DACTM: Drift-Adaptive Continuous Thought Machines for Robust Cross-Day Decoding in Invasive Brain-Computer Interfaces

1st Chuwen Jiang#
The Center for Excellence in Brain Science and Intelligence Technology, Chinese Academy of Sciences,
Shanghai, China
ec41139@163.com

2nd Liyuan Han#
The Center for Excellence in Brain Science and Intelligence Technology, Chinese Academy of Sciences,
Shanghai, China
hanliyuan@ion.ac.cn

3rd Tielin Zhang*
Center for Excellence in Brain Science and Intelligence Technology, Chinese Academy of Sciences
State Key Laboratory of Brain Cognition and Brain-inspired Intelligence Technology
Institute of Neuroscience, Chinese Academy of Sciences
School of Artificial Intelligence, University of Chinese Academy of Sciences
The Key Laboratory of Cognition and Decision Intelligence for Complex Systems,
Institute of Automation, Chinese Academy of Sciences
Shanghai, China
zhangtielin@ion.ac.cn

***Abstract*—Invasive brain–computer interfaces (iBCIs) require stable decoding of cortical activity over long periods, yet neural signals exhibit strong temporal structure and substantial cross-day non-stationarity. Here, we present a drift-adaptive continuous thought machine (DACTM) method for long-term stable neural decoding. DACTM combines multi-scale temporal convolution with drift-context modeling to capture neural dynamics across time scales while adaptively tracking cross-day distributional shifts. A dual-path prediction architecture integrates accumulated temporal states with instantaneous neural context, while drift-aware attention queries modulate attention according to signal drift. We further develop a spiking-enhanced variant (S-DACTM) that embeds spiking mechanisms into temporal, attentional and prediction modules. Across macaque M1 datasets involving center-out reaching and isometric wrist movement, DACTM and S-DACTM outperform vanilla continuous thought machine (CTM), Long Short-Term Memory (LSTM) and time-aware feed-forward (TAFF) baselines, particularly under cross-day evaluation. These results highlight drift-adaptive modeling as a promising strategy for robust long-term iBCI decoding.**

***Index Terms*—Invasive Brain–Computer Interface (iBCI), cross-day decoding, continuous thought machine, drift-adaptive modeling, and spiking neural networks.**

I. Introduction

Invasive brain-computer interfaces (iBCIs) decode cortical activity to enable direct communication between the brain and external devices, with broad potential in motor rehabilitation, neural prosthetics, and intelligent control systems [1]. However, intracortical signals are temporally dependent and highly non-stationary across days. Session-to-session distributional shifts caused by neural drift and recording variability can markedly degrade long-term decoding stability [2], [3]. Robust cross-day generalization, therefore, remains a key challenge for practical iBCI systems [4].

Deep neural decoders, including convolutional, recurrent, Long Short-Term Memory (LSTM) and Transformer-based models, have improved neural decoding by capturing local and long-range temporal dependencies [5]–[7]. However, most existing models are not explicitly designed to adapt to cross-day neural drift or session-dependent distributional changes [2], [4], [8]. Traditional machine learning methods further rely on handcrafted features and have limited capacity to model nonlinear neural dynamics [9].

Continuous Thought Machines (CTMs) offer a biologically inspired framework for temporal representation learning through iterative internal computation and synchronization dynamics [10]. Yet vanilla CTM was mainly developed for visual inputs and lacks mechanisms for intracortical neural recordings, which are channel-dependent, temporally structured and non-stationary across sessions. To address this gap, we propose DACTM, a drift-adaptive continuous thought machine for neural decoding. DACTM combines multi-scale temporal encoding, drift-adaptive modeling, drift-aware attention and dual-path prediction to capture neural dynamics and cross-day distributional shifts.

We further develop S-DACTM, a spiking-enhanced variant that integrates spiking neural network (SNN) mechanisms into the temporal encoding and attention pathways while preserving decoding performance [11], [12]. We evaluate DACTM and S-DACTM on macaque M1 cortical datasets

This work was supported by projects 2024YFF1400600 and 2024YFF1400604. # Co-first authors. Corresponding author: Tielin Zhang

involving center-out reaching and isometric wrist movement tasks [13]. The results show that both models consistently outperform baseline methods under cross-day neural drift conditions, achieving improved decoding accuracy and stronger generalization.

The main contributions of this work are summarized as follows:

- We propose DACTM, a drift-adaptive continuous thought machine for robust intracortical neural decoding under cross-day non-stationarity.
- We integrate multi-scale temporal encoding, drift-adaptive modeling, drift-aware attention and dual-path prediction to capture neural dynamics and session-dependent distributional shifts.
- We develop S-DACTM, a spiking-enhanced variant of DACTM that incorporates spike-based computation while preserving decoding performance.
- We validate the proposed methods on macaque M1 cortical datasets, demonstrating superior decoding accuracy and cross-day generalization over baseline approaches.

