A General Framework for Neural Adaptive Sensing of Dynamical Fields

Felix Köster and Atsushi Uchida
Department of Information and Computer Sciences
Saitama University, 255 Shimo-Okubo, Sakura-ku,
Saitama City, Saitama 338–8570, Japan
Email: felixk@mail.saitama-u.ac.jp, auchida@mail.saitama-u.ac.jp

***Abstract*—Physical dynamical systems can be considered as natural information processors, where their trajectories encode past inputs in high-dimensional continuous states. Reservoir computing exploits this property by driving a fixed nonlinear dynamical system and learning only a linear output mapping, however traditional approaches assume fixed measurement locations and weights, which limits robustness across different regimes and substrates. This work introduces a general framework for adaptive measurement of continuous physical dynamics using a partial-differential-equation–based reservoir, where both the spatial positions of the measurements and the output weighting are learned jointly. The method dynamically determines how and where to observe the state of the system and optimizes sensor placement end-to-end. Experiments on forecasting tasks involving multiple chaotic systems demonstrate substantial improvements in accuracy compared to fixed-measurement reservoir computing approaches, indicating that adaptive measurement provides an effective interface for extracting task-relevant information from continuous dynamical systems.**

***Index Terms*—Reservoir computing, attention mechanism, dynamical systems, adaptive measurement, partial differential equations, forecasting**

I. Introduction

Many physical systems can be understood as high-dimensional information reservoirs. When such a system is driven by an input signal, its state trajectory implicitly stores a linear or nonlinear, temporal representation of the past drive [1]. This perspective underlies *reservoir computing* (RC), which uses a fixed nonlinear dynamical system (the reservoir) together with a simple, typically linear readout to solve supervised learning tasks [2]–[4].

Classical RC has been successfully implemented in software and in hardware, for example in photonic, electronic, mechanical, or soft robotic substrates [3], [5]–[9]. The standard setting assumes a fixed input coupling, a fixed measurement interface (e.g., a set of sensors or sampled coordinates), and a linear readout trained by ridge regression. While this yields efficient training and can exploit fast physical dynamics, it also imposes a strong constraint: the measurement operator is not adapted to the task or to changes in the dynamical regime.

Recent work has introduced *Attention–Enhanced Reservoir Computing* (AERC), in which the linear readout is replaced by a small neural network with an attention mechanism [10]. Instead of reading out the reservoir state with fixed weights, the system learns a state-dependent weighting that can emphasize task-relevant coordinates and suppress irrelevant ones. This connects RC to the broader class of attention mechanisms in deep learning [11], [12]. However, the existing AERC formulation typically operates on discrete, finite-dimensional reservoir states and assumes that the set of measured coordinates is fixed.

In many physical implementations, the reservoir is inherently *continuous* in space and time. Examples include optical fields propagating in space, reaction–diffusion media, and partial differential equation based analog computing substrates. In such settings, the choice of measurement locations (e.g., sensor positions) and actuators (e.g., injection points) is itself a crucial design parameter.

The central idea of this work is to treat measurement as a *trainable, differentiable interface* between a continuous dynamical substrate and a task-specific neural readout. We generalize AERC to continuous spatiotemporal reservoirs, formalize the resulting Continuous AERC (CAERC) framework, and demonstrate its effectiveness on forecasting tasks involving multiple attractors. Conceptually, this links RC to adaptive sensing and sensor placement approaches [13], [14], and to learnable spatial transformations used in spatial transformer and deformable convolutional networks [15], [16], however applied here to continuous physical dynamics.

