Experimental study on parallel and analog optical reservoir computing with delayed feedback system for physical implementation

Tadashi Okumura\(^{1a)}\), Mitsuharu Tai\(^{1}\), and Masahiko Ando\(^{1}\)

\(^{1}\) Hitachi UTokyo Laboratory, Research \\& Development Group, Hitachi, Ltd.
4-6-1 Komaba, Meguro-ku, Tokyo 153-8904, Japan

\(^{a)}\) tadashi.okumura.tx@hitachi.com

Received July 10, 2018; Revised November 20, 2018; Published April 1, 2019

Abstract: Optical reservoir computing (RC) with delayed feedback is expected to achieve high-speed data processing. However, in the parallel RC framework, the digital pre-process limits the actual processing speed. An analog-based, simple pre-processing method was developed and implemented in an optical RC architecture to overcome the bottleneck, and the performance in calculating two types of parallel task was evaluated. One was two independent benchmark tasks; the other was an integrative multi-input and multi-output odor identification task. We successfully demonstrated that these two parallel tasks can be physically processed with the network parameters optimized to maximize the RC performance. These results strongly suggest the potential of the optical reservoir system for future high-speed multimodal data processing applications.

Key Words: reservoir computing, physical implementation, parallel computing, time-series analysis

1. Introduction

Our brains have a great ability to process several independent tasks simultaneously. The five senses are analyzed integratively in the frontal lobe of our brain followed by processing of each sense individually. We speculate that our brain does not feel each sense individually, but it totally analyzes biological information and understands the environment. Reservoir computing (RC) is one of the neuromorphic computing framework that partially try to mimic the neural behavior in brains [1, 2]. The reservoir is constructed with a random and untrained recurrent neural network (RNN). RC has an advantage in its ease of learning because an RC training algorithm requires adjusting only readout connections and not the entirety of network connections. This advantage should be useful to perform the above mentioned parallel processing of the brain, such as speech recognition [3]. For that reason, we think RC is a promising artificial approaches toward an energy-efficient parallel processor for multi-cognitive tasks simultaneously. In addition, RC can utilize the dynamics of the physical system as a reservoir.
Research on exploiting the dynamics of soft materials [4–6], quantum system [7] and several kinds of optical device [3, 8–10] are currently being reported. The soft material or a mechanical RC seems very interesting and unique technology. Body motions, for instance of robots or even an octopus arm, are extracted as calculation resources [4, 5]. From the perspective of robotics, multi-degree of freedom structure is effective to expand adaptability of its operation. However, the degree of freedom of soft material or body makes it difficult to control the body motions. The mechanical RC is expected to overcome the difficulty by turning the disadvantage of complex body dynamics into big advantage as abundant computational resources [6]. A quantum physical mechanism is also expected as physical reservoir resources. Time transient dynamics of quantum superposition of states are exploited as calculation resources [7]. For example, ensemble of nuclear spins using nuclear magnetic resonance system is applied. In this quantum systems, each basis state is regarded as a node of the physical reservoir. The quantum system has a potential to realize large scale network because N-qubits quantum systems construct as much as $2^N$ nodes of the physical reservoir. High computational performance has been achieved by using 5–7 qubits RC comparable to echo state networks (ESNs) [1] of 100–500 nodes for benchmark tasks [7]. An optical network or an optical device is one of representative physical reservoir component. Various kinds of optical device characteristics such as a relaxation oscillation frequencies of semiconductor laser [8] or the Kerr effect of photonic crystal cavity [9] can be applied as nonlinear, time transient physical resources. Advantages in the optical reservoir are generally regarded as low power, high speediness and parallel processing capability. The optical reservoir is mainly classified by two architectures. One is an optical integrated circuit [10] including large number of nonlinear optical devices as nodes. The other architecture is a delayed feedback system by using optical delay line [3, 8]. In this systems, optical pulse streams in the delay line are utilized as virtual nodes of the reservoir and the high-speediness enables to expand network size without worrying about power dissipation and delay in data processing. In addition, conventional commercial optical devices for optical fiber communications can be easily employed for computing.