II. Method

The original Continuous Thought Machine (CTM) is primarily designed for visual perception tasks and cannot be directly applied to intracortical neural decoding. Directly flattening neural recordings destroys temporal dependencies in neural activity and reduces the effectiveness of temporal representation learning [14]. In addition, cross-day neural drift causes substantial distributional shifts across recording sessions, leading to degraded long-term decoding performance. To address these limitations, this work proposes a drift-adaptive continuous thought machine for intracortical neural decoding, termed DACTM. The overall framework consists of three major components: a temporal drift-adaptive encoder, an iterative attention and prediction module, and a spiking-enhanced extension for biologically inspired neural decoding. The overall framework of the proposed DACTM is illustrated in Fig. 1.

A. Overview

Given an intracortical neural recording sequence $x_0 \in \mathbb{R}^{B\times T\times C}$, where B, T, and C denote batch size, temporal length and channel number, respectively. DACTM first transforms the raw neural signal into temporal token representations through a multi-scale temporal encoder. A drift-adaptive modeling mechanism is then introduced to capture cross-day neural distribution shifts and generate drift-aware contextual representations. The encoded representations are subsequently processed by iterative CTM attention dynamics with drift-aware attention queries. Finally, a dual-path prediction architecture jointly models synchronization states and contextual neural dynamics for decoding.

A spiking-enhanced variant termed S-DACTM is further introduced to improve biological plausibility and neuromorphic deployment capability by integrating spike-based computation into temporal encoding, attention projection, and prediction modules.

B. Temporal Drift-Adaptive Encoding

To preserve temporal structures in intracortical neural recordings, a temporal convolutional input module, termed *BCITemporalConvStem*, is introduced. Given an input neural sequence x_0, LayerNorm is first applied along the channel dimension:

$$x = \text{LayerNorm}(x_0) \tag{1}$$

The temporal and channel dimensions are then transposed to match the Conv1D input format.

To capture neural dynamics at multiple temporal scales, three parallel temporal convolution branches with kernel sizes $\{3, 5, 7\}$ are employed:

$$b_i = \text{Conv1d}_{k_i}(x_{\text{in}}), \quad k_i \in \{3, 5, 7\} \tag{2}$$

Multi-scale temporal features are subsequently fused through residual convolutional refinement to generate temporal token representations:

$$x_{\text{tokens}} = \text{BCITemporalConvStem}(x_0) \tag{3}$$

To improve robustness against cross-day neural drift, channel-wise statistical features including mean and standard deviation are computed for each neural recording. The normalized signal is defined as:

$$\hat{x} = \frac{x - \mu_c}{\sigma_c} \tag{4}$$

A drift representation vector d is further constructed from channel-wise and global statistical features. Both original and normalized signals are independently processed by the temporal encoder:

$$x_{\text{tokens}} = \text{BCITemporalConvStem}(x) \tag{5}$$

$$\hat{x}_{\text{tokens}} = \text{BCITemporalConvStem}(\hat{x}) \tag{6}$$

A fusion gate network generates adaptive fusion weights from the drift representation:

$$g = \text{fusion\_gate}(d) \tag{7}$$

The final drift-adaptive representation is computed as:

$$f_{\text{tokens}} = x_{\text{tokens}} + g \odot \hat{x}_{\text{tokens}} \tag{8}$$

To encode global drift conditions, a drift-context network generates a drift-aware contextual representation:

$$d_{\text{ctx}} = \text{drift\_context}(d) \tag{9}$$

Unlike the vanilla CTM positional encoding designed for visual spatial structures, DACTM only applies one-dimensional positional encoding along the temporal dimension to preserve sequential ordering while avoiding artificial spatial correlations between recording channels.