II. Reservoir Computing and Adaptive Measurement

A. Classical Reservoir Computing

In classical RC, a discrete-time input sequence $\{\mathbf{x}_t\}$, with $\mathbf{x}_t \in \mathbb{R}^{D_x}$, drives a nonlinear dynamical system [2]–[4]

$$\mathbf{r}_{t+1} = F(\mathbf{r}_t, \mathbf{x}_t), \tag{1}$$

where $\mathbf{r}_t \in \mathbb{R}^N$ is the reservoir state, and $F : \mathbb{R}^N \times \mathbb{R}^{D_x} \to \mathbb{R}^N$ denotes the (fixed) reservoir dynamics. After a transient, the state trajectory is used as a feature representation of the input. The output is computed by a linear readout,

$$\hat{\mathbf{y}}_t = \mathbf{W}_{\text{out}} \mathbf{r}_t, \tag{2}$$

where $\hat{\mathbf{y}}_t \in \mathbb{R}^{D_y}$ and $\mathbf{W}_{\text{out}} \in \mathbb{R}^{D_y \times N}$ is trained by ridge regression,

$$\mathbf{W}_{\text{out}} = \arg\min_{\mathbf{W}} \|\mathbf{Y} - \mathbf{W}\mathbf{R}\|_2^2 + \lambda \|\mathbf{W}\|_2^2, \tag{3}$$

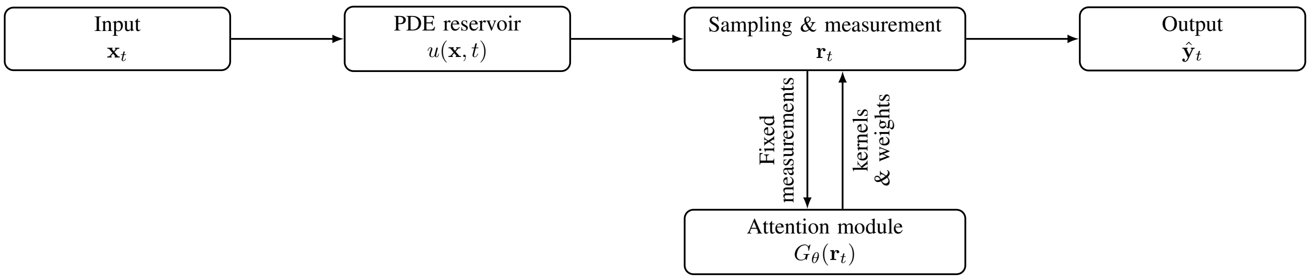

Figure 1. Conceptual architecture of the Continuous Attention–Enhanced Reservoir Computer (CAERC). A discrete input $\mathbf{x}_t$ drives a fixed spatiotemporal PDE reservoir $u(\mathbf{x}, t)$. The reservoir state is sampled via fixed measurement kernels to form $\tilde{\mathbf{r}}_t$, while the trainable attention module $G_\theta(\mathbf{r}_t)$ predicts adaptive kernel parameters (i.e., *where to observe*) and combination weights (i.e., *how to weigh*), which inform improved measurement during prediction.

where $\mathbf{R} \in \mathbb{R}^{N \times T}$ collects reservoir states across T time steps, $\mathbf{Y} \in \mathbb{R}^{D_y \times T}$ are the target outputs, and λ is a regularization parameter.

The reservoir itself can be a recurrent neural network, a delay line, or any physical dynamical system with suitable fading memory properties [3], [5]–[9]. The key assumption is that the reservoir dynamics are rich enough that a linear combination of its state coordinates suffices to approximate the desired mapping from input history to output.

B. Attention–Enhanced Reservoir Computing

AERC generalizes the static readout by introducing a *state-dependent* projection [10]. Instead of fixed weights, the readout uses a small attention network G_θ that maps the current reservoir state to output-specific attention weights,

$$\mathbf{A}_t = G_\theta(\mathbf{r}_t), \tag{4}$$

where we interpret $\mathbf{A}_t \in \mathbb{R}^{D_y \times D_\phi}$ as a collection of D_y row vectors $\mathbf{A}_t^{(k)} \in \mathbb{R}^{D_\phi}$, one for each output dimension $k = 1, \ldots, D_y$, and $G_\theta : \mathbb{R}^N \to \mathbb{R}^{D_y \times D_\phi}$. These are used as dynamic combination coefficients such that

