A Redox-Based Ion-Gating Reservoir, Utilizing Double Reservoir States in Drain and Gate Nonlinear Responses

Tomoki Wada, Daiki Nishioka, Wataru Namiki, Takashi Tsuchiya,* Tohru Higuchi, and Kazuya Terabe

Herein, physical reservoir computing with a redox-based ion-gating reservoir (redox-IGR) comprising Li$_x$WO$_3$ thin film and lithium-ion conducting glass ceramic (LICGC) is demonstrated. The subject redox-IGR successfully solves a second-order nonlinear dynamic equation by utilizing voltage pulse driven ion-gating in a Li$_x$WO$_3$ channel to enable reservoir computing. Under the normal conditions, in which only the drain current ($I_D$) is used for the reservoir states, the lowest prediction error is $8.15 \times 10^{-4}$. Performance is enhanced by the addition of $I_G$ to the reservoir states, resulting in a significant lowering of the prediction error to $5.39 \times 10^{-4}$, which is noticeably lower than other types of physical reservoirs (memristors and spin torque oscillators) reported to date. A second-order nonlinear autoregressive moving average (NARMA2) task, a typical benchmark of reservoir computing, is also performed with the IGR and good performance is achieved, with a normalized mean square error (NMSE) of 0.163. A short-term memory task is performed to investigate an enhancement mechanism resulting from the $I_G$ addition. An increase in memory capacity, from 2.35 without $I_G$ to 3.57 with $I_G$, is observed in the forgetting curves, indicating that enhancement of both high dimensionality and memory capacity is attributed to the origin of the performance improvement.

1. Introduction

Recently, the research on and development of physical reservoir computing has seen increased activity due to the tremendous potential for it to significantly reducing computation resources compared to conventional machine learning approaches, which are based solely on semiconductor integrated circuits.$^{[1,2]}$ Various materials and devices, including soft bodies, optical devices, analogue circuits, spin torque oscillators, memristors, nanowire networks, and ion-gating transistors, have been reported to function as physical reservoirs that require nonlinearity, high dimensionality, and short-term memory.$^{[2-26]}$ while the demonstrated computing performance has been far from satisfactory to date. One common characteristic of a physical reservoir that is proving to be extremely difficult to achieve is the securing of high dimensionality, which is in essence the obtaining of a sufficient number of reservoir states from the output of a physical reservoir. This is because the outputs of physical reservoirs are measured as small numbers of time-series responses with a limited number of detecting probes (e.g., electrodes, sensors), which are attached to or arranged in some manner with the reservoir under serious geometrical constraints. This is in direct contrast to fully simulated reservoirs, in which unrestricted access to the reservoir states of nodes is enabled. Virtual node methods are useful in compensating for said lack of high dimensionality, and are thus widely used.$^{[7-10]}$ In such method, postprocessing allows for a lot of virtual nodes to be obtained from given time-series data. However, there is a known trade-off relationship between increasing the number of virtual nodes and the diversity of each virtual node.$^{[27]}$ It is thus not straightforward to secure a sufficient number of diverse virtual nodes from a limited number of time-series response, which makes physical reservoirs impracticable. Therefore, to achieve practical use, it is necessary to explore physical reservoirs that have diverse outputs.

Here, we report a redox-based ion-gating reservoir (redox-IGR) composed of all-solid-state redox transistors$^{[28-41]}$ which can derive double reservoir states from drain and gate current response, based on ion insertion and desertion (redox) through a solid electrolyte$^{[42-46]}$ at a Li$^+$–electron mixed conductor, Li$_x$WO$_3$. The redox mechanism, as well as the electric double layer mechanism, is useful for the conductance modulation of semiconductor channels.$^{[47-53]}$ Using sequential gate voltage pulse trains, a drain current (electronic current) flows through a Li$_x$WO$_3$ thin film channel, where it is modified by a redox...
reaction with a Li$^+$-ion conducting glass ceramic (LICGC) substrate through the modulation of the conducting electron density, so as to generate a nonlinear time-series response in the drain current. Simultaneously, a relatively large gate current for the redox process (lithium-ion current) can provide another time-series response. In the normal measurement configuration of transistor devices, two responses (drain and gate) with different characteristics are easily obtained. This increases the number of virtual nodes by overcoming the said trade-off relationship. By employing a redox-IGR, second-order nonlinear dynamical tasks and a second-order nonlinear autoregressive moving average (NARMA2) were successfully solved, with normalized mean square errors (NMSEs) of $5.39 \times 10^{-4}$ and 0.163, respectively. Said IGR structure, with inorganic materials, is useful as a building block for semiconductor integrated circuits. Therefore, it is shown that the approach described herein can contribute to the physical implementation of physical reservoirs in practical devices that require compatibility together with high computational performance and high-density integration.

The novelty and differences of the redox-IGR with respect to the electric double layer (EDL)-IGR reported in our previous work are the operating mechanisms of the IGR. While the EDL-IGR utilizes electric response of EDL transistors on the basis of EDL charging/discharging at the channel/electrolyte interface,[21] the subject redox-IGR utilizes electric response of redox transistors on the basis of ion insertion/desertion into the channel (redox). The feature of the redox-IGR realizes double reservoir states in drain and gate nonlinear responses, as discussed later.