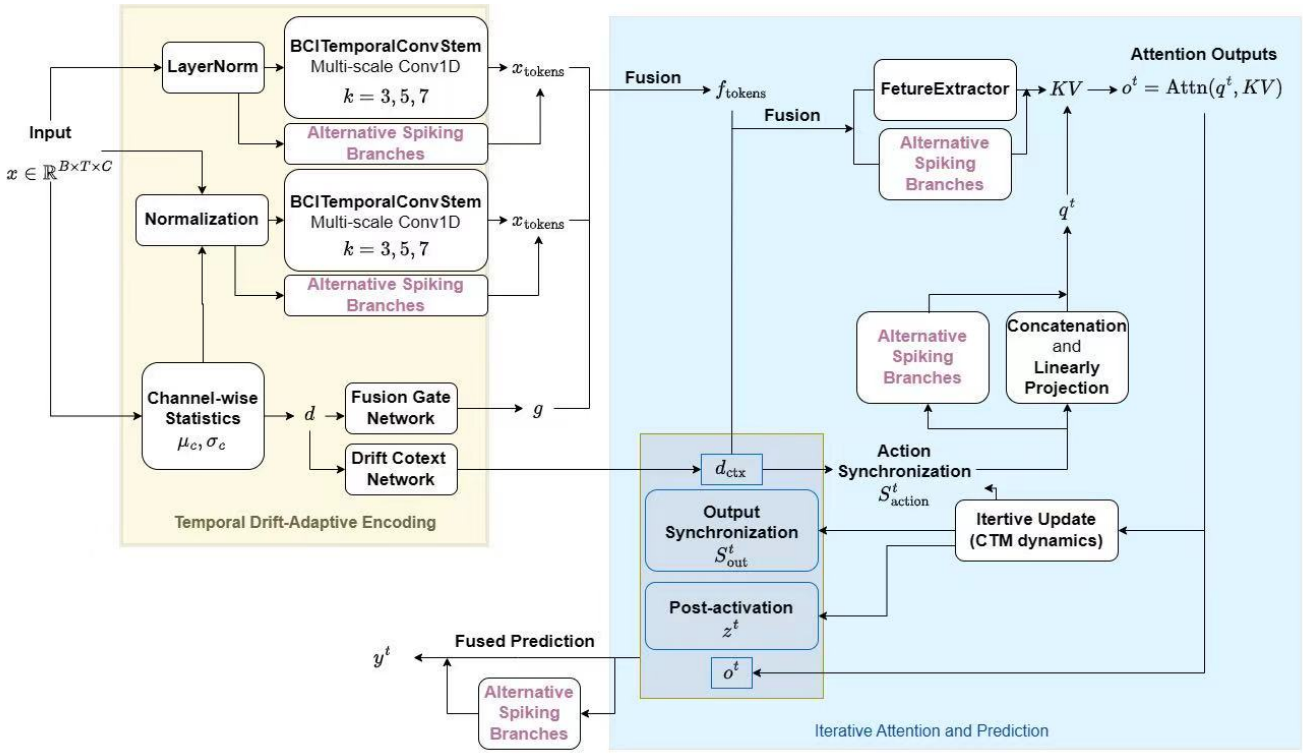

Fig. 1. Overview of the proposed DACTM and S-DACTM frameworks. The architecture corresponds to DACTM when the Alternative Spiking Branches module is disabled, and to S-DACTM when the module is enabled.

C. Iterative Attention and Prediction

The drift-adaptive token sequence is processed by iterative CTM attention dynamics. To improve sensitivity to instantaneous neural variations and cross-day drift conditions, the synchronization representation and drift-context vector are concatenated to construct drift-aware attention queries:

$$q^t_{\text{in}} = \text{concat}(S^t_{\text{action}}, d^t_{\text{ctx}}) \tag{10}$$

The resulting query representation is projected into the attention space:

$$q^t = W_{\text{in}} q^t_{\text{in}} \tag{11}$$

The drift context vector d_{ctx} is projected into the token feature space and concatenated with f_{tokens} to form the augmented token representation F_{tokens}, which is then used for key-value projection:

$$KV_{\text{ana}} = \text{FeatureExtractor}(F_{\text{tokens}}) \tag{12}$$

The attention output at iteration t can be written as

$$o^t = \text{softmax}\left(\frac{q^t K^T}{\sqrt{d_k}}\right) V \tag{13}$$

where d_k denotes the dimensionality of the query and key representations. Therefore, drift information influences the attention weights through the query construction process.

To jointly model long-term synchronization dynamics and instantaneous neural responses, a dual-path prediction architecture is introduced. The synchronization branch follows the vanilla CTM prediction mechanism:

$$y^t_{\text{syn}} = W_{\text{syn}} S^t_{\text{out}} \tag{14}$$

A contextual prediction branch further integrates attention outputs, post-activation states, synchronization states, and drift-context representations:

$$y^t_{\text{ctx}} = W_{\text{ctx}} C^t \tag{15}$$

The final prediction is obtained through adaptive fusion:

$$y^t = y^t_{\text{syn}} + \tanh(\alpha) y^t_{\text{ctx}} \tag{16}$$

D. Spiking-Enhanced CTM

To improve biological plausibility and neuromorphic deployment capability, S-DACTM introduces spike-based computation into the temporal encoder, attention projection, and prediction modules while preserving the stable optimization characteristics of DACTM.