$$\hat{\mathbf{y}}_t = \mathbf{A}_t \, \Phi(\mathbf{r}_t), \tag{5}$$

where $\Phi : \mathbb{R}^N \to \mathbb{R}^{D_\phi}$ is a fixed or learned feature map (e.g., linear, nonlinear, or low-rank), so that $\Phi(\mathbf{r}_t) \in \mathbb{R}^{D_\phi}$ and $\hat{\mathbf{y}}_t \in \mathbb{R}^{D_y}$.

This architecture keeps the reservoir dynamics F fixed while allowing the readout to adapt the effective measurement distribution over the reservoir state. The attention parameters θ are trained by gradient-based optimization, in line with modern attention-based sequence models [11], [12]. Compared to classical RC, AERC can exploit different parts of the state space depending on the current regime and has been shown to improve robustness and predictive performance without modifying the underlying physical substrate [10].

C. PDE-Based Spatiotemporal Reservoir

To move towards continuous reservoirs, we consider a spatiotemporal field $u(\mathbf{x}, t)$ on a domain $\Omega \subset \mathbb{R}^d$, governed by a PDE of the form

$$\partial_t u(\mathbf{x}, t) = \mathcal{L}[u](\mathbf{x}, t) + \mathcal{I}\big(\mathbf{x}; \mathbf{x}_t\big), \tag{6}$$

where $\mathcal{L}$ encodes intrinsic dynamics (e.g., diffusion, advection, reaction) and $\mathcal{I}$ injects the discrete-time input $\mathbf{x}_t$ via spatial actuators. For example, $\mathcal{L}$ may include a diffusion term $\nu \Delta u$ and a nonlinear reaction $g(u)$, and $\mathcal{I}$ may be realized as localized sources or sinks at specified injection points.

In numerical simulations, the PDE is discretized in space and time (e.g., finite differences or finite elements combined with explicit or implicit time-stepping), which yields a high-dimensional state vector $\mathbf{r}_t$ at each discrete time t. In a physical implementation, $u(\mathbf{x}, t)$ is directly realized by a medium such as an optical or chemical system.

D. Continuous Attention–Enhanced Reservoir Computing

In the continuous setting, attention serves a dual function: it learns (i) *where to sense the field* and (ii) *how to combine sensed features to produce the prediction*. To this end, we distinguish two types of measurement functionals.

Fixed measurements (input to attention).: A set of stationary kernels $\{\psi^{(j)}(\mathbf{x})\}_{j=1}^{N_{\text{fix}}}$ evaluates the PDE reservoir state,

$$\tilde{r}_t^{(j)} = \int_\Omega \psi^{(j)}(\mathbf{x}) \, u(\mathbf{x}, t) \, d\mathbf{x}, \qquad \tilde{\mathbf{r}}_t = \big(\tilde{r}_t^{(1)}, \ldots, \tilde{r}_t^{(N_{\text{fix}})}\big)^\top, \tag{7}$$

providing the input feature vector to the attention network.

Adaptive measurements (learned sensors).: Conditioned on $\tilde{\mathbf{r}}_t$, the attention module $G_\theta(\mathbf{r}_t)$ predicts parameters $\theta_t^{(i)}$ of time-varying spatial kernels,

$$\phi_t^{(i)}(\mathbf{x}) = \kappa(\mathbf{x}; \theta_t^{(i)}), \quad i = 1, \ldots, N, \tag{8}$$

which act as dynamic sensors. These produce an adaptive measurement state

$$r_t^{(i)} = \int_\Omega \phi_t^{(i)}(\mathbf{x}) \, u(\mathbf{x}, t) \, d\mathbf{x}, \qquad \mathbf{r}_t = \big(r_t^{(1)}, \ldots, r_t^{(N)}\big)^\top \in \mathbb{R}^N. \tag{9}$$

