Applied Physics Express

Inverted input method for computing performance enhancement of the ion-gating reservoir

OPEN ACCESS

Applied Physics Express 17, 024501 (2024) LETTER

https://doi.org/10.35848/1882-0786/ad2906

Inverted input method for computing performance enhancement of the ion-gating reservoir

Yu Yamaguchi[1,2], Daiki Nishioka[1,2], Wataru Namiki[1], Takashi Tsuchiya[1*], Masataka Imura[3], Yasuo Koide[4], Tohru Higuchi[2], and Kazuya Terabe[1]

[1]Research Center for Materials Nanoarchitectonics (MANA), National Institute for Materials Science (NIMS), 1-1 Namiki, Tsukuba, Ibaraki, 305-0044, Japan
[2]Department of Applied Physics, Faculty of Science, Tokyo University of Science, 6-3-1 Niijuku, Katsushika, Tokyo, 125-8585, Japan
[3]Research Center for Functional Materials, NIMS, 1-1 Namiki, Tsukuba, Ibaraki, 305-0044, Japan
[4]Research Center and Facility Services Division, NIMS, 1-2-1 Sengen, Tsukuba, Ibaraki, 305-0047, Japan

*E-mail: TSUCHIYA.Takashi@nims.go.jp

Received January 21, 2024; accepted February 12, 2024; published online February 27, 2024

Physical reservoir computing (PRC) is useful for edge computing, although the challenge is to improve computational performance. In this study, we developed an inverted input method, the inverted input is additionally applied to a physical reservoir together with the original input, to improve the performance of the ion-gating reservoir. The error in the second-order nonlinear equation task was 7.3×10^{-5}, the lowest error in reported PRC to date. Improvement of high dimensionality by the method was confirmed to be the origin of the performance enhancement. This inverted input method is versatile enough to enhance the performance of any other PRC. © 2024 The Author(s). Published on behalf of The Japan Society of Applied Physics by IOP Publishing Ltd

Supplementary material for this article is available online

Among various state-of-art neuromorphic computing approaches,[1–6] physical reservoir computing (PRC) is a particularly promising computation scheme to reduce energy consumption in AI-based information processing.[1] In this scheme, materials and devices are utilized as physical reservoirs to efficiently process information, by mapping input to a higher-dimensional feature space using their own intrinsic nonlinearity.[1] Whereas PRC using various materials and devices has been reported so far, the performance is not sufficient.[1,3–23] Although nonlinearity, high dimensionality, and short-term memory (STM) are known to be required to be improved for high-performance PRC,[1] a universal methodology to improve them has not yet been established. In particular, improving high dimensionality is difficult in PRC because of limitations in the structure of PRC devices. A masking process improves high dimensionality by adding a certain amount of modulation to the desired signal and inputting it into the physical reservoir to maintain its transient state.[20,24] However, this has its drawbacks, such as requiring complex processing with adjustment of hyperparameters and making the data length of input very long. This causes an urge for the development of different approaches for the improvement of high dimensionality. Herein, we demonstrate an inverted input method as a new approach for high dimensionality improvement and evaluate its effect on computational performance. An inverted input method is a simple method in which inverted input is generated by an inversion (linear transformation) of the original input and additionally applied to PRC with original input, leading to far lower processing costs for maintaining transient states of the reservoir than general masking methods. An "ion-gating reservoir (IGR)" which uses the nonlinear response of transistors with ions,[20–23] was employed as a model case to evaluate the effect of using inverted inputs on the prediction error of a second-order nonlinear equation task. Furthermore, to investigate the cause of the error reduction by the inverting input, we quantitatively evaluated high dimensionality and STM, which are requirements for PRC.[1,25]

A schematic diagram of the IGR is shown in Fig. 1(a). This IGR is based on an electric double-layer (EDL) transistor[22,23,26,27] fabricated with a Li-Si-Zr-O (LSZO) lithium ion-conducting amorphous electrolyte thin film, which is deposited by pulsed laser deposition, and a hydrogen-terminated diamond homoepitaxial (100) substrate, which is deposited by microwave-plasma chemical vapor deposition. Electronic carrier density on the diamond channel's surface is tuned by EDL formation due to Li^+ transport driven by applied gate voltage (V_g).[22,23,26–30] As shown in Fig. 1(a), the drain current (I_d) is largely modulated by the V_g application. Please refer to the Supplementary Material for details of the device's fabrication and electrical measurements.