In the temporal encoder, spiking temporal branches based on repeated Conv1D-LIF operations are employed to extract spike-based temporal representations:

$$s_i = \text{SpikingTemporalBranch}(x_{\text{in}}) \tag{17}$$

Hybrid analog-spiking fusion is further introduced into Q/K/V projection:

$$KV^t = KV^t_{\text{ana}} + \tanh(\delta_1) KV^t_{\text{spk}} \tag{18}$$

$$q^t = q^t_{\text{ana}} + \tanh(\delta_2) q^t_{\text{spk}} \tag{19}$$

Similarly, hybrid spiking projection is incorporated into the prediction head:

$$W_{\text{hybrid}} = W_{\text{ana}} + \tanh(\zeta) W_{\text{spk}} \tag{20}$$

During end-to-end training, gradients are propagated through both analog and spiking pathways. The analog pathway is initialized from pretrained DACTM checkpoints, whereas the newly introduced spiking modules are optimized jointly with the analog parameters. The gated fusion mechanisms defined in (12)–(20) scale the contribution of the spiking pathway in the forward pass and correspondingly modulate its gradient magnitude during backpropagation. As a result, the spike-based modules are introduced progressively as residual corrections to the pretrained analog representations. This warm-start and gated-fusion strategy stabilizes optimization while enabling both pathways to be trained in an end-to-end manner.

The proposed hybrid architecture preserves stable analog optimization during training while introducing spike-based computation as a potential route towards neuromorphic deployment.

E. Computational Efficiency

Although DACTM introduces drift-adaptive and spiking pathways, it preserves efficient temporal processing through lightweight one-dimensional convolutions and parameter-shared iterative attention. Unlike large Transformer-based sequence models, the framework avoids full-sequence attention and instead uses compact iterative state updates. In S-DACTM, spiking pathways further support event-driven computation, offering a potential route towards neuromorphic deployment.

III. Experimental Results and Analysis

To evaluate the proposed DACTM and S-DACTM frameworks, we conducted experiments on two non-human primate intracortical neural decoding datasets. The analysis focuses on decoding accuracy, cross-day generalization and robustness under neural drift.

A. Datasets and Preprocessing

We evaluated the proposed methods on two macaque M1 intracortical recording datasets: the center-out reaching task (CO) and the isometric wrist task (Jango). Both datasets were sampled at 30 kHz and involved motor control based on reaching or isometric wrist-force modulation [13].

Neural recordings were segmented into temporal sequences and used directly as model inputs without handcrafted feature extraction. During training, data augmentation, including gain perturbation, offset perturbation, trend perturbation, additive noise and channel dropout, was applied to improve robustness to cross-day neural drift.

B. Experimental Settings

DACTM and S-DACTM were compared with three baseline models: vanilla CTM, Long Short-Term Memory (LSTM) and Temporal-Aware Feed-Forward network (TAFF). For the CO dataset, we used a contiguous-tail split with 15% of the training data for validation. For the Jango dataset, recordings from 20150730∼20150805 were used for training, 20150806 for validation, and 20150808∼20151102 for testing.

All experiments were repeated with ten random seeds (159, 264, 265, 314, 323, 338, 358, 777, 846, 979). Models were trained on a single NVIDIA RTX A6000 GPU using AdamW, early stopping and learning-rate warmup. Unless otherwise specified, the default configuration used *d_model=128*, four attention heads, memory length 10, learning rate 3×10^{-4}, weight decay 2×10^{-4} and batch size 96. The training objective combined the final-step loss, the highest-confidence-step loss, and the averaged iterative-step loss. Both spiking modules are simulated for two discrete timesteps ($T_s = 2$) using LIF neurons.

C. Statistical Analysis

Classification accuracy was used as the primary metric. Results are reported as mean accuracy across ten random seeds. Median accuracy across testing days was also reported to assess robustness under cross-day drift. Statistical significance was evaluated using paired two-sided t-tests with a threshold of $p < 0.05$.

D. Warm-Start Initialization for S-DACTM

Because direct training of SNNs can be unstable, S-DACTM was initialized from pretrained DACTM checkpoints. Parameter-matched analog pathways were transferred from DACTM, whereas newly introduced spiking modules were randomly initialized. This warm-start strategy stabilizes early training and allows the spiking pathways to learn complementary corrections. Specifically, S-DACTM H was initialized from DACTM B, and S-DACTM G from DACTM A.

E. Ensemble Strategy

To further improve robustness, ensemble predictions were obtained by weighted averaging. Ensemble weights were selected on the validation set through grid search with a step size of 0.05. Three DACTM variants were evaluated, with model B enabling drift-adaptation modules and model C using increased capacity and dropout regularization. Two high-performing S-DACTM variants, G and H, were selected for ensemble evaluation.

F. Results on the CO Dataset

Table I reports the accuracy on the held-out testing days of the CO dataset, and Fig. 2 shows the corresponding cross-day decoding trends.