2. Results and Discussion

2.1. General concept of the redox-IGR

In order to perform the efficient pattern recognition required by neuromorphic computing, including common deep learning and physical reservoir computing, time-series data are processed as input through their mapping to high dimensional feature space.[1–2] An example of a general neuromorphic computer mapping scheme is shown in Figure 1a. In physical reservoir computing, which is the main concern of the present study, such mapping can be performed by inputting time-series data to a “physical reservoir” so as to generate output schematically the same as the physical reservoir computing shown in Figure 1b. Utilizing inherent functions of the physical reservoir, time-series data are nonlinear-transformed and used to computing various pattern recognition tasks.[1–26] To achieving high performance with diverse outputs, physical reservoir computing requires nonlinearity, high dimensionality, and short-term memory. We examined the computing performance of a redox-based ion-gating transistor as a physical reservoir candidate; referred to in this study as a redox-IGR. Figure 1c is a schematic diagram of the redox-IGR operating on the redox of a Li$_x$WO$_3$ channel. The nonlinear $I$–$V$ characteristic of the redox-IGR, explained in detail later, is used to map input signals to high dimensional feature space. The subject redox-IGR was fabricated by RF sputtering to deposit a LiCoO$_2$ thin film (200 nm) on a 0.15 mm-thick LICGC substrate, with a large storage capacity Li-ion/Pt thin film (50 nm) as a gate electrode, Pt thin film (50 nm) as drain and source electrodes, and WO$_3$ thin film (100 nm) as a channel, respectively. As pretreatment for the IGR prior to commencing reservoir computing operations, a constant voltage of 2.5 V was applied between the gate and source electrodes for an hour. This operation inserted Li ions into the WO$_3$ channel to create a Li$_x$WO$_3$ phase with a smooth Li$^+$ insertion/desertion characteristic. The $I_D$ (drain current)–$V_G$ (gate voltage) and $I_C$ (gate current)–$V_C$ characteristics of the redox-based IGR are shown in Figure 1d. The gate voltage is swept from 0.5 to 1.5 V and then back to 0.5 V at various sweep rates, ranging from 5 mV s$^{-1}$ (slow) to 250 mV s$^{-1}$ (fast). $I_D$ is normalized at the initial value for comparison to those measured under different sweeping rate conditions. The $I_D$ is modulated by the application of $V_C$ because the conducting electron is doped (or removed) in the Li$_x$WO$_3$ by the redox reaction (Li$^+$ insertion or desertion) (Equation (1)).

$$\text{Li}_x\text{WO}_3 + y\text{Li}^+ + ye^{-} \leftrightarrow \text{Li}_{x+1}\text{WO}_3$$  \(1\)

The $I_D$–$V_C$ curves exhibited hysteresis at sweep rates ranging from 5 to 250 mV s$^{-1}$ (a frequency range of from 2.5 to 125 MHz). In the subject redox-IGR transistor, Li$^+$ transport in the Li$_x$WO$_3$ channel is much slower than in the electrolyte; a rate-limiting step of the overall Li$^+$ transport is Li$^+$ transport in the Li$_x$WO$_3$ channel. Another main cause of the hysteresis characteristics observed in the subject transistor is the delay of Li$^+$ transport in the Li$_x$WO$_3$ channel relative to gate voltage sweep, which is the origin of the short-term memory of the redox-IGR. In addition to these short-term memory characteristics, as the sweep rate decreases, the modulation of the $I_D$ response to the applied gate voltage increases and the nonlinearity becomes stronger. Therefore, by changing the sweep rate of the gate voltage (or the frequency of the input signal applied to the gate), the nonlinearity in the redox-IGR can be modulated, which is an important characteristic for a physical reservoir to have.

To perform time-series tasks using the transistor, $I_D$ response to gate voltage pulse is useful for mapping input signals to higher dimensional feature space. $V_C$ pulse streams can be used to deal with sequential time-series signals. The upper and middle panels of Figure 1e show an example of $I_D$ response with respect to $V_C$ pulse streams, which are input signals to the transistor. When one $V_C$ pulse (corresponding to one point in a time-series dataset) is input, 20 reservoir states $X_i$ ($i = 1, ..., 20$) can be obtained from the $I_D$ response by the virtual node method. Conventional IGRs use only the $I_D$,[21] but the subject redox-IGR can use $I_C$ as a reservoir state. This is due to the significant gate current present, which is much larger than the one found in an electric double layer-IGR.[21] While very small $I_C$ in an electric double layer-IGR suffers from noise floor and thus makes it extremely difficult to obtain reliable reservoir states from $I_C$, $I_C$ in the subject redox-IGR, which is comparable to $I_D$, is in fact suitable for obtaining reliable reservoir states. Therefore, 20 additional reservoir states $X_i$ ($i = 21, ..., 40$) can be obtained from the $I_C$ response, as shown in the lower panel of Figure 1e. The doubled reservoir states can be utilized to perform reservoir computing with enhanced high dimensionality, as schematically shown in Figure 1f. Recently, mixed reservoir properties, with different dynamical characteristics, were theoretically predicted to show high-performance reservoir computing.[54] Due to their different characteristics, the
doubled reservoir states obtained by $I_D$ and $I_C$ responses can derive such a mixed reservoir property effect and result in high computation performance. The details of computation performance with specific tasks will be discussed later.

Previously, a SiO$_x$ (doped with Ag)-based diffusive memristor was applied to reservoir computing to identify handwritten digits from the Modified National Institute of Standards and Technology (MNIST) dataset, the results of which were a highly accurate 83%.[26] Quickly fading current in the memristor, which functions as a short-term memory for resistor–capacitor (RC), was obtained by resistance modulation of the memristor due to voltage stimulated fast Ag diffusion and the resultant volatile redox reactions in the SiO$_2$ film, which is a typical diffusive memristor behavior.[55,56] In the subject redox-IGR, the similarly quick fading $I_C$ observed is due to Li$^+$ diffusion inside the electrolyte and Li$^+$ transfer between the electrolyte and Li$_x$WO$_3$ ion–electron mixed conducting channel layer.

