baseline threshold value for our large scale experiment.

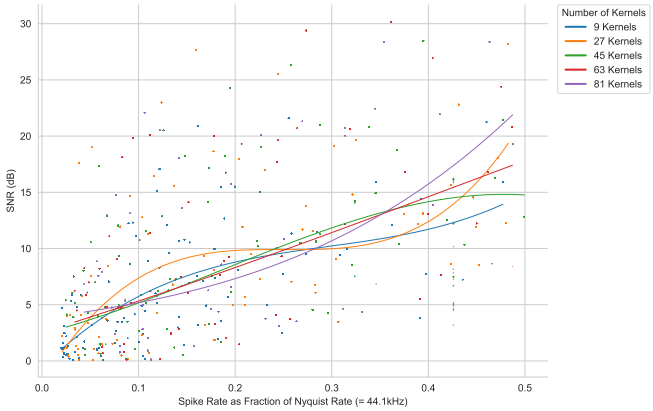
Choice of Kernels: The theoretical framework presented up until Section VI is agnostic to the specific choice of kernels, i.e., the framework has been developed for a generalized set of kernels satisfying certain constraints that are easy to satisfy

TABLE I
PARAMETER SETTINGS USED IN GRID SEARCH

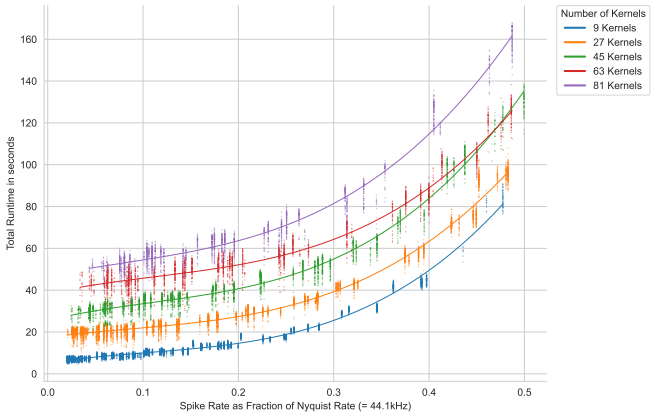
Baseline Threshold C	*Ahp* Increment M (as multiples of C)	Refractory Period δ (unit of $1/44.1kHz$)
5×10^{-5}	1	1000
5×10^{-6}	10	700
5×10^{-7}	100	400
5×10^{-8}	1000	100
5×10^{-9}		90
5×10^{-10}		80
		70
		60
		50
		40
		30
		20
		10

in real biological systems. However, going by Theorem III, it is important to note that for better representation of a certain class of signals, a suitable set of kernels ought to be chosen so that they represent the signals well. In our experiment on audio signals we chose to use gammatone kernels since they are widely used as a reasonable model of cochlear filters in auditory systems[4]. The gammatone filterbank, mathematically of the form: $at^{n-1}e^{-2\pi bt}\cos(2\pi ft+\phi)$, with f denoting the center frequency and b the bandwidth, is implemented with chosen center frequencies commonly spaced uniformly along the ERB scale[5]; the ERB scale is defined by the relationship: $ERB(f) = 24.7(4.37\frac{f}{1000}+1)$, when the order n is chosen to be 4. We adopted this implementation of the gammatone filterbank as our set of kernels with the exact values of the center frequencies and bandwidths shown in Table II. Going by Theorem III again, the more kernels we choose, the greater is the representational capacity of the set of kernels, and therefore, the better is the reconstruction accuracy. However, increasing the number of kernels also increases the overall spike rate of the system and the overlap between spikes, compromising the stability of the system. Also, from an implementation standpoint, increasing the number of kernels increases both the runtime and memory requirements for our experiment. Also, an increase in overlap between spikes leads to a decrease in the stability of the system, and therefore, based on Theorem VI requires larger window sizes, which in turn increases the runtime further. Finally, an issue that comes into play while using our framework for communicating signals in an online setup is that one needs to convey spike indices along with the spike times to fully communicate the code, and therefore increasing the size of the kernel set increases the bit length required to communicate the kernel index and therefore increases the overall bandwidth for communicating the spike ensemble code.

To empirically evaluate the trade-offs between accuracy, spike-rate and runtime complexity, we ran a small set of experiments on about 20 audio snippet of short length (typically of duration 250ms). Results shown in Fig. 4a and 4b demonstrate the above trade-off as the number of kernels increase. As the figures suggest, increasing the kernel count marginally increases the average accuracy at a given spike-rate and increases the runtime rapidly. Based on these results, we empirically chose the kernel count to be 50 for conducting our large scale experiments. Of those 50 kernels the top 10% of the longer kernels with the lowest center frequencies were discarded to speed up the process further.