TABLE I
ACCURACY ON THE HELD-OUT TESTING DAYS OF THE CO DATASET

Method	11/28	11/29	12/02	12/05	12/08	12/09	12/13	12/14	12/15	12/19	12/30	Med.	Mean
vanilla CTM	0.623	0.599	0.607	0.657	0.669	0.765	0.674	0.694	0.422	0.587	0.526	0.628	0.620
LSTM	0.571	0.492	0.517	0.526	0.564	0.604	0.644	0.539	0.383	0.487	0.447	0.520	0.525
TAFF	0.606	0.582	0.596	0.608	0.672	**0.870**	0.651	0.710	0.480	0.648	**0.577**	0.616	0.636
DACTM (Single)	0.648	0.624	0.785	0.735	0.772	0.805	0.723	0.746	**0.522**	0.713	0.536	0.708	0.691
S-DACTM (Single)	0.594	0.631	0.689	0.559	0.667	0.642	0.645	0.508	0.436	0.683	0.460	0.622	0.592
DACTM (Two-Ensemble)	0.654	0.629	0.769	0.729	0.785	0.787	0.720	**0.756**	0.509	**0.716**	0.537	0.717	0.690
S-DACTM (Two-Ensemble)	0.637	**0.679**	0.778	**0.781**	**0.808**	0.816	0.672	0.709	0.511	0.714	0.519	0.702	0.693
DACTM (Three-Ensemble)	**0.668**	0.655	**0.795**	0.767	0.799	0.790	**0.725**	0.739	0.516	0.714	0.541	**0.721**	**0.701**

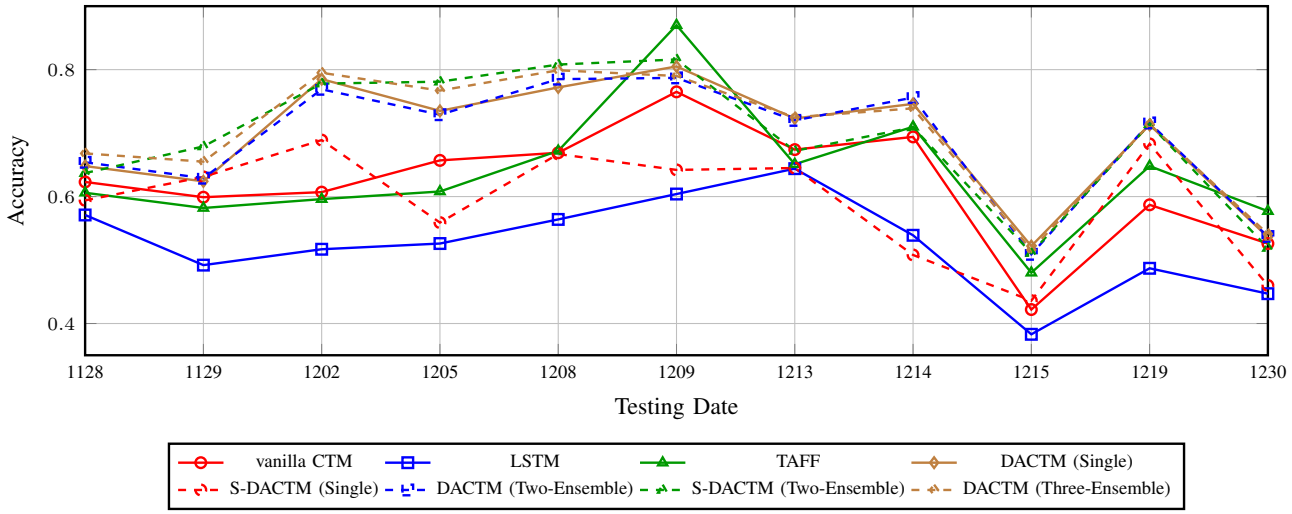

Fig. 2. Performance of cross-day decoding on the CO dataset.

DACTM consistently outperformed the baseline models across most testing days, with particularly clear gains on strongly drifted sessions such as 20221215 and 20221230.

On 20221215, single-model DACTM achieved the highest accuracy (0.5224), significantly outperforming vanilla CTM ($p = 6.23 \times 10^{-6}$), LSTM ($p = 2.8 \times 10^{-4}$) and TAFF ($p = 7.6 \times 10^{-3}$). On 20221230, the three-model DACTM ensemble achieved the best accuracy (0.5407), exceeding both LSTM and vanilla CTM. Across all testing days, the DACTM ensemble obtained the highest median and mean accuracies, indicating improved robustness to cross-day neural drift.