In conventional transistors, low gate current is always preferred so as to keep energy consumption low, especially in standby states. In the subject redox-IGR, no voltage or current was obtained by resistance modulation of the memristor due to voltage stimulated fast Ag diffusion and the resultant volatile redox reactions in the SiO$_2$ film, which is a typical diffusive memristor behavior.[55,56]

Figure 1. a) General scheme for the mapping of input to high-dimensional feature space in neuromorphic computing. b) General scheme of physical reservoir computing. c) Schematic image of a Li$_x$WO$_3$-based redox-IGR. d) Normalized drain current, and gate current measured during $V_G$ sweeping from 0.5 to 1.5 V. e) Gate voltage pulse stream, drain current response, and gate current response during operation of the redox-IGR. 40 reservoir states $X_i$ ($i = 1, ..., 40$) are obtained as shown in the panels to the right. f) General concept of a reservoir computing system with redox-IGRs. $W_i$ denotes the read-out weight.
application to the redox-IGR is required in the standby state; both $I_D$ and $I_C$ are used only during information processing. Therefore, the significant $I_C$ of the redox-IGR is unlikely to cause any serious issues.

### 2.2. Solving a Second-Order Nonlinear Dynamic Equation

Reservoir computing is advantageous for time-series data analysis due to the nonlinearity, short-term memory, and high dimensionality of the reservoir for input signals. Therefore, we evaluated the computational performance of the subject redox-IGR in time-series data analysis by solving a second-order nonlinear equation task. The general concept of a process flow diagram for the second-order nonlinear equation task solved is shown in Figure 2a.\[7,8\] The target time-series $y_t(k)$ for this task is generated by the second-order nonlinear dynamic equation shown in Equation (2).

$$y_t(k) = f[y_t(k-1), y_t(k-2), u(k)]$$

where $u(k)$ and $k$ are a random input ranging from 0 to 0.5 and the discrete time, respectively. Equation (2) contains a second-order nonlinearity and a two-step forward term which, in order to solve the equation, are required to be expressed by the reservoir as linearly separable.\[7]

A random input $u(k)$ was linearly converted to a voltage pulse stream, with a pulse period of $T$ (2–100 s) and a duty rate of 50%, which was input to the subject redox-IGR transistor under a constant $V_D$ of 0.1 V.

The pulse intensity $V_C(k)$ ranged from 0.5 to 1.5 V ($V_C(k) = 2u(k) + 0.5 V$) and a constant $V_C$ of 1 V was applied during the pulse interval. The gate voltage pulse stream, drain current response, and gate current response are shown in Figure 1e. As already discussed, a total of 40 reservoir states $X_i$ were obtained from both the $I_D$ ($i = 1, \ldots, 20$) and the $I_C$ ($i = 21, \ldots, 40$) responses by the virtual node method.

By combining physical and virtual nodes with different characteristics, the unique electrical behavior of the subject redox-IGR, caused by redox reactions (electronic current through channel and ion currents through electrolyte, which are associated with $\text{Li}^+$ transport), can be extracted as reservoir states that are valid for reservoir computing and can further be mapped to a high-dimensional feature space.\[21\] As a result, the reservoir output $y(k)$ is obtained by the following equation

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

where $N$, $w_i$, and $b$ are the size of the reservoir ($= 40$), read-out weights, and bias, respectively (see Experimental Section for additional details on the learning algorithm).

In the test phase, in order to evaluate the generalization performance of the subject redox-IGR, we checked whether the reservoir output (Equation (3)) with fixed $w_i$ matched the learned equation (Equation (2)) for inputs different from those in the training phase. The prediction error defined below was used to evaluate the computational performance of the subject redox-IGR in this task

$$\text{Prediction error} = \frac{\sum_{k=1}^{L} (y(k) - \hat{y}(k))^2}{\sum_{k=1}^{L} \hat{y}(k)^2}$$

where $L$ ($= 150$) is a data length.

Figure 2b,c shows the target and predicted waveforms when the subject redox-IGR transistor was operated at different pulse periods of $T$ of 2 and 40 s in the test phase. It is particularly noteworthy that the target and predicted waveforms are in excellent agreement for $T = 40$ s, as shown in Figure 2c. That is, Equation (2) was successfully solved by the subject redox-IGR, with a prediction error of $5.39 \times 10^{-4}$, which is sufficiently low when compared to other physical reservoirs reported to date ($1.31-3.13 \times 10^{-4}$)\[7,8] as shown in Figure 2d. Therefore, it is suggested that the IGR system is a mechanism that enables high-performance reservoir characteristics. On the other hand, the prediction error worsened when the subject IGR was operated with $T$ of 2 s, as shown in Figure 2b. This is because the relaxation process in the subject redox-IGR is correlated with the sweep rate of the gate input, as detailed in Figure 1d. To evaluate the correlation between the operating conditions of the subject redox-IGR and its computational performance as a reservoir, we investigated the relationship between $T$ and the prediction error, as shown in Figure 2e. The red line in the figure shows the results when only the $I_D$ is used for the reservoir states (i.e., $X_i$, $i = 1, \ldots, 20$), and the purple line shows the results when both the $I_C$ and the $I_D$ are used for the reservoir states (i.e., $X_i$, $i = 1, \ldots, 40$). It is found that the prediction performance is best at $T = 40$ s, regardless of the presence or absence of a gate current. This is because, when the pulse period is short, the redox reaction shown in Equation (1) proceeds with too great a delay, and the ion current associated with ion transport in the electrolyte dominates. Therefore, the short-term memory characteristics and nonlinearity due to the resistance modulation of $\text{Li}_2\text{WO}_4$ are lost, and the only current response obtained is a simple one similar to the relaxation process of a RC parallel circuit.