A general scheme of PRC is shown in Fig. 1(b). In the input layer, time series data $u(k)$ are inputted, where k is the time step. Then, in the reservoir layer, input time series data is transformed nonlinearly to high dimensional space as reservoir states $X_i(k)$ at given node i ($i = 1, 2, \ldots, N$).[1,31] For the output layer, the readout weight w_i connecting $X_i(k)$ and output $y_i(k)$ is trained by linear regression to obtain the desired output. The reservoir output $y(k)$ is described as a linear combination of $X_i(k)$ and the w_i, as follows,

$$y(k) = \sum_{i=1}^{N} w_i X_i(k) + b \quad (1)$$

where N and b are the number of the reservoir state and bias, respectively.

As shown in Fig. 1(c), we prepared the input $u(k)$ as a random value ($0.0 \leqq u(k) \leqq 0.5$) and inverted input $u'(k)$ ($u'(k) = 0.5\text{-}u(k)$). By combining both input $u(k)$ and $u'(k)$, the inverted input method enhances high dimensionality in PRC. The principle is explained below. The function of the reservoir in reservoir computing is to map the input to a

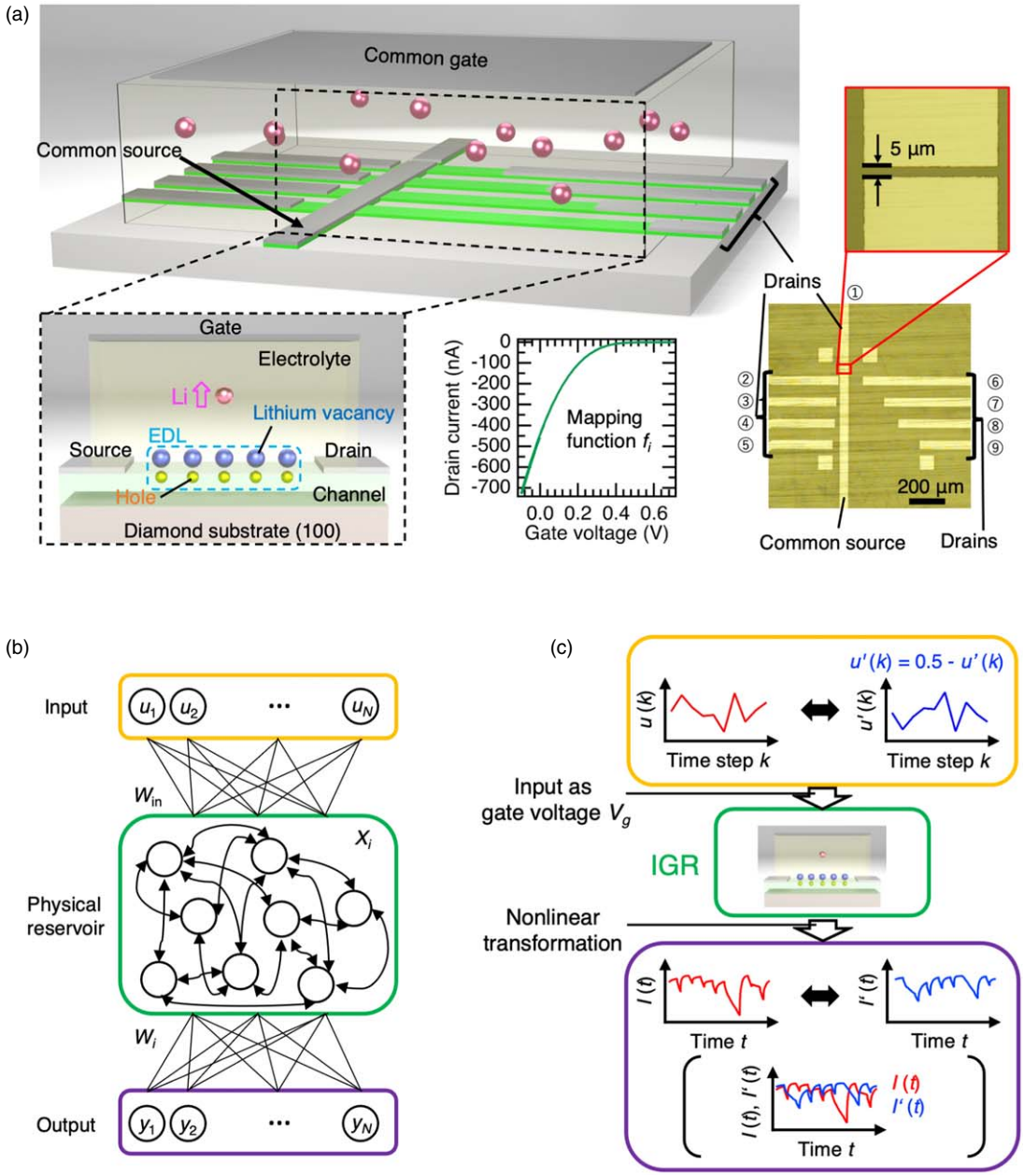

Fig. 1. (a) Schematic illustration of electric double layer effect-based IGR. (b) General scheme of PRC. (c) Basic role of inverted input in PRC.

high-dimensional feature space.[1,31] The reservoir's mapping function f_i for the input u_i and reservoir state X_i is shown in Fig. 1(a). In a simulation reservoir such as an echo state network, nonlinear functions such as tanh are employed identically for all nodes (i.e., $f_i = f_{j\neq i}$).[25,31] On the other hand, in PRC, a variety of nonlinear functions that originate from nonlinear dynamics that are inherent in the physical system are employed as mapping functions for each node (i.e., for each measurement terminal or sampling time/virtual node).[1,3–23] These mapping functions change dynamically according to the driving conditions (e.g., input intensity and time scale) of the physical system (i.e., $f_i \neq f_{j\neq i}$). In this IGR, the I_d-V_g and gate current (I_g)-V_g characteristic were utilized as mapping functions for the nonlinear transformation to high dimensional space. The relationship between the reservoir state and the mapping function is defined by the following equation.