(a)



(b)

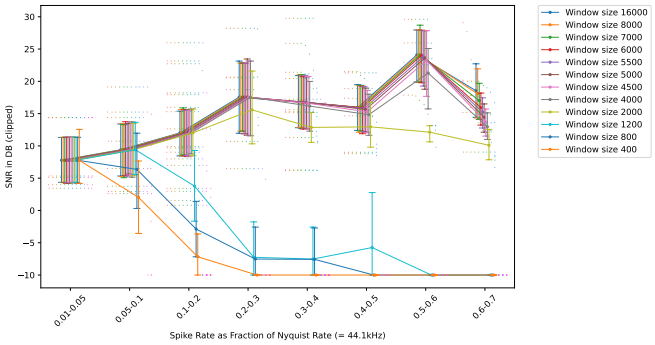
Fig. 4. (a) Scatter plot of reconstructions for 20 audio snippets with varying kernel counts (Section VII). The x-axis represents spike rate as a fraction of the Nyquist rate, and the y-axis shows SNR (dB). Different colors indicate different kernel counts. Reconstruction accuracy slightly improves with more kernels, though the improvement is limited due to the trade-off between representation capacity and kernel overlap (Section VII). (b) Reconstruction runtimes corresponding to (a), measured on a 10-core Intel(R) Xeon(R) CPU E5-2650 v3 server. Runtime increases with the number of kernels.

Choice of Window Size: As discussed in Section VI, choosing the window size involves a trade-off between reconstruction accuracy, runtime efficiency, and stability. A suitable window size ensures faster and stable reconstruction. On the other hand, if the window size is too small, the windowed reconstruction significantly deviates from the optimal solution formulated in Section III. According to Windowing Theorem VI, the optimal window size primarily depends on the spike rate of the ensemble. To empirically establish this relationship, we conducted experiments using 20 audio snippets of medium length ($\approx$ 500 ms) with the same set of kernels ($\approx$ 50) used in the large scale experiments, with varying window size. Figure 5 furnishes the results of this sensitivity test with respect to window sizes. Clearly, smaller window sizes suffer higher reconstruction errors, and as the spike rate increases, larger

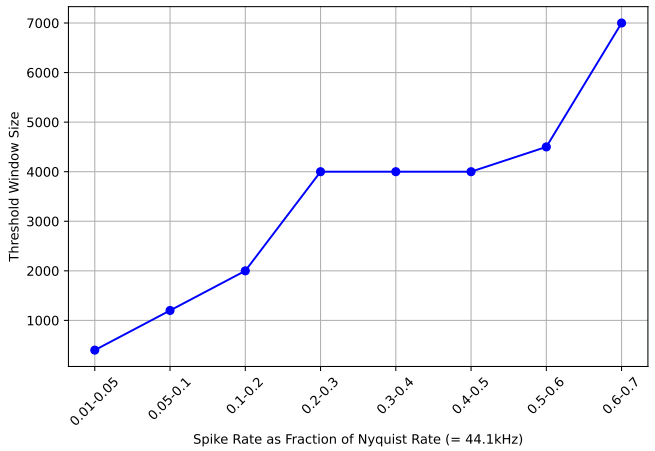
Center Frequency (Hz)	Bandwidth (Hz)
20.22	2.21
43.11	4.68
68.11	7.38
95.41	10.32
125.22	13.54
157.76	17.05
193.30	20.89
232.11	25.08
274.48	29.65
320.75	34.65
371.27	40.10
426.44	46.05
486.67	52.56
552.45	59.66
624.27	67.41
702.69	75.87
788.32	85.12
881.83	95.21
983.93	106.23
1095.41	118.26
1217.15	131.40
1350.07	145.75
1495.22	161.42
1653.70	178.52
1826.76	197.20
2015.72	217.60
2222.06	239.87
2447.36	264.19
2693.38	290.74
2962.01	319.74
3255.33	351.40
3575.62	385.97
3925.35	423.72
4307.23	464.94
4724.22	509.95
5179.54	559.10
5676.71	612.76
6219.59	671.36
6812.38	735.35
7459.66	805.21
8166.44	881.50
8938.19	964.80
9780.89	1055.76
10701.05	1155.09
11705.80	1263.54
12802.92	1381.96
14000.89	1511.27
15308.98	1652.46
16737.33	1806.64
18296.98	1974.98

TABLE II
CENTER FREQUENCIES AND CORRESPONDING BANDWIDTHS OF THE KERNELS CHOSEN IN OUR COMPREHENSIVE EXPERIMENT.