In addition, if the pulse period is too long, the interaction between the virtual nodes is suppressed, which results in poor computational performance.\[10\] In order to compare the utilization of double reservoir states alone, the results of solving a task with 20 nodes consisting of $I_D$ and $I_C$ (i.e., $X_i$, $i = 1, 3, \ldots, 37, 39$) are shown with black line in Figure 2e. The predicted performance at each pulse period was higher than with a single reservoir state using $I_D$ or $I_C$, and effective nodes were successfully obtained by adding $I_C$. The utilization of $I_C$ in addition to $I_D$ not only lowers the said error, but also moderates the dependence of the computational performance on the input conditions. This is because, in addition to increased expressive power due to the increased reservoir size, the $I_D$-derived $X$ and $I_C$-derived $X$ utilize complementary features that are necessary for task execution, which is an important feature of the subject redox-IGR, which utilizes different physical nodes as computational resources. This feature of the subject redox-IGR, which provides good computational performance regardless of slight changes in its operating conditions, is also an extremely significant practical advantage for reservoir computing implementation.
The high performance shown in Figure 2 is modulable nonlinearity in both $I_D$ and $I_G$ as discussed below. In the $V_G$ sweeping measurement shown in Figure 1d, $I_D$ is modified by electronic carrier density change due to Li$^+$ insertion/desertion from the Li$_x$WO$_3$ layer. As the Li$^+$ insertion/desertion process is driven by a gradient of the electrochemical potential of Li$^+$ (composed of the chemical potential of Li$^+$ and the local electrostatic potential) in the Li$_x$WO$_3$ layer, the local flux of Li$^+$ through the electrolyte/Li$_x$WO$_3$ interface is strongly influenced by the electrolyte and by both the Li$^+$ density (the chemical potential of Li$^+$) profile and the electrical potential profile in the Li$_x$WO$_3$ layer. Under Li$^+$ flux, the Li$^+$ density profile variation follows $V_G$ sweeping with notable delay due to the relatively slow Li$^+$ diffusion kinetics in the Li$_x$WO$_3$ layer, so the extent of the delay thus depends on the $V_G$ sweep rate. Furthermore, local Li$^+$ density variation is accompanied by local electron density variation due to charge compensation with Li$^+$ [shown in Equation (1)], making the electrical potential profile vary with delay. These lead to modulable nonlinearity of $I_D$. On the other hand, the total Li$^+$ flux (corresponding to $I_G$) through the electrolyte/Li$_x$WO$_3$ interface also follows $V_G$.
sweep with delay, which is influenced by the potential profiles discussed above. Therefore, both $I_C$ and $I_G$ show modulable nonlinearity. These modulable nonlinearities hold under $V_G$ pulse stream applied conditions with different pulse periods ($T$). Furthermore, said $I_D$ and $I_C$ are sensitive to the $V_G$ input history because the history is stored in unique Li$^+$ density profiles in the Li$_2$WO$_3$ layer. We believe that these are the origin of the observed high performance of the subject redox-IGR.

Note that in addition to the $I_D$ and $I_C$ responses of the subject redox-IGR, it is entirely possible for other nonlinear data to be used to obtain the waveforms in the RC tasks. However, depending on the dynamical characteristics of the device (the source of the nonlinear data) used as the reservoir, performance in physical RC may change significantly. The three requirements for physical reservoirs—nonlinearity, short-term memory, and high dimensionality—are of great importance for the achievement of high-performance RC.

In comparing the subject redox-IGR and the EDL-IGR,\cite{21} there is a notable difference in their respective prediction errors, which may be due to the inherent properties of the channel material used for the subject redox-IGR. The Li$_2$WO$_3$ used in the subject redox-IGR is a well-known material used in electrochromic windows, which require their redox ON and OFF states to be highly repeatable. Because of this, Li$_2$WO$_3$ was expected to be a suitable channel material in the subject redox-IGR. However, the occurrence of irreversible Li$^+$ trapping during repetition of ON and OFF states has been recently pointed out.\cite{56,57,58} Although such irreversibility appears to be not significant in our redox-IGR, we suspect that it is possible that Li$^+$ trapping can cause increased or decreased losses of the echo state property, which is required for high-performance computing, and thus may lead to increased prediction errors. Further to this point, it is expected that errors in the subject redox-IGR can be reduced by using alternative ion–electron mixed conductors, in which such ion trapping does not occur.

### 2.3. Evaluation of Prediction Performance for a NARMA2 Task

To further evaluate the performance of time-series prediction using the subject redox-IGR, we have performed a NARMA2-task, which is more difficult than the second-order nonlinear equation task performed in the previous section, as well as being a typical benchmark task for both full-simulation reservoir computing and physical reservoir computing.\cite{4,6,22,23,25} The time-series prediction generated by the NARMA2 model, with specific parameters defined by Equation (5), is a popular benchmark for the development of physical reservoirs.\cite{4,6,22,23,25}

$$y_t(k + 1) = 0.4y_t(k) + 0.4y_t(k)y_t(k - 1) + 0.6u^1(k) + 0.1$$ \hspace{1cm} (5)