$$X_i(k+1) = f_i[u(k+1), X_i(k)] \quad (2)$$

As shown in Fig. 1(c), the relation between $u(k)$ and $u'(k)$ is linear. Yet, the nonlinearity of the f_i causes them to be different outputs from each other (i.e., two outputs are not in a linear relationship), allowing them to map to a higher dimensional feature space than is possible with a single $u(k)$. This is the mechanism of higher performance with inverted inputs, which is a particularly effective scheme for physical systems with diverse and dynamic mapping functions, compared to simulation reservoirs with uniform mapping functions. In addition, the effective number of nodes in the reservoir increases with the improvement in high dimensionality; therefore, STM can also be improved.[1,25]

In this present study, the effect of the inverted input has been investigated with a second-order nonlinear equation task, which is a typical benchmark task of PRCs.[16–22] The target output $y_t(k)$ for the task is determined from the

equation below.

$$y_t(k) = 0.4y_t(k-1) + 0.4y_t(k-1)y_t(k-2) + 0.6u^3(k) + 0.1 \quad (3)$$

For the input data, input $u(k)$ and inverted input $u'(k)$ were prepared. As referenced in Fig. S2, $u(k)$ and $u'(k)$ were converted to a voltage pulse stream with a pulse period T (10–80 ms) and duty rate D (40%–80%). The constant drain voltage V_d was set to 0.1 V. Voltage pulse stream was applied to the common gate and the output current was measured from 9 terminals of drains (i.e., 9 drain current) which have different channel lengths and 1 terminal of the gate (i.e., 1 gate current), in a total of 10 physical nodes. Under conditions without $u'(k)$, 20 virtual nodes were extracted with even spacing from each current response from 10 terminals. Therefore, 200 virtual nodes (10 physical nodes × 20 virtual nodes) were utilized in total for the task. Under condition with $u'(k)$, 10 virtual nodes were extracted with even spacing from each current response from 10 terminals. Thus, 200 virtual nodes (2 ways of input × 10 physical nodes × 10 virtual nodes) were utilized similarly in total for the task. By utilizing both physical and virtual nodes, unique diverse reservoir states can be extracted, which enables the higher dimensionality necessary for high performance.