These results show that drift-adaptive temporal modeling improves long-term decoding stability. The weaker performance of vanilla CTM under severe drift further suggests that the original visual-oriented CTM architecture is insufficient for robust intracortical neural decoding without task-specific temporal and drift-adaptive modifications. Ensemble learning generally improved DACTM stability, although single large-scale models occasionally achieved comparable performance, possibly due to validation–test distribution mismatch.

G. Ablation Study on Drift Adaptation

To quantify the contribution of the drift adaptation module, we performed an ablation study by disabling drift adaptation while keeping all other components unchanged. Results were averaged across ten independent runs.

TABLE II
ABLATION STUDY OF THE DRIFT ADAPTATION MODULE ON THE CO DATASET

Model	Median Accuracy	Mean Accuracy
DACTM-B (w/o Drift Adaptation)	0.638	0.636
DACTM-B (Full)	**0.702**	**0.679**
Improvement ↑	10%	6.8%

As shown in Table II and Fig. 3, removing drift adaptation reduced the median accuracy from 0.7019 to 0.6383 and the mean accuracy from 0.6789 to 0.6357. Performance degradation was observed on most testing days, with particularly large drops on 20221205, 20221209, and 20221219.

The largest relative gains were observed on sessions affected by substantial cross-day distribution shifts. For example, accuracy improved from 0.5754 to 0.6857 on 20221219 and from 0.6415 to 0.7160 on 20221205 after enabling drift adaptation. These results indicate that the proposed module effectively captures day-dependent neural statistics and improves robustness to neural drift.

Overall, the ablation study confirms that drift adaptation is a key contributor to the superior cross-day decoding perfor-

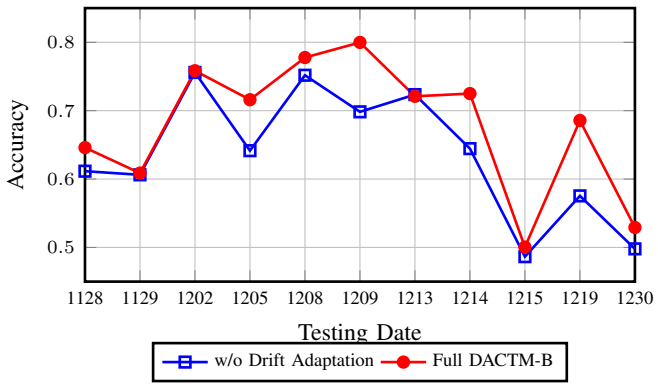

Fig. 3. Ablation study of the drift adaptation module on the CO dataset.

mance of DACTM.

H. Results on the Jango Dataset

Table III reports the accuracy on the held-out testing days of the Jango dataset, and Fig. 4 shows the corresponding cross-day decoding trends.

Both DACTM and S-DACTM outperformed vanilla CTM and LSTM across nearly all testing sessions. Although TAFF achieved competitive accuracy on several individual dates, the proposed models maintained higher median and mean accuracies across the full testing period.

The three-model DACTM ensemble achieved the best overall performance, with a mean accuracy of 0.8378 and a median accuracy of 0.8794. Consistent with the CO results, CTM-based models showed higher median than mean accuracies, suggesting that performance degradation was mainly driven by a small number of severely drifted sessions rather than broad instability across days.

S-DACTM also achieved strong decoding performance. Although single-model S-DACTM was slightly weaker than analog DACTM, likely due to the optimization difficulty of spiking pathways, the two-model spiking ensemble substantially improved performance and approached analog DACTM. Overall, these results demonstrate that the proposed drift-adaptive framework improves cross-day decoding stability and generalization over recurrent and feed-forward baselines.

IV. Discussion

The main advantage of DACTM lies in its explicit treatment of cross-day neural drift during temporal representation learning. By inheriting the iterative synchronization dynamics of CTM, DACTM maintains synchronization-aware representations and models the evolution of contextual neural states during decoding. Combined with drift-adaptive temporal modeling, this design improves robustness to cross-session distributional shifts and helps preserve stable neural dynamics across recording days. Experiments on the CO and Jango datasets show that DACTM achieves higher median and average decoding accuracy than baseline methods, particularly under severe cross-day drift. Compared with standard Transformer architectures with quadratic self-attention complexity, the proposed synchronization-guided attention updates also provide a more efficient strategy for long neural sequence modeling.

S-DACTM further extends DACTM by introducing hybrid analog-spiking computation into temporal encoding and attention projection modules. Although S-DACTM does not consistently outperform analog DACTM in single-model settings, it shows complementary performance in ensemble experiments and offers a biologically plausible extension of the proposed framework. Several limitations remain. More comprehensive ablation studies are needed to quantify the contribution of each component. Spike-path optimization remains challenging because spike discretization may introduce information loss and training instability, and the current evaluation is limited to two intracortical motor datasets. Broader validation across additional subjects, tasks, and public datasets will be needed to assess the generality of the proposed approach.