**Figure 3a,b** shows the results of waveform prediction utilizing $I_D + I_C$ (40 nodes) at $T = 2$ s and $40$ s, respectively. The error value, a normalized mean square error (NMSE), between the target waveform (blue line) and the predicted waveform (orange line) is defined by Equation (6)

$$\text{NMSE} = \frac{1}{L} \sum_{k=1}^{L} \frac{(y_t(k) - \hat{y}_t(k))^2}{\sigma^2(y_t(k))}$$ \hspace{1cm} (6)

where $L$ (150) is a data length. The computational performance of the subject redox-IGR in performing the NARMA2 task was enhanced by the higher dimension, combined with $I_C$. As shown in the comparison of the two conditions ($T = 2$ and $40$ s) using $I_D + I_C$ (40 nodes) in Figure 3a,b, the minimum value of NMSE is 0.163 at $T = 40$ s, whereas the maximum NMSE is 0.321 at $T = 2$ s. These two waveforms were chosen as examples to show how the predicted waveforms, with relatively high and low NMSE, differ from each other. As can be seen from a comparison of the two, deviation of the predicted waveform from the target waveform appears less significant in Figure 3b, giving support to a predicted accuracy higher at $T = 40$ s than at $T = 2$ s. Figure 3c shows the relationship between NMSE and pulse period at each reservoir state. When using only $I_D$ as the reservoir, as shown by the red line in Figure 3c, the best predicted NMSE performance was $0.212$ ($T = 40$ s). As in the case of the second-order nonlinear equation task in Figure 2, the positive effects from the utilization of the double reservoir states alone were evaluated by excluding the collateral effect of increasing the number of nodes to 40. The result for a double reservoir state consisting of 20 nodes ($I_D + I_C$) is indicated by the black line in Figure 3c. The relationship between NMSE and pulse period under all conditions was the same as for the second-order nonlinear equation task in Figure 2. The general tendency, in which utilizing $I_C$ in addition to $I_D$ gives better performance over the whole pulse period range, is quite similar to the case for the second-order nonlinear equation task, which gives support to our assumption that the present approach is sufficiently versatile to achieve an information processing ability in the subject redox-IGR.

In order to further investigate the underlying mechanism of the enhancement effect elicited by the addition of $I_C$ to the reservoir states, we performed a short-term memory task, which task measures the ability of our redox-IGR to reconstruct past time series data input to the redox-IGR. Here, as in the time series analysis task described in Figure 2 and 3, a voltage-transformed random input $u(k)$ is applied to the subject redox-IGR and the input $u(k-r)$ before the delay time $\tau$ is reconstructed by a linear combination of reservoir states and weights obtained from the current response of the subject redox-IGR (Equation (3)). The agreement between the target waveform $u(k-r)$ and the reconstructed waveform $y(k)$ by the reservoir was evaluated using the following coefficient of determination $r^2$ \cite{24}

$$r^2(\tau) = \frac{\text{Cov}^2(u(k-r), y(k))}{\text{Var}(u(k)) \times \text{Var}(y(k))}$$ \hspace{1cm} (7)

where Cov() and Var() are the covariance and variance, respectively. Figure 3d shows the forgetting curve (determination coefficient vs delay) of the subject redox-IGR using $I_D$ and $I_D + I_C$ for utilization of 40(20) nodes at $T = 40$ s.\cite{25} The determination coefficient $r^2$ (i.e., the ability for reconstruction) decreases as the delay increases. That is a universal feature of short-term memory. Memory capacities (MC) for the three conditions are calculated to be 2.35 for $I_D$, 2.80 for $I_D + I_C$ (20 nodes), and 3.57 for $I_D + I_C$ (40 nodes), respectively, by integration of the curves in Figure 3d as follows

$$\text{MC} = \sum_{r=1}^{\infty} r^2(\tau)$$ \hspace{1cm} (8)

As shown in Figure 3e, by comparing to MCs under the three conditions, it is revealed that the double reservoir states
combining $I_G$ and $I_D$ enhance both the high dimensionality and the MC of the reservoir.

### 2.4. Evaluation of Output Versatility with Correlation Efficient Between Each Node

Increasing the number of virtual nodes with sufficient diversity can enhance high dimensionality, leading to high-performance reservoir computing. The apparent difference of the $I_D$ and $I_G$ characteristics observed in Figure 1e and the significant performance improvement accompanied by the addition of $I_G$ [observed in Figure 2 and 3] indicate that adding $I_G$ causes an increase in diverse nodes, which are not strongly correlated to existing $I_D$ nodes. However, this mechanism is not evidenced in the above discussion. In order to clarify the enhancement mechanism for $I_G$ addition in relation to the diversity of nodes, the correlation between each node $i$ under given conditions (e.g., utilizing of $I_D$ only, or $I_D + I_G$, $T$) was quantified by the Pearson correlation coefficient $r(X_i, X_j)$, using the following equation