Among the physical RC implementations, the optical RC with delayed feedback system is seen as a promising approach for extremely high-speed serial processor beyond the 1-Gbyte/s rate [11]. This system is originally based on an Ikeda model [12], utilizing the chaotic behavior of light from a ring cavity. A number of experimental studies related to electric and optical hybrid cavities have been reported [13, 14]. The cavity structure was also applied to a recurrent neural network [3]. In the reservoir with delayed feedback systems [15], a number of nonlinear nodes and complex connections of conventional neural networks are replaced by a single nonlinear device and a delay line. However, the digital pre-processing has a bottleneck, preventing high speed computation. In previous research, a fully analog architecture with one of pre-processing, called mask processing, was experimentally reported for time-multiplexing [16]. However, additional digital pre-processing, called signal multiplexing [15], is required for parallel computation. All of the pre-processing should be carried out in the analog systems to maintain the high-speediness of the optical reservoir.

In this study, to overcome this bottleneck of the optical reservoir for parallel computation, we introduced a small time delay adding to each input signal as a pre-processing step for high-speed analog computing in real time. We experimentally analyzed parallel computation performance using the optical reservoir with delayed feedback. Two different benchmark tasks were carried out concurrently as an independent parallel process. Then, odor information from multiple sensors were processed as integrative analysis.

2. RC Model with delayed feedback system and optical implementation

We adapted a reservoir model on the basis of delayed feedback systems [15]. Figure 1 shows the scheme of RC utilizing a nonlinear node with delayed feedback. Node dynamics $x(t)$ can be modeled using a delay differential equation

$$\frac{dx(t)}{dt} = -x(t) + f(x(t - T), u(t))$$

(1)
where $T$ is the delay time, $u(t)$ is the input signal, and $f$ is a nonlinear function. Optical delayed feedback systems have been demonstrated to be able to accomplish excellent results in complex problems such as nonlinear auto-regressive moving average (NARMA) and speech recognition tasks for a physical implementation. The delay time was divided into $N$ intervals. A high-dimensional space of the reservoir was extended into the time space of the delay line. Nonlinear optical devices are already commercially available technologies used for high-speed broadband optical networks, and continuous technological development should increase the range of photonics applications including optical reservoirs in the future.

The setup of the optical reservoir is shown in Fig. 2(a), which is similar to those of prior research [3], as mentioned in the previous section. The input signal was extended to 100-step pulse signals using a preprocessing algorithm with a mask process [15] as shown in top of Fig. 3. This process generates $100(=N)$ virtual nodes. In this experiment, the input signal was generated by Arbitrary Waveform Generator (AWG) and the maximum input signal intensity was adjusted to a setting driving voltage of the AWG. A telecom Mach-Zehnder (MZ) modulator was used as an electrical/optical converter and a nonlinear function $f$ as shown in Eq. (1). The output optical signal intensity $P$ of the MZ modulator varies sinusoidally with the input voltage $V$ as shown in Eq. (2). A phase $\phi$ of the sine wave was tuned by DC bias as shown in Fig. 2(a). Figure 2(b) shows an example of optical output from the MZ modulator as a function of input voltage. A half-wave voltage $V_{\pi}$ of the MZ modulator was experimentally estimated to 4 V and corresponding angular frequency $\omega$ was $\pi/(4V)$, respectively. The input voltage generated by the AWG was tunable parameter depending on experimental conditions. In this experiment, maximum input signal intensity was adjusted to the AWG setting voltage $V_{in}$. As
a result, the AWG voltage amplitude $V_{pp}$ shown in top of Fig. 3 corresponds to $2V_{in}$.

$$P(V) = 1 + \sin(\omega V + \phi)$$

The output optical pulse signals were coupled to the delay line, which was a 1 km-optical fiber and an RF electronic cable. 100 pulse signals with a 50.9 ns-width generated by the mask process were kept in the delay line as virtual nodes. Light propagation time through the delay line corresponds to the delay time $T$. The optical signal streams after propagating the delay fiber were converted to electrical pulses in the photoreceiver and then divided into two directions. One was the readout port, and the other was the feedback port. The feedback signal intensity $V_{fb}$ is a parameter adjusted by a voltage amplifier and an attenuator. As this feedback system could show bifurcation and chaotic properties because of the feedback signals, the intensity was kept low in order to sustain a non-chaotic state [13]. The readout signal was observed and processed using an oscilloscope and a control computer. Figure 3 shows a part of one scalar input $u(t)$ divided using a 100-step binary mask process and the corresponding reservoir dynamics $x(t)$ as an example. These time transient input and output streams are regarded as 100 dimensional vectors $u(t)$ and $x(t)$, respectively. When dealing with multiple ($n$-arrays) input signals, an input matrix is commonly applied as a separate mask for each signal [15]. However, multiple signals were overlapped in this experiment after adding a different small time delay (a mask pulse width delay between each signal) [17]. The reason is that a fully analog input pre-process is crucial to keep up with the optical reservoir network, as mentioned in the previous section. This analog physical simple pre-process for multiple signal inputs is preferred to a digital matrix operation. An updated input signal vector after the overlapping process is expressed with Eq. (3).