window sizes also gradually begin to deviate from the optimal reconstruction. To determine the required window size at a given spike rate, we plotted the threshold window size–defined as the minimum window size incurring less than a 10% deviation from the average optimal reconstruction error for all spike rate bins lower than or equal to a given spike rate–as a function of spike rate in Fig. 5b. Clearly, the general trend is that as the spike rate increases, larger window sizes are necessary to keep the windowing error in check. Based on this observation, we set the window size in our large scale experiment to be inversely proportional to the refractory period (δ), which governs the spike rate. The exact formula for the window size was empirically chosen as: $w = \frac{5\times10^6}{\delta\times44.1kHz}$. Admittedly, the required window size also depends on the nature of the kernels. For instance, high-frequency gammatone kernels with shorter effective lengths have significantly less mutual overlap at the same spike rate compared to low frequency kernels with longer effective lengths. To confirm this, we conducted the same set of experiments to test windowing sensitivity, but this time we used only half of the kernels, specifically the higher-frequency ones from Table II. These have shorter effective lengths and thus were expected to have lower inter-spike overlap. The results shown in Fig. 6 reveal that using only the higher-frequency kernels, we achieve significantly smaller threshold window sizes compared to the previous experiment involving all kernels from Table II. This observation is consistent with the theory.

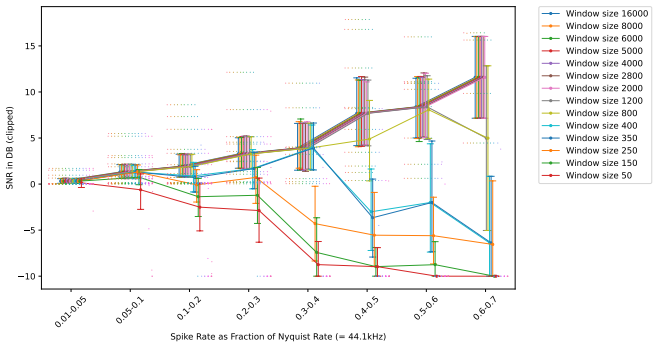


(a)

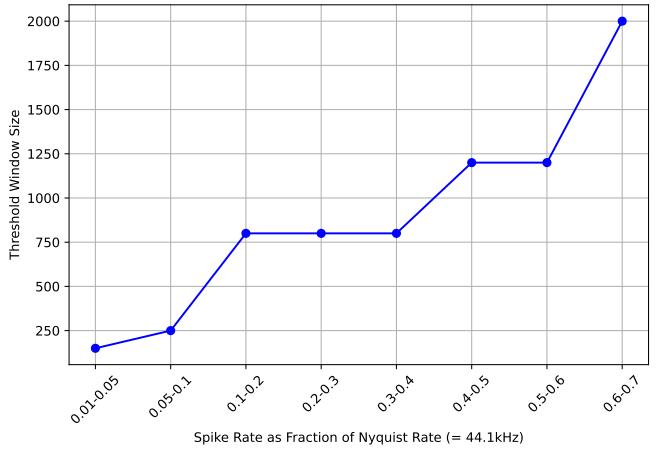


(b)

Fig. 5. Results of sensitivity test of our proposed reconstruction based on different window sizes, conducted on a set of 20 audio of snippets of medium length ($\approx$ 500 ms). **(a)** Strip plot of SNR values of reconstruction against spike rates bins. Different colors correspond to different window sizes, as specified by the legend on the right side. The SNR values are clipped at $-10dB$ to mitigate scaling issues caused by large windowing errors at smaller window sizes. Clearly, as the window size increases, the average SNR line closely approximates the optimal line. Smaller window sizes deviate from the optimal line as the spike rate increases; the smaller the window size, the faster it deviates from optimal. **(b)** Plot of the threshold window size as a function of spike rate, where the threshold window size at a given spike rate is defined from (a) above as the minium window size necessary to obtain an average SNR value within 10% of the optimal average SNR for all spike rate bins up to and including the given spike rate. Clearly, the threshold window size steadily increases with spike rate.



(a)



(b)

Fig. 6. Similar to Fig. 5, this figure shows the sensitivity test results with respect to window size in reconstruction, except this time we only use the bottom half (high frequencies) kernels of the Table II. **(a)** Strip plot shows the SNR values of reconstruction against spike rates bins, where different colors represent different window sizes. **(b)** Plot of the threshold window size as a function of spike rate. Clearly, threshold window sizes here are significantly lower compared to those shown in Fig. 5b.

REFERENCES

[1] A. Chattopadhyay and A. 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)*, 2023, pp. 1–5.

[2] B. Schölkopf, R. Herbrich, and A. J. Smola, "A generalized representer theorem," in *International Conference on Computational Learning Theory*. Springer, 2001, pp. 416–426.

[3] C. da Fonseca and J. Petronilho, "Explicit inverses of some tridiagonal matrices," *Linear Algebra and its Applications*, vol. 325, no. 1-3, pp. 7–21, 2001.

[4] R. Patterson, I. Nimmo-Smith, J. Holdsworth, and P. Rice, "An efficient auditory filterbank based on the gammatone function," *APU Report*, vol. 2341, 1987.

[5] B. R. Glasberg and B. C. J. Moore, "Derivation of auditory filter shapes from notched-noise data," *Hearing Research*, vol. 47, no. 1–2, pp. 103–138, 1990.