To evaluate the performance of the task, we calculated the normalized mean squared error (NMSE) of the $y_t(k)$ from Eq. (3) and the prediction output $y(k)$ trained by Eq. (1), as follows.

$$\mathrm{NMSE} = \frac{\sum_{k=1}^{L}\{y_t(k) - y(k)\}^2}{\sum_{k=1}^{L} y_t(k)^2} \quad (4)$$

where $L(=500)$ is a data length.

Figures 2(a) and 2(b) show $y_t(k)$ and $y(k)$ of the second order nonlinear equation task without $u'(k)$ and with $u'(k)$ when the NMSE was at a minimum under the best conditions of T and D (i.e., $T = 70$ ms and $D = 70\%$), which suggest the long T and large D enhance at least one of the three properties required for PRC, leading to precise prediction performance in the task. Please refer to the Supplementary Material for details of T and D dependence on the performance. By combining $u(k)$ and $u'(k)$, it was possible to predict more precisely which achieved the NMSE of 7.3×10^{-5}, whereas it was 1.4×10^{-4} under conditions without $u'(k)$. Moreover, the NMSE under condition with $u'(k)$ was the smallest NMSE in the second-order nonlinear equation task among reported PRC.[16–22)]

In order to clarify the origin of such performance enhancement of the IGR due to the introduction of the inverted input method, we quantitatively analyzed the high dimensionality and STM of the system under the two conditions (i.e., input without $u'(k)$ and with $u'(k)$). As discussed above, introducing an inverted input provides diverse reservoir states due to the nonlinearity and asymmetry of the mapping function. Figures 3(a) and 3(b) show all reservoir states used for the task under conditions without $u'(k)$ and with $u'(k)$ respectively. The system diversity is clearly enhanced by adding the reservoir state $X'(k)$ for the inverted input $u'(k)$ as shown in Fig. 3(b) to the reservoir state $X(k)$ for the original input $u(k)$ as shown in Fig. 3(a). In particular, $X(k)$ and $X'(k)$ are not inversely symmetric due to

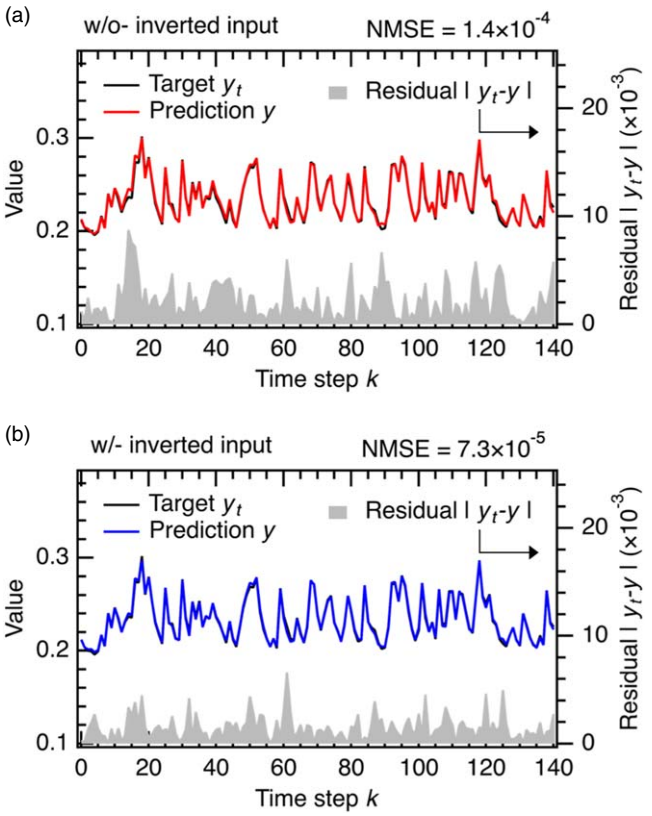

Fig. 2. The target and prediction waveforms under conditions without $u'(k)$ (a) and with $u'(k)$ (b) in pulse period of 70 ms and duty rate of 70%.