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

where $X_i$, $k$, and $L$ are the reservoir state of node $i$, and the discrete time and data length, respectively. Figure 4a shows the reservoir state waves for $X_5$ (black line) and $X_7$ (red line) obtained from the $I_D$ response. The $X_i$ and $X_j$ waveforms were similar to each other, and a high correlation was confirmed with a calculated $r$ of 0.95, as shown in Figure 4b. Although $r$ can express positive correlations ($r > 0$) and negative correlations ($r < 0$) between each node by taking values in the range of $-1$ to $1$, $1 - |r|$ ($0 \leq -1 - |r| \leq 1$) is rather useful for evaluating the extent of correlation, regardless of the sign, between each node. $1 - |r|$ for the $X_5$ and $X_7$ waveforms was calculated to be 0.05, indicating high correlation and poor versatility. On the other hand, the waveforms for $X_5$ (black line) from $I_D$ and $X_{15}$ (blue line) from $I_G$ appear completely different, as shown Figure 4c, and the calculated $1 - |r|$ of 0.92 ($r = 0.08$) indicated almost no correlation, as shown in Figure 4d. From these comparisons, it was confirmed that $1 - |r|$ can be a useful index for evaluating the versatility of virtual nodes. $1 - |r|$ can thus be understood as an uncorrelated coefficient for a specific combination of two nodes. Figure 4e shows a heatmap representing $1 - |r|$ between each node ($X_i$ to $X_{60}$ in the vertical axis vs $X_i$ to $X_{55}$ in the horizontal axis) measured at $T = 40$ s. The heatmap has linear symmetry with respect to a diagonal line because a pixel for a specific combination ($X_i$, $X_j$) is equivalent to the one for the corresponding combination ($X_j$, $X_i$). If $1 - |r|$ for a specific combination (e.g., $X_5$ vs $X_{15}$...
Figure 4. a) $X_5$ for $I_D$ (black line) and $X_7$ for $I_D$ (red line) waveforms at $T = 40$ s. b) The scatter plot between $X_5$ and $X_7$ with high correlation ($r = 0.95$). c) $X_5$ for $I_D$ (black line) and $X_{35}$ for $I_G$ (blue line) waveforms. d) The scatter plot between $X_5$ and $X_{35}$ without correlation ($r = 0.08$). e) The heatmap of $1/|C_0|r$ for 40 nodes ($X_1, X_2, \ldots, X_{40}$) at $T = 40$ s. f) The heatmaps of $1/|C_0|r$ for 40 nodes ($X_1, X_2, \ldots, X_{40}$) measured under all $T$ conditions. g) The relationship between memory capacity and sum of $1/|r|$. Memory capacity dependence of h) prediction error for second-order nonlinear dynamic equation task and i) NMSE for a NARMA2 task. Each orange fitting curve is inserted for easier understanding of the characteristic.
indicated by a green circle) is close to 1, the color of the corresponding pixel becomes dark, and expresses that the correlation has low and high versatility, and vice versa. The regions surrounded by red, blue, and purple squares represent $I_D$ ($X_1$ to $X_{20}$) versus $I_D$ ($X_1$ to $X_{20}$), $I_C$ ($X_{21}$ to $X_{40}$) versus $I_C$ ($X_{21}$ to $X_{40}$), and $I_D$ ($X_1$ to $X_{20}$) versus $I_C$ ($X_{21}$ to $X_{40}$) correlations, respectively. In each region, there is a notable distribution of $1/|r|$. Figure 4f shows several $1/|r|$ heatmaps measured under various operation conditions from $T = 2$ to 100 s. As $T$ is increased from 4 to 10 s, $1/|r|$ in a part of the $I_D$ versus $I_D$ region ($X_1$ to $X_{10}$ vs $X_{11}$ to $X_{19}$) becomes much higher (darker) than the ones at $T = 2$ and 4 s. In addition, $1/|r|$ in a part of the $I_D$ versus $I_G$ region ($X_1$ to $X_{20}$ vs $X_{21}$ to $X_{40}$) also becomes slightly higher. More significantly, at $T = 20$ s or above, $1/|r|$ of the high $1/|r|$ domain becomes very high and much broadened in both the $I_D$ versus $I_D$ and $I_D$ versus $I_G$ regions, meaning that the $I_D$ and $I_G$ nodes become more uncorrelated to enhance high dimensionality. As $1/|r|$ between each node can express an effectiveness of the nodes, a sum of $1/|r|$, described by $\sum_{i \neq j} \sum_{\ell \neq i,j} (1/|r_{ij}|)$, can be an index to compare versatility and high dimensionality in overall outputs. We compare the sum of $1/|r|$ and MC so as to analyze the relationship between versatility (high dimensionality) and MC. Figure 4g shows MC versus the sum of $1/|r|$ plots under only $I_D$, $I_D + I_G$ (20 nodes), and $I_D + I_G$ (40 nodes) conditions. Positive correlation is clearly found between MC and the sum of $1/|r|$, regardless of

![input V_G](image)

**Figure 5.** $I_D$ and $I_G$ responses of device A and B, with the same device dimensions, under identical input $V_G$ pulse train applied conditions. The left (right-)hand side shows the result in the beginning (last) part of the input $V_G$ pulse train.
conditions, meaning that the high versatility (high dimensionality) caused by $I_C$ addition surely contributes to the strengthening of MC. We further investigated the relationship between MC and computing performance for a second-order nonlinear dynamic equation and NARMA2, as shown in Figure 4h,i. For both tasks, computation performance is improved as MC increases from below 2.0 to about 4.0. These results evidence that $I_C$ addition enhances the high dimensionality (versatility) of the output, leading to higher computation performance accompanied by MC increase. This is consistent with the fact that computing performance in both the second-order nonlinear dynamic equation task and the NARMA2 task is higher at $T = 4$ s than $T = 2$ s.

2.5. Repeatability and Stability of the Redox-IGR

In order to investigate repeatability and stability of our redox-IGR, we compared the $I_D$ and $I_C$ responses of two devices, each with the same device dimensions (device A and B), under identical $V_C$ pulse train applied conditions. The left-hand side of Figure 5 shows the $I_D$ and $I_C$ responses of device A and B in the beginning part of the input $V_C$ pulse train. When comparing the $I_D$ and $I_C$ responses, the two devices were found to give very similar responses, with repeated spiking and relaxation. Moreover, during the last part of the $V_C$ pulse train, as shown on the right-hand side of Figure 5, device A and B continued showing such similar $I_D$ and $I_C$ responses. From said responses, we were able to obtain stable multiple 40 states, as well as those shown in Figure 2 and 3. This result supports that the subject redox-IGR has sufficient repeatability and stability to performing reservoir computing.