V. Conclusion

This work presents DACTM, a drift-adaptive Continuous Thought Machine framework for intracortical neural decoding, together with its spiking-enhanced extension, S-DACTM. By integrating multi-scale temporal encoding, drift-adaptive representation fusion, synchronization-guided iterative attention and hybrid analog–spiking computation, the proposed framework improves temporal modeling and cross-day decoding robustness for non-stationary intracortical signals.

Experiments on the CO and Jango intracortical neural datasets show that DACTM and S-DACTM consistently outperform conventional CTM, LSTM and feed-forward temporal baselines in decoding accuracy and cross-session generalization. DACTM achieves robust performance under neural drift, while ensemble experiments further indicate that model diversity improves decoding stability, particularly for spiking-enhanced variants. Repeated runs across multiple random seeds further indicate stable model performance.

References

[1] J. J. Shih, D. J. Krusienski, and J. R. Wolpaw, "Brain-computer interfaces in medicine," *Mayo Clinic Proceedings*, vol. 87, no. 3, pp. 268–279, 2012.

[2] Y. Dong, S. Wang, Q. Huang, R. W. Berg, G. Li, and J. He, "Neural decoding for intracortical brain–computer interfaces," *Cyborg and Bionic Systems*, vol. 4, p. 0044, 2023.

[3] J. Dai, P. Zhang, H. Sun, X. Qiao, Y. Zhao, J. Ma, S. Li, J. Zhou, and C. Wang, "Reliability of motor and sensory neural decoding by threshold crossings for intracortical brain–machine interface," *Journal of neural engineering*, vol. 16, no. 3, p. 036011, 2019.

[4] W. Wu and N. G. Hatsopoulos, "Real-time decoding of nonstationary neural activity in motor cortex," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 16, no. 3, pp. 213–222, 2008.

[5] Y. Song, Q. Zheng, B. Liu, and X. Gao, "Eeg conformer: Convolutional transformer for eeg decoding and visualization," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 31, pp. 710–719, 2023.

[6] F. Liu, S. Meamardoost, R. Gunawan, T. Komiyama, C. Mewes, Y. Zhang, E. Hwang, and L. Wang, "Deep learning for neural decoding in motor cortex," *Journal of Neural Engineering*, vol. 19, no. 5, p. 056021, 2022.

TABLE III
ACCURACY ON THE HELD-OUT TESTING DAYS OF THE JANGO DATASET

Method	08/08	08/09	08/20	08/24	08/25	08/26	08/27	08/28	08/31	09/05	09/06	09/08	10/29	11/02	Med.	Mean
vanilla CTM	0.879	0.714	0.718	0.743	0.620	0.627	0.623	0.645	0.710	0.493	0.551	0.344	0.634	0.551	0.637	0.632
LSTM	0.890	0.689	0.853	0.722	0.532	0.612	0.514	0.633	0.577	0.370	0.575	0.344	0.590	0.565	0.603	0.605
TAFF	**0.982**	0.814	**0.981**	**0.910**	0.738	0.704	0.668	0.744	0.719	0.465	0.665	0.541	0.806	0.626	0.732	0.740
DACTM (Single)	0.962	0.870	0.950	0.898	0.832	0.891	0.897	0.923	0.792	0.619	0.749	0.497	0.875	0.702	0.862	0.818
S-DACTM (Single)	0.955	**0.895**	0.980	0.835	0.751	0.881	0.842	0.894	0.725	0.584	0.700	0.580	0.838	0.650	0.831	0.794
DACTM (Two-Ensemble)	0.973	0.872	0.955	0.902	0.821	0.906	**0.908**	**0.924**	0.794	0.624	0.749	0.566	0.882	0.693	0.871	0.826
S-DACTM (Two-Ensemble)	0.976	0.894	0.974	0.881	0.797	0.902	0.890	0.918	0.789	0.555	0.731	**0.636**	0.870	0.722	0.873	0.824
DACTM (Three-Ensemble)	0.973	0.874	0.943	0.893	**0.838**	**0.919**	0.904	0.905	**0.838**	**0.665**	**0.791**	0.572	**0.888**	**0.727**	**0.879**	**0.838**