the nonlinearities of the mapping functions, despite $u(k)$ and $u'(k)$ being inversely symmetric. Therefore, the inverted input surely improved the high dimensionality of this system. The high dimensionality in PRC is achieved by the presence of multiple nodes that behave independently, which can be characterized by a correlation coefficient r between each node X_i and $X_{j\neq i}$ defined by following equation[20,21)]

$$r(X_i, X_j) = \frac{\sum_{k=1}^{L=150}(X_i(k) - \bar{X}_i)\left(X_j(k) - \bar{X}_j\right)}{\sqrt{\sum_{k=1}^{L=150}(X_i(k) - \bar{X}_i)^2}\sqrt{\sum_{k=1}^{L=150}\left(X_j(k) - \bar{X}_j\right)^2}} \quad (5)$$

where $\bar{X}_i$ is the average value of X_i. Please refer to Supplementary Material for the details of the correlation coefficient analysis.

Figure 3(c) shows the correlation coefficients between each node under the conditions with $u(k)$ and $u'(k)$ combined. $X_1 \sim X_{100}$ are nodes under the condition with $u(k)$ and $X'_1 \sim X'_{100}$ are nodes under the condition with $u'(k)$. In addition, $X_{90} \sim X_{100}$ and $X'_{90} \sim X'_{100}$ were extracted from the I_g response, and all other reservoir states were extracted from the I_d response, resulting in lower correlation coefficients between reservoir states extracted from I_g and I_d. Overall, the correlation coefficients between $X_1 \sim X_{100}$ and $X'_1 \sim X'_{100}$ were low compared to pairs of both $X_1 \sim X_{100}$ and both $X_1 \sim X_{100}$, which suggests that $X(k)$ from $u(k)$ and $X'(k)$ from $u'(k)$ are not linear relationships due to the nonlinear transformation by mapping function of the IGR. Therefore, the inverted input method effectively boosts PRC's high-dimensionality, which maps the input data to high-dimensional feature space based on the nonlinear dynamics inherent in the physical system.[20,21)]