Adaptive weighting (readout).: The same network outputs output-specific attention weights

$$\mathbf{A}_t = G_\theta(\tilde{\mathbf{r}}_t), \qquad \mathbf{A}_t \in \mathbb{R}^{D_y \times N}, \tag{10}$$

such that each row $\mathbf{A}_t^{(k)} \in \mathbb{R}^N$ defines how the N adaptive sensor values are combined to compute output component k. The final prediction is then

$$\bar{\mathbf{y}}_t = \mathbf{A}_t \, \mathbf{r}_t, \qquad \bar{\mathbf{y}}_t \in \mathbb{R}^{D_y}. \tag{11}$$

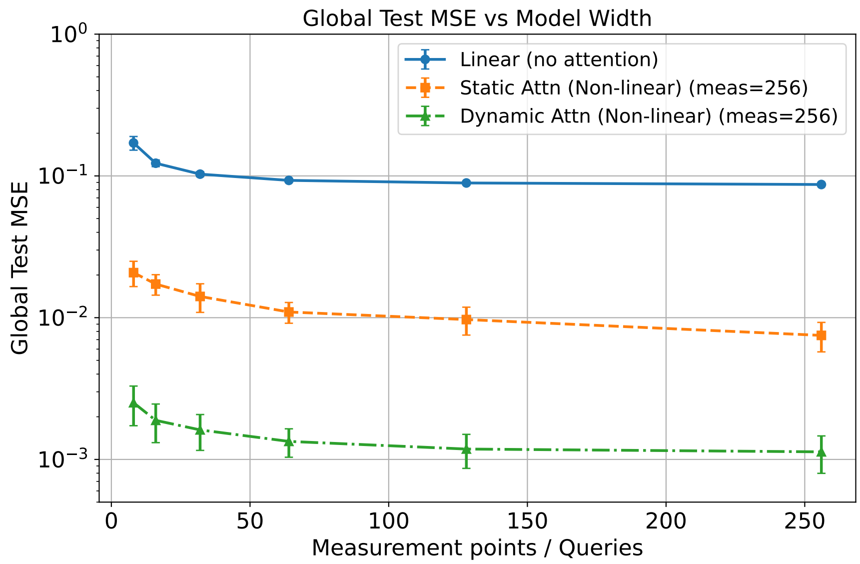

Figure 2. **Prediction error (normalized root-mean-square error, NRMSE) as a function of the number of measurement points.** All methods use the same PDE reservoir. (Blue solid line) Classical reservoir computing with ridge regression; (Orange dashed line) AERC with fixed measurement positions; (Green dash–dotted line) CAERC with adaptive sensing.

Only the attention parameters are updated by gradient descent; the PDE dynamics remain fixed and receive no gradients.

If the spatial kernels $\phi_t^{(i)}$ are held fixed and equal to $\psi^{(j)}$, CAERC reduces to discrete AERC [10]. Replacing the PDE with a discrete update further recovers the standard RC/AERC model.

III. Experimental Setup

We illustrate the CAERC framework by predicting multi-attractor dynamics. We compare three configurations: classical linear RC with fixed measurements, discrete AERC with state-dependent readout weights, and the proposed CAERC with adaptive measurement locations.

In the multi-attractor setting, a single reservoir is driven by several distinct chaotic dynamical systems. Typical examples include the Lorenz, Rössler, Duffing, Mackey–Glass, Hénon, and logistic systems [17]–[21]. Each attractor is simulated under standard parameters and sampled at a fixed timestep, and the resulting state trajectories serve as driver signals for the reservoir.

The task is to forecast the next step of a designated state variable (e.g., $x(t)$ of the Lorenz system) in an open-loop, one-step-ahead fashion. During training, trajectories from all attractors are interleaved so that the reservoir and readout must handle transitions between qualitatively different dynamical regimes. Performance is evaluated in terms of mean squared error (MSE) and normalized error on held-out trajectories.