$$U(t) = \sum_{i=1}^{n} u_i(t - \frac{(i-1)T}{N})$$

The input stream $U(t)$ was converted to $N$ dimensional node activity $x(t)$ as shown in Eq. (1). RC output was obtained with only a weighted linear summation of each virtual node, as shown in Eq. (4):

$$y(t) = \sum_{k=1}^{N} w_k x_k(t)$$

The weight $w_k$ was obtained by minimizing the mean square error of the training data. In this research, output signals from the reservoir were once stored by the oscilloscope as shown in Fig. 2(a) and the data were transferred to the control computer for readout process shown with Eq. (4). For real time computation, high speed processing at the reservoir output port should be developed in future work.

### 3. Independent task

The benchmarks used were the Santa-Fe (SF) time-series prediction task [18] and NARMA10 [19] to evaluate the parallel computing performance with this system. The SF time-series data were obtained
from a far-infrared laser experimentally operating in a chaotic state. This task requires predicting a
discrete time-series laser intensity \( y(t+1) \) from a previous laser intensity \( y(t) \). The NARMA10 task
is nonlinear system modelling that is commonly used in the field of RC. The output \( y(t) \) is given by
Eq. (5) against time-series input value \( u(t) \), which is randomly and uniformly chosen between 0 and
0.5.

\[
y(t + 1) = 0.3y(t) + 0.05y(t) \sum_{i=0}^{9} y(t - i) + 1.5u(t - 9)u(t) + 0.1
\]

(5)

The output \( y(t+1) \) is determined using a past input and output stream, so this task requires a
memory of 10 past steps.

In this parallel experiment, signal numbers \( u_1(t) \) was assigned to SF time-series data, \( u_2(t) \) and
\( u_3(t) \) were assigned to two different random inputs as shown in Fig. 4(a). Then one or two out of
the three signals were input to the reservoir network system as shown in Fig. 4(b). For example,
the input signal stream data \( u_1(t) \) and \( u_2(t) \) were prepared separately by control computer and the
two signals were generated by the AWG (Model WW2572, Tabor Electronics Ltd), at output port
\#1 and \#2, respectively. Before combining these two signals, coaxial cables with length 10.05 m was
inserted between port \#2 of and BNC power divider as a delay line for the preprocess. The delay
length corresponded to one virtual node width of 50.9 ns. In this way, the serial input stream \( U(t) \)
described in Eq. (3) was prepared. The processing network was entirely shared through these two
inputs, and target value sets were given independently. In these experiments on benchmark tasks,
the output weights were trained during 2000 points, and evaluations were performed in the following
1000 points. The performance was evaluated based on the normalized mean square error (NMSE)
defined with Eq. (6):

\[
NMSE = \frac{\text{ave}[y_{\text{target}}(t) - y(t)]}{\text{var}(y_{\text{target}}(t))}
\]

(6)

\( \text{ave} \) and \( \text{var} \) stand for the average and the variance, respectively.

Figure 5 shows examples of raster plots, output signal and calculation errors for SF time-series
prediction task, NARMA10, and both tasks in parallel at input voltage condition of 0.7 V, respectively.
In a parallel calculation, the input signal voltage was provided using the sum of slightly delayed SF
time-series data and NARMA10 input, as shown in Fig. 4(b), with an amplitude simply reduced to a
half. In this experiment, the input voltage \( V_{\text{in}} \) was limited below 1 V for the following reason. It
was required to limit the MZ voltage below the \( V_{\pi} \) in order to suppress a bifurcation or a chaotic mode [13].
Fig. 5. Time transients of internal node state, output, and calculation error of time-series prediction; NARMA10 task with single input; and mixed tasks, respectively.

Fig. 6. Bifurcation diagram and estimated Lyapunov exponents of this system.