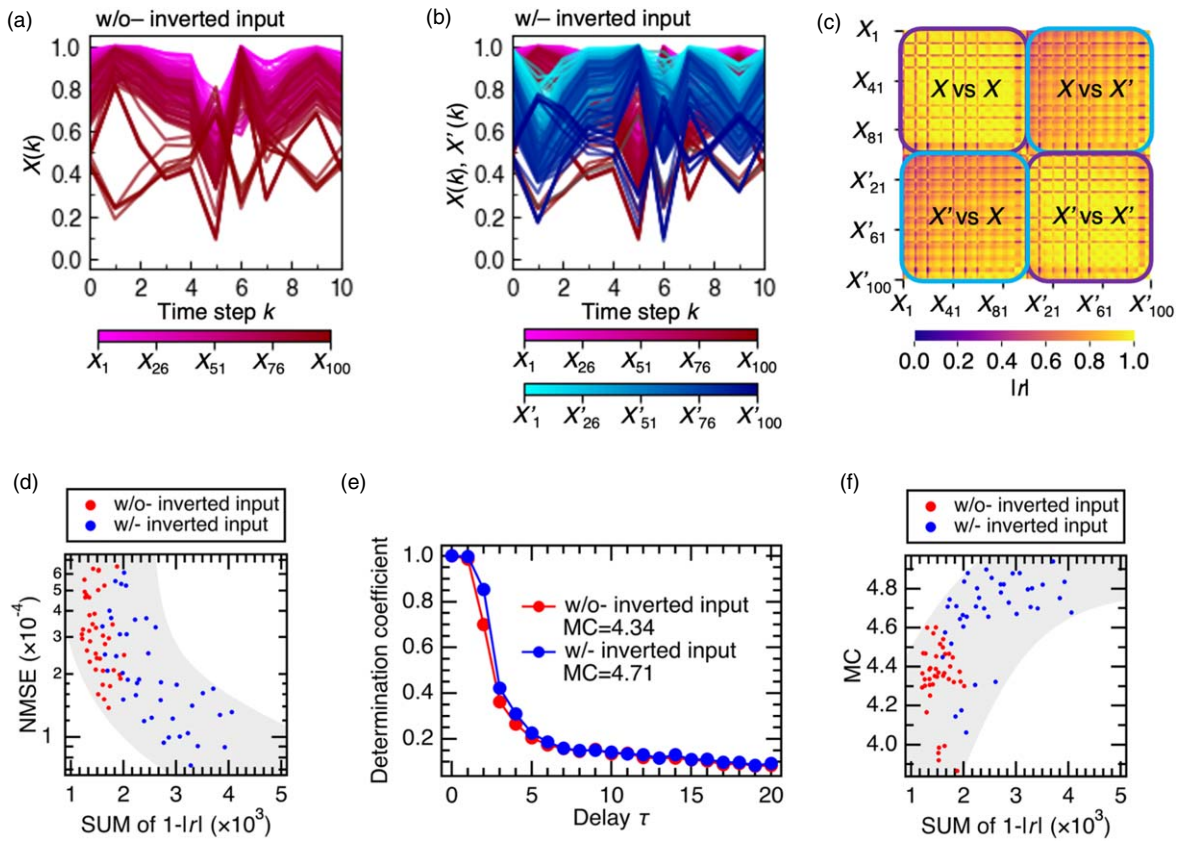

Fig. 3. (a) All reservoir states X under conditions without $u'(k)$ (a) and with $u'(k)$ (b), (c) The heatmap of r for 200 nodes consists of $u(k)$ and $u'(k)$ at $T = 70$ ms and $D = 70\%$. $X_1 \sim X_{100}$ corresponds to condition without $u'(k)$. $X'_1 \sim X'_{100}$ corresponds to condition with $u'(k)$ (d) The relation between NMSE of second order nonlinear equation task and SUM of 1-|r| (e) The forgetting curve under both conditions without $u'(k)$ and with $u'(k)$ at $T = 70$ ms and $D = 70\%$ (f) The relation between MC and SUM of 1-|r|.

High dimensionality is also dependent on the diversity of each node and the number of nodes.[1)] To evaluate the overall high dimensionality of the system, the sum of 1-|r| is rather suitable for quantification.[20,21)] Figure 3(d) shows the relationship between NMSE of the second-order nonlinear equation task under all T and D conditions and the sum of 1-|r| under both conditions without $u'(k)$ (red dot) and with $u'(k)$ (blue dot). The sums of 1-|r| tended to be larger with $u'(k)$ than without $u'(k)$, meaning that the overall high dimensionality is enhanced with $u'(k)$. More importantly, there appears to be a clear trend where NMSE is low when the sum of 1-|r| is high. Therefore, it is evidenced that utilizing the inverted input method improves the high dimensionality, leading to high performance.

Moreover, improvement of the high dimensionality can further increase STM, which can be evaluated with a delay task in which the reservoir reconstructs historical time series data.[1)] The input of this task is the same random input as used in the second-order nonlinear equation task, and the $u(k\text{-}\tau)$ before the delay time τ was reconstructed by a linear combination of $X(k)$ and w obtained from the current response. The difference between the target waveform $u(k\text{-}\tau)$ and the reconstructed waveform $y(k)$ was evaluated by the following equation of determination coefficient r^2

$$r^2(\tau) = \frac{\mathrm{Cov}^2(u(k-\tau), y(k))}{\mathrm{Var}(u(k) \times \mathrm{Var}(y(k)))} \tag{6}$$