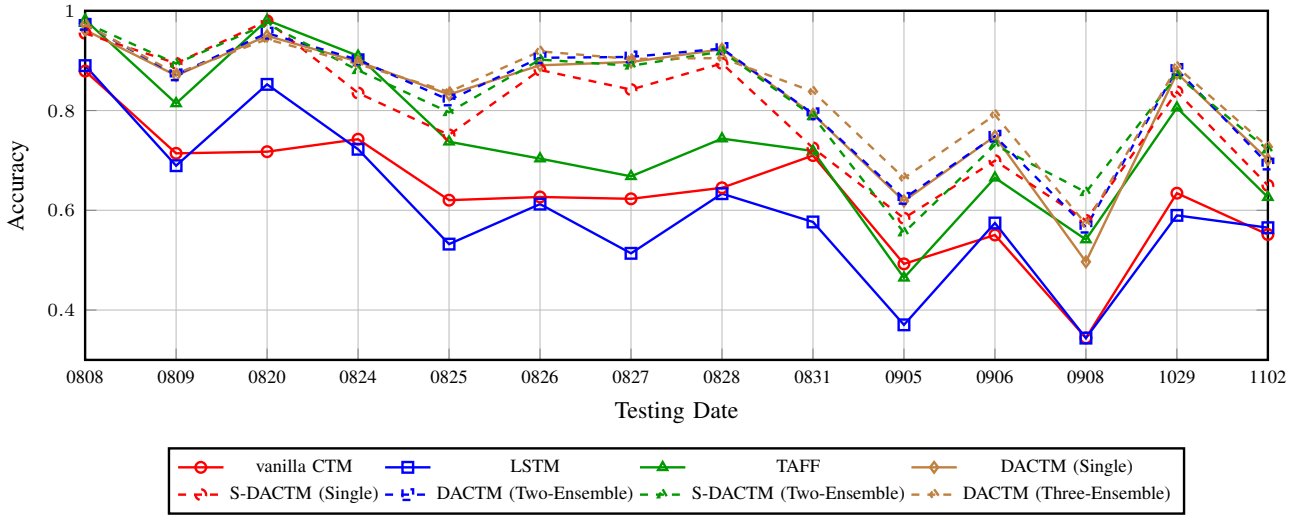

Fig. 4. Performance of cross-day decoding on the Jango dataset.

[7] D. K. Luu, A. T. Nguyen, M. Jiang, J. Xu, M. W. Drealan, J. Cheng, E. W. Keefer, Q. Zhao, and Z. Yang, "Deep learning-based approaches for decoding motor intent from peripheral nerve signals," *Frontiers in Neuroscience*, vol. 15, p. 667907, 2021.

[8] T. Hosman, T. K. Pun, A. Kapitonava, J. D. Simeral, and L. R. Hochberg, "Months-long high-performance fixed lstm decoder for cursor control in human intracortical brain-computer interfaces," in *2023 11th International IEEE/EMBS Conference on Neural Engineering (NER)*, 2023, pp. 1–5.

[9] R. B. Vallabhaneni, P. Sharma, V. Kumar, V. Kulshreshtha, K. J. Reddy, S. S. Kumar, V. S. Kumar, and S. K. Bitra, "Deep learning algorithms in eeg signal decoding application: A review," *IEEE Access*, vol. 9, pp. 125 778–125 786, 2021.

[10] L. Darlow, C. Regan, S. Risi, J. Seely, and L. Jones, "Continuous thought machines," in *Advances in Neural Information Processing Systems*, vol. 38. Curran Associates, Inc., 2025, pp. 21 548–21 594. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2025/file/1f1628d502c62ef3725fb3b0b8eb4219-Paper-Conference.pdf

[11] T. Liu, Y. Chua, Y. Ning, P. Liu, Y. Zhang, T. Li, G. Wan, Z. Wan, W. Chen, and S. Zhang, "motorsrnn: A spiking recurrent neural network inspired by brain topology for the effective and efficient decoding of cortical spike trains," *Biomedical Signal Processing and Control*, vol. 99, p. 106745, 2025.

[12] Q. T. Pham, T. Q. Nguyen, P. C. Hoang, Q. H. Dang, D. M. Nguyen, and H. H. Nguyen, "A review of snn implementation on fpga," in *2021 International Conference on Multimedia Analysis and Pattern Recognition (MAPR)*, 2021, pp. 1–6.

[13] X. Ma, F. Rizzoglio, K. L. Bodkin, E. Perreault, L. E. Miller, and A. Kennedy, "Using adversarial networks to extend brain computer interface decoding accuracy over time," *eLife*, vol. 12, p. e84296, 2023.

[14] W. Wu, M. J. Black, Y. Gao, M. Serruya, A. Shaikhouni, J. P. Donoghue, and E. Bienenstock, "Neural decoding of cursor motion using a kalman filter," in *Advances in Neural Information Processing Systems*, S. Becker, S. Thrun, and K. Obermayer, Eds., vol. 15. MIT Press, 2002, pp. 117–124. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2002/file/169779d3852b32ce8b1a1724dbf5217d-Paper.pdf