3. Conclusions

Physical reservoir computing, with a redox-IGR composed of Li$_x$WO$_3$ thin film and LICGC, has been demonstrated. The subject redox-IGR successfully solved a second-order nonlinear dynamic equation, with a lowest prediction error of $8.15 \times 10^{-4}$, under a normal condition where only $I_D$ is used for reservoir states. Performance was enhanced by the addition of $I_C$ to the reservoir states, resulting in a significant lowering of the prediction error to $5.39 \times 10^{-4}$, which is noticeably lower than other types of physical reservoirs reported to date. NARMA2, a typical reservoir computing benchmark, was also performed with the subject redox-IGR. Better performance was achieved, with an NMSE of 0.163, by the addition of $I_C$ to the reservoir states, which reveals that $I_C$ is a useful source for obtaining better reservoir properties. A short-term memory task was performed to investigate enhancement mechanism resulting from the addition of $I_C$. The forgetting curves of the subject redox-IGR show that MC was enhanced from 2.35 with $I_D$ to 3.57 with $I_D + I_C$. The enhancement of both high dimensionality and MC resulting from the addition of $I_C$ to the reservoir states is attributed to the origin of the performance improvement. Physical reservoir computing is, from a certain viewpoint, an attempt to utilize as many inherent properties of a material/device as is possible so as to achieve efficient information processing. For a transistor, $I_C$ is usually regarded as of no use; nevertheless, it includes certain internal and temporal information about the transistor. The present technique is useful in harnessing such internal information in a device so as to realize the efficient mapping of input to higher dimensional feature space. This approach can be applied to a wide range of multicomponent physical reservoir systems.

4. Experimental Section

Fabrication of a Li$_x$WO$_3$-Based Redox Transistor: The Li$_x$WO$_3$-based redox transistor, schematically shown in Figure 1c, was fabricated on a 0.15 mm-thick LICGC substrate. First, the drain and source electrodes, made of 50 nm-thick Pt films, were deposited by the RF sputtering method at room temperature. Next, a 100 nm-thick WO$_3$ thin film was deposited by the RF sputtering method, using a 99.9% pure, sintered stoichiometric WO$_3$ target, with a supply of pure Ar and O$_2$ gases at fixed flow rates of 10 and 0.6 sccm, respectively. A 200 nm-thick LiCoO$_2$ thin film was then deposited as a gate electrode on the opposite side of the LICGC substrate, against the WO$_3$ side, with a supply of pure Ar and O$_2$ gases at fixed flow rates of 9 and 3 sccm, respectively. Finally, the current collector on the gate electrode, made of 50 nm-thick Pt film, was deposited by the RF sputtering method at room temperature. Prior to measurements being made, a constant voltage of 2.5 V was applied between the gate and source electrodes so as to insert Li ions into the WO$_3$ channel.

Measurement of $I_D$ and $I_C$: All electrical measurements of the subject redox-IGR were carried out at room temperature in a vacuum chamber and carried out using the source measure unit (SMU) of a semiconductor parameter analyzer (4200 A-SCS, Keithley). A random input $u(k)$ was linearly converted to the voltage pulse streams, with a pulse period of $T = 0.5$ ms and a duty rate of 50%, which was input to the subject redox-IGR transistor under constant $V_G$ of 0.5 V. The pulse intensity $I_C(k)$ ranged from 0.5 to 1.5 V ($V_C(k) = 2u(k) + 0.5V$) and constant $V_G$ of 1 V was applied during pulse intervals. The $I_D$ and $I_C$ responses of the subject redox-IGR were monitored, and 20 virtual nodes were extracted from each response. Thus, 40 reservoir states were obtained from input $u(k)$ by the subject redox-IGR. Said reservoir states were normalized from 0 to 1 for calculation, as shown in Equation (3).

Ridge Regression for Time-Series Data Analysis Tasks: In the time-series data analysis tasks, such as the solving of the second-order nonlinear dynamic task and the NARMA2 task shown in Figure 2 and 3, the readout network of the subject redox-IGR was trained by ridge regression. Here, we describe the algorithm used for said ridge regression. The reservoir output $y(k)$ shown in Equation (3) can also be defined as follows:

$$y(k) = w \cdot x(k)$$

where $w = (w_1, w_2, \ldots, w_N)$ and $x(k) = (1, x_1(k), \ldots, x_N(k))^T$ are the weight vector and the reservoir state vector with a reservoir size of $N$, respectively. The cost function $J(w)$ in ridge regression is defined as follows:

$$J(w) = \frac{1}{2} \sum_{k=1}^{L} (y(k) - y(k))^2 + \frac{\lambda}{2} \sum_{i=1}^{N} w_i^2$$

where $L$, $\lambda$, and $y(k)$ are the data length in the training phase, the ridge parameter, and the target output generated by Equation (2) or (5), respectively. The data length and ridge parameter were $L = 150$ and $\beta = 5 \times 10^{-4}$ for the second-order nonlinear equation task. The NARMA2 task used $\beta = 1 \times 10^{-4}$. The trained weights $w$ that minimize cost function $J(w)$ are given by the following equation:

$$w = (X^T(X^T + \beta I)^{-1}$$

where $Y = (y_1(1), y_1(2), \ldots, y_1(L))$, $X = (x(1), x(2), \ldots, x(L))$, and $I \in \mathbb{R}^{(N+1) \times (N+1)}$ are the target output vector, the reservoir state matrix, and the identity matrix, respectively.
Acknowledgements

T.W. and D.N. contributed equally to this work and treated as co-first authors. This work was in part supported by Japan Society for the Promotion of Science (JSPS) KAKENHI Grant Number JP22H04625 (Grant-in-Aid for Scientific Research on Innovative Areas “Interface Ionics”) and JP21J21982 (Grant-in-Aid for JSPS Fellows). A part of this work was supported by the Yazaki Memorial Foundation for Science and Technology.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Keywords

all-solid-state, ion-gating reservoir, ion-gating transistor, lithium ion, nanoionics, neuromorphic, reservoir computing