where Cov() and Var() are covariance and variance respectively. Figure 3(e) shows the forgetting curve, the τ dependence of r^2. The ability to reconstruct past data, represented by r^2, decreased with increasing τ. Memory capacities (MCs) were calculated by integration of the forgetting curves described in the following equation

$$\mathrm{MC} = \sum_{\tau=1}^{\infty} r^2(\tau) \tag{7}$$

MCs were calculated as 4.34 without $u'(k)$ and 4.71 with $u'(k)$. The theoretical limit of the MC is the same as the reservoir size (the number of nodes), but generally, in physical systems, the MC is much lower than the number of nodes.[25)] That is because, in a physical system, there are many nodes that behave similarly, which causes a lower effective reservoir size than a number of nodes.[1)] Thus, it is essential to improve high dimensionality, which corresponds to an effective reservoir size to increase the MC. To evaluate overall STM, we analyzed the relationship between the MC and the sum of 1-|r|, as shown in Fig. 3(f). Under the condition with $u'(k)$, both the sum of 1-|r| and the MC tend to be large. Conversely, under the condition without $u'(k)$, the sum of 1-|r| and the MC are small. Therefore, it is suggested that the MC increased due to the improvement in high dimensionality despite using the same number of nodes. Introducing the inverted input significantly enhances the performance of this IGR due to the improvement in high dimensionality and STM.

In conclusion, the inverted input method was applied to EDL-based IGR for PRC performance improvement. The NMSE in a second-order nonlinear equation task was

significantly reduced from 1.4×10^{-4} to 7.3×10^{-5}, confirming the effectiveness of the additional inverted input and achieving the best performance of any physical reservoir reported to date. High dimensionality was evaluated by calculating the correlation coefficient r, r between each node notably decreased by applying the additional inverted input. Higher dimensionality due to additional inverted input was the main origin of performance improvement of the reservoir by analyzing the relationship between NMSE and the sum of 1-|r|. Also, the MC increased from 4.34 to 4.71 by applying the inverted input method. The MC increase is accompanied by a decrease in r. This inverted input method is versatile enough to be applied to any physical reservoir and may improve performance. Compared to masking, this method is easy to use and improves computational performance at a low computational cost without hyperparameters, because it can generate diverse reservoir states with a simple process and avoid increasing the input data length.

Acknowledgments This work was supported in part by the Japan Society for the Promotion of Science (JSPS) KAKENHI Grant No., JP22H04625 (Grant-in-Aid for Scientific Research on Innovative Areas "Interface Ionics") and JP22KJ2799 (Grant-in-Aid for JSPS Fellows), and JST PRESTO (Grant No., JPMJPR23H4). A part of this work was supported by the Iketani Science and Technology Foundation and JFE 21st century foundation. A part of this work was supported by "Advanced Research Infrastructure for Materials and Nanotechnology in Japan (ARIM)" of the Ministry of Education, Culture, Sports, Science and Technology (MEXT). Proposal Number JPMXP1223NM5072.

ORCID iDs Wataru Namiki https://orcid.org/0000-0003-4053-7366
Takashi Tsuchiya https://orcid.org/0000-0002-6950-6160
Kazuya Terabe https://orcid.org/0000-0003-3988-3456