In this experiment, maximum value of the feedback signal intensity voltage $V_{fb}$ was normalized by maximum value of $V_{in}$. Then, input voltage $V_{in}$ was set below 1 V based on Eqs. (7) and (8).

\[ V_{in} + V_{fb} < \frac{V_{pp}}{2} < \frac{V_{p}}{2} = 2V \]  \hspace{1cm} (7)

\[ V_{in} = V_{fb} \]  \hspace{1cm} (8)

Figure 6 shows a bifurcation diagram of the optical delayed feedback systems. A bifurcation mode was observed around relative feedback intensity of 1.0 and chaotic mode was occurred around 1.3. Flat upper limit of the output was determined by a saturation level of the voltage amplifier in Fig. 2(a). In addition, a pseudo-Lyapunov exponent $\lambda$ were estimated to be $-0.001$ and $+0.007$ at relative feedback intensity of 1.0 and 1.3, respectively [20]. Figures 7(a) and (b) show the experimentally obtained calculation error as a function of feedback signal intensity at various input signal amplitudes, when $u_1(t)$ and $u_2(t)$ were input as shown in Fig. 4(b). Figure 7(c) is also parallel calculation error of two different NARMA10 tasks with input signals of $u_2(t)$ and $u_3(t)$. Though the NARMA10 task was not sensitive to another input signal, the SF time-series task performance was degraded in parallel processing.

Absolute calculation errors of the SF time-series prediction task were shown in Figs. 5(c) and (i) with green for comparisons. In terms of errors at an abrupt intensity change point around 2200–2230, the magnitudes of error were in the same range. However, in the periodic operating condition from 2100 to 2200, the calculation errors increased when parallel signals were processed. The SF
Fig. 7. Calculation errors of time-series prediction and NARMA10 task at various optical reservoir condition.

Fig. 8. Numerical calculation errors of time-series prediction and NARMA10 task by using 100-node ESNs.

time-series task is considered to be a noise sensitive task [11] because the required outputs are a quasi-continuous value. Thus, the random input signals of NARMA10 caused a negative effect on the SF time-series prediction task. Figures 8(a)–(c) show calculation results of same benchmark tasks with the 100-node conventional ESNs [1] for comparison. In this calculation, an average and a standard deviation of nonzero matrix elements were 0 and 0.33, respectively. Calculation errors were plotted as a function of spectral radius (SR). The SR was roughly equivalent to the relative feedback signal intensity shown in Figs. 7(a)–(c). Input scaling was also relatively changed in accordance with the physical reservoir. These performances indicated a similar tendency compared to those of experimental results. Figures 8(d) and (e) show NMSE depending on the number of ESNs node at fixed input and SR condition. These results showed that more accurate performance can be expected by increasing the number of node for both tasks. The SF time-series task showed clear improvement by ESNs simulations compared to the experimental results. It is also considered that the noise sensitive analog time traces were affected by fluctuated node states due to noise in the system. Calculation performance of NARMA10 task seemed severely dependent on the feedback signal. Appropriate
feedback intensity was observed slightly less than those of bifurcation state and the error seemed to increase rapidly around chaotic state as shown in Fig. 8.

For a NARMA10 task, the NMSE of single and parallel processing were nearly the same value, while the NMSE of the time-series prediction task showed significant degradation with the parallel processing method as shown in Figs. 7. The memory capacity (MC) was also investigated to understand the feedback signal effect of the parallel NARMA10 task. The MC evaluates the ability of a dynamical system to hold past inputs in a recurrent network [21]. The MC was obtained using Eqs. (9) and (10). In the MC task, the input signals used for NARMA10 task were applied and only target values were replaced by \( y_k(t) \) shown in Eq. (10).

\[
MC = \sum_{k=1}^{N} [1 - \text{NMSE}(y_k(t))] \quad (9)
\]

\[
y_k(t) = u(t - kT) \quad (10)
\]

Figures 9(a) and (b) show the relationship between NMSE of NARMA10 versus MC against same random input, namely, x-axis values in Fig. 9(a) correspond to y-axis values shown in Fig. 7(b). The closed dots indicate parallel processing results with input signals of \( u_1(t) \) and \( u_2(t) \), while the open circle shows a single processing result with input signal of \( u_2(t) \) for comparison. A clear trend emerged in which the calculation error was reduced by at least an NMSE of 0.5 or less. Smaller NMSEs were obtained when the MC was around 8