For the PDE-based reservoir used in this experiment, the input is injected at a fixed set of actuation points, while the measurement strategy differs between the three configurations. In classical RC, a fixed set of measurement points is specified in advance and only a linear readout is trained. In discrete AERC, the same fixed measurements are used, but the readout weights become state dependent through an attention mechanism. In CAERC, both the locations of the measurement points and the attention-based combination of their outputs are learned jointly with the task-specific readout.

IV. Results

The results are shown in Fig. 2, where the normalized root mean square error (NRMSE) over the whole dataset of 8 different dynamical systems is shown. CAERC consistently improves forecasting performance compared to classical RC and discrete AERC for a fixed reservoir and comparable parameter counts.

In the multi-attractor setting, classical RC (blue curve in Fig. 2) with fixed measurements can achieve reasonable one-step-ahead predictions when the reservoir is tuned for a specific attractor. However, its performance degrades when the reservoir must handle multiple attractors with different geometries and time scales. The learned linear readout is forced to compromise across regimes, and no mechanism exists for emphasizing different state components in different regions of the attractor space, thus the normalized root-mean-square error (NRMSE) plateaus at around 0.1.

Discrete AERC (orange curve in Fig. 2) mitigates this limitation by introducing state-dependent weighting over the measured coordinates [10]. Empirically, this reduces prediction error across most attractors and yields more robust performance when switching between regimes. Nevertheless, the set of measured coordinates remains fixed, thus the error roughly is only reduced by one order of magnitude.

CAERC (green curve in Fig. 2) provides an additional degree of optimization by adjusting not only the weights but also the spatial locations of measurements. In the experiments, this leads to two orders-of-magnitude reduction in error over classical RC and one order-of-magnitude improvement over discrete AERC for the same number of measurement points.

Overall, these observations support the view that treating measurement as a trainable interface significantly enhances the expressivity of reservoir computing without altering the underlying physical dynamics. Adaptive measurement can be seen as learning an optimal coordinate system on the manifold of reservoir states, tailored to the task of forecasting.

V. Conclusion

We proposed Continuous Attention–Enhanced Reservoir Computing (CAERC), a general framework for adaptive measurement of continuous physical dynamical systems. By combining a PDE-based reservoir, differentiable measurement kernels with trainable locations, and an attention-based readout, CAERC learns where and how to observe a spatiotemporal field in order to forecast target signals.

Numerical simulations on multi-attractor forecasting demonstrate that CAERC substantially improves prediction accuracy over classical linear RC and discrete AERC, even when the underlying reservoir and the number of trainable parameters are kept comparable. The results suggest that learning sensor placement and measurement structure is a powerful and flexible mechanism for interfacing with physical information-processing substrates.

Future work includes applying CAERC to real physical reservoirs such as photonic, electronic, or mechanical systems [3], [5]–[9], extending the framework to control and inference tasks, and exploring regularization strategies and architectural variants tailored to specific PDEs and hardware constraints.

Acknowledgment

This study was supported in part by JSPS KAKENHI (JP22H05195, JP24KF0179, JP25H01129) and JST CREST (JPMJCR24R2) in Japan.

References

[1] J. Dambre, D. Verstraeten, B. Schrauwen, and S. Massar, "Information processing capacity of dynamical systems," *Scientific Reports*, vol. 2, p. 514, 2012.

[2] H. Jaeger, "The "echo state" approach to analysing and training recurrent neural networks-with an erratum note," *Bonn, Germany: German National Research Center for Information Technology GMD Technical Report*, vol. 148, no. 34, p. 13, 2001.

[3] L. Appeltant, M. C. Soriano, G. Van der Sande, J. Danckaert, S. Massar, J. Dambre, B. Schrauwen, C. R. Mirasso, and I. Fischer, "Information processing using a single dynamical node as complex system," *Nature Communications*, vol. 2, p. 468, 2011.