Received: March 9, 2023
Revised: April 18, 2023
Published online: June 29, 2023

[1] H. Jaeger, H. Haas, Science 2004, 304, 78.
[2] G. Tanaka, T. Yamane, J. B. Heroux, R. Nakane, N. Kanazawa, S. Takeda, H. Numata, D. Nakano, A. Hirose, Neural Netw. 2019, 175, 100.
[3] Y. Paquot, F. Duport, A. Smerieri, J. Dambre, B. Schrauwen, M. Haelterman, S. Massar, Sci. Rep. 2012, 2, 287.
[4] K. Nakajima, H. Hauser, T. Li, R. Pfeifer, Sci. Rep. 2015, 5, 10487.
[5] S. Kan, K. Nakajima, Y. Takeshima, T. Asai, Y. Kuwahara, M. Akai-Kasaya, Phys. Rev. Appl. 2021, 15, 024030.
[6] M. Akai-Kasaya, Y. Takeshima, S. Kan, K. Nakajima, O. Tya, T. Asai, Neuromorphic. Comput. Eng. 2022, 2, 014003.
[7] C. Du, F. Cai, M. A. Zidan, W. Ma, S. H. Lee, W. D. Lu, Nat. Commun. 2017, 8, 2204.
[8] W. Jiang, L. Chen, K. Zhou, L. Li, Q. Fu, Y. Du, R. H. Liu, Appl. Phys. Lett. 2019, 115, 192403.
[9] J. Torrejon, M. Riou, F. A. Araujo, S. Tsunegi, G. Khalsa, D. Querlioz, P. Bortolotti, V. Cros, K. Yakuushii, A. Fukushima, K. Kubota, S. Yuasa, M. D. Stiles, J. Grollier, Nature 2017, 547, 428.
[10] Y. Zhong, J. Tang, X. Li, B. Gao, H. Qian, H. Wu, Nat. Commun. 2021, 12, 408.
[11] M. S. Kulkarni, C. Teuscher, in IEEE/ACM Int. Symp. on Nanoscale Architectures NANOARCH, IEEE, Piscataway, NJ 2012, pp. 226–232.
[12] J. Moon, W. Ma, J. H. Shin, F. Cai, C. Du, S. H. Lee, W. D. Lu, Nat. Electron. 2019, 2, 480.
[13] X. Zhu, Q. Wang, W. D. Lu, Nat. Commun. 2020, 11, 2439.
[14] L. Sun, Z. Wang, J. Jiang, Y. Kim, B. Joo, S. Zheng, S. Lee, W. J. Yu, B.-S. Kong, H. Yang, Sci. Adv. 2021, 7, eabg1455.
[15] X. Liu, K. K. Parhi, ACS Synth. Biol. 2022, 11, 2202.
[16] G. Tanaka, R. Nakane, Sci. Rep. 2012, 12, 9868.
[17] J. Hochstetter, R. Zhu, A. Loeffler, A. Diaz-Alvarez, T. Nakayama, Z. Kuncic, Nat. Commun. 2021, 12, 4008.
[47] T. Tsuchiya, K. Terabe, M. Aono, Appl. Phys. Lett. 2013, 103, 073110.
[48] T. Tsuchiya, K. Terabe, M. Aono, Adv. Mater. 2014, 26, 1087.
[49] T. Tsuchiya, M. Ochi, T. Higuchi, K. Terabe, M. Aono, ACS Appl. Mater. Interfaces 2015, 7, 12254.
[50] A. Lu, J. Sun, J. Jiang, Q. Wan, Phys. Lett. 2009, 95, 222905.
[51] T. Tsuchiya, M. Takayanagi, K. Mitsuishi, M. Imura, S. Ueda, Y. Koide, T. Higuchi, K. Terabe, Commun. Chem. 2021, 4, 117.
[52] T. Tsuchiya, S. Moriyama, K. Terabe, M. Aono, Appl. Phys. Lett. 2015, 107, 013104.
[53] M. Lübben, F. Cüppers, J. Mohr, M. von Witzleben, U. Breuer, R. Waser, C. Neumann, I. Valov, Sci. Adv. 2020, 6, eaaz9079.
[54] M. Inubushi, K. Yoshimura, Sci. Rep. 2017, 7, 10199.
[55] J. J. Yang, D. B. Strukov, D. R. Stewart, Nat. Nanotechnol. 2013, 8, 13.
[56] Z. Wang, S. Joshi, S. E. Savelev, H. Jiang, R. Midya, P. Lin, M. Hu, N. Ge, J. P. Strachan, Z. Li, Q. Wu, M. Barnell, G.-L. Li, H. L. Xin, R. S. Williams, Q. Xia, J. J. Yang, Nat. Mater. 2017, 16, 101.
[57] R. T. Wen, C. G. Granqvist, G. A. Niklasson, Nat. Mater. 2015, 14, 996.
[58] R. T. Wen, S. Malmgren, C. G. Granqvist, G. A. Niklasson, ACS Appl. Mater. Interfaces 2017, 9, 12872.
[59] T. Tsuchiya, T. Tsuruoka, S.-J. Kim, K. Terabe, M. Aono, Sci. Adv. 2018, 4, eaau2057.
[60] K. Terabe, T. Tsuchiya, T. Tsuruoka, Adv. Electron. Mater. 2022, 8, 2100645.
[61] K. Terabe, T. Tsuchiya, T. Tsuruoka, Jpn. J. Appl. Phys. 2022, 61, SM0803.
[62] T. Tsuchiya, T. Nakayama, K. Ariga, Appl. Phys. Express 2022, 15, 100101.
[63] W. S. Wang, L. Q. Zhu, Sci. Technol. Adv. Mater. 2023, 24, 2.
[64] T. Chen, J. van Gelder, B. van de Ven, S. Amitonov, B. de Wilde, H.-C. Ruiz Euler, H. J. Broersma, P. A. Bobbert, F. A. Zwanenburg, W. G. van der Wiel, Nature 2020, 577, 341.