1) G. Tanaka, T. Yamane, J. B. Heroux, R. Nakane, N. Kanazawa, S. Takeda, H. Numata, D. Nakano, and A. Hirose, A Review. Neural Netw. **115**, 100 (2019).
2) N. Hagiwara, T. Asai, K. Ando, and M. Akai-Kasaya, Adv. Funct. Mater. **33**, 2300903 (2023).
3) S. Kan, K. Nakajima, Y. Takeshima, T. Asai, Y. Kuwahara, and M. Akai-Kasaya, Phys. Rev. Appl. **15**, 024030 (2021).
4) S. Kan, K. Nakajima, T. Asai, and M. Akai-Kasaya, Adv. Sci. **9**, 2104076 (2022).
5) M. Akai-Kasaya, Y. Takeshima, S. Kan, K. Nakajima, T. Oya, and T. Asai, Neuromorph. Comput. Eng. **2**, 014003 (2022).
6) Y. Usami et al., Adv. Mater. **33**, 2102688 (2021).
7) R. Nakane, G. Tanaka, and A. Hirose, IEEE Access. **6**, 4462 (2018).
8) J. Torrejon et al., Nature **547**, 428 (2017).
9) G. Van der Sande, D. Brunner, and M. C. Soriano, Nanophotonics **6**, 561 (2017).
10) M. Nakajima, K. Inoue, K. Tanaka, Y. Kuniyoshi, T. Hashimoto, and K. Nakajima, Nat. Commun. **13**, 7847 (2022).
11) K. Nakajima, H. Hauser, T. Li, and R. Pfeifer, Sci. Rep. **5**, 10487 (2015).
12) Y. Zhong, J. Tang, X. Li, B. Gao, H. Qian, and H. Wu, Nat. Commun. **12**, 408 (2021).
13) G. Milano, G. Pedretti, K. Montano, S. Ricci, S. Hashemkhani, L. Boarino, D. Lelmini, and C. Ricciardi, Nat. Matter. **21**, 195 (2022).
14) T. Tsuchiya, T. Nakayama, and K. Ariga, Appl. Phys. Express **15**, 100101 (2022).
15) D. Banerjee, S. Azhari, Y. Usami, and H. Tanaka, Appl. Phys. Express **14**, 105003 (2021).
16) D. Nishioka, Y. Shingaya, T. Tsuchiya, T. Higuchi, and K. Terabe, Sci. Adv. (2024).
17) W. Jiang, L. Chen, K. Zhou, L. Li, Q. Fu, Y. Du, and R. H. Liu, Appl. Phys. Lett. **115**, 192403 (2019).
18) C. Du, F. Cai, M. A. Zidan, W. Ma, S. H. Lee, and W. D. Lu, Nat. Commun. **8**, 2204 (2017).
19) W. Namiki, D. Nishioka, Y. Yamaguchi, T. Tsuchiya, T. Higuchi, and K. Terabe, Adv. Intell. Syst. **5**, 2300228 (2023).
20) K. Shibata, D. Nishioka, W. Namiki, T. Tsuchiya, T. Higuchi, and K. Terabe, Sci. Rep. **13**, 21060 (2023).
21) T. Wada, D. Nishioka, W. Namiki, T. Tsuchiya, T. Higuchi, and K. Terabe, Adv. Intell. Syst. **5**, 2300123 (2023).
22) D. Nishioka, T. Tsuchiya, W. Namiki, M. Takayanagi, M. Imura, Y. Koide, T. Higuchi, and K. Terabe, Sci. Adv. **8**, eade1156 (2022).
23) M. Takayanagi, D. Nishioka, T. Tsuchiya, M. Imura, Y. Koide, T. Higuchi, and K. Terabe, Mater. Today Adv. **18**, 100393 (2023).
24) L. Appeltant, G. Van der Sande, J. Danckaert, and I. Fischer, Sci. Rep. **4**, 3629 (2014).
25) M. Inubushi and K. Yoshimura, Sci. Rep. **7**, 10199 (2017).
26) T. Tsuchiya, K. Terabe, and M. Aono, Appl. Phys. Lett. **103**, 073110 (2013).
27) T. Tsuchiya, K. Terabe, and M. Aono, Adv. Mater. **26**, 1087 (2014).
28) M. Takayanagi, T. Tsuchiya, D. Nishioka, M. Imura, Y. Koide, T. Higuchi, and K. Terabe, Mater. Today Phys. **31**, 101006 (2023).
29) T. Tsuchiya, M. Takayanagi, K. Mitsuishi, M. Imura, S. Ueda, Y. Koide, T. Higuchi, and K. Terabe, Commun. Chem. **4**, 117 (2021).
30) K. Terabe, T. Tsuchiya, and T. Tsuruoka, Adv. Electron. Mater. **8**, 2100645 (2022).
31) T. Akiyama and G. Tanaka, IEEE Access. **10**, 28535 (2022).