[4] G. Tanaka, T. Yamane, J. A. He, R. Nakane, S. Kanazawa, H. Takeda, H. Numata, D. Nakano, and A. Fujii, "Recent advances in physical reservoir computing," *Neural Networks*, vol. 115, pp. 100–123, 2019.

[5] J. Torrejon, M. Riou, F. A. Araujo, S. Tsunegi, G. Khalsa, D. Querlioz, P. Bortolotti, V. Cros, K. Yakushiji, A. Fukushima, H. Kubota, S. Yuasa, M. D. Stiles, and J. Grollier, "Neuromorphic computing with nanoscale spintronic oscillators," *Nature*, vol. 547, no. 7664, pp. 428–431, 2017. [Online]. Available: https://doi.org/10.1038/nature23011

[6] F. Köster, S. Yanchuk, and K. Lüdge, "Master memory function for delay-based reservoir computers with single-variable dynamics," *arXiv preprint arXiv:2108.12643*, 2021. [Online]. Available: https://arxiv.org/abs/2108.12643

[7] E. Picco, L. Jaurigue, K. Lüdge, and S. Massar, "Efficient optimisation of physical reservoir computers using only a delayed input," *Communications Engineering*, vol. 3, no. 1, p. e00340, 2025.

[8] K. Nakajima, H. Hauser, T. Li, and R. Pfeifer, "Information processing via physical soft body," *Scientific Reports*, vol. 5, May 2015.

[9] T. Kotooka, S. Lilak, A. Stieg, J. Gimzewski, N. Sugiyama, Y. Tanaka, H. Tamukoh, Y. Usami, and H. Tanaka, "Ag2se nanowire network as an effective in-materio reservoir computing device," 2021, preprint. [Online]. Available: https://www.researchsquare.com/article/rs-322405/v1

[10] F. Köster, K. Kanno, J. Ohkubo, and A. Uchida, "Attention-enhanced reservoir computing," *Phys. Rev. Appl.*, vol. 22, p. 014039, Jul 2024. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevApplied.22.014039

[11] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in Neural Information Processing Systems*, 2017, pp. 5998–6008.

[12] S. Bai, J. Z. Kolter, and V. Koltun, "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling," *arXiv preprint arXiv:1803.01271*, 2018.

[13] K. Manohar, B. W. Brunton, J. N. Kutz, and S. L. Brunton, "Data-driven sparse sensor placement for reconstruction," *arXiv preprint arXiv:1701.07569*, 2017. [Online]. Available: https://arxiv.org/abs/1701.07569

[14] A. Krause, A. Singh, and C. Guestrin, "Near-optimal sensor placements in gaussian processes: Theory, efficient algorithms and empirical studies," *Journal of Machine Learning Research*, vol. 9, pp. 235–284, 2008.

[15] M. Jaderberg, K. Simonyan, A. Zisserman, and K. Kavukcuoglu, "Spatial transformer networks," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2015.

[16] J. Dai, H. Qi, Y. Xiong, Y. Li, G. Zhang, H. Hu, and Y. Wei, "Deformable convolutional networks," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 764–773.

[17] E. N. Lorenz, "Deterministic nonperiodic flow," *Journal of Atmospheric Sciences*, vol. 20, no. 2, pp. 130–141, 1963.

[18] O. E. Rössler, "An equation for continuous chaos," *Physics Letters A*, vol. 57, no. 5, pp. 397–398, 1976.

[19] G. Duffing, *Forced Oscillations, with Special Reference to Mechanical Systems: Collected Papers*. Vieweg & Sohn, Braunschweig, 1918.

[20] M. Henon, "A two-dimensional mapping with a strange attractor," *Communications in Mathematical Physics*, vol. 50, no. 1, pp. 69–77, 1976.

[21] M. C. Mackey and L. Glass, "Oscillation and chaos in physiological control systems," *Science*, vol. 197, no. 4300, pp. 287–289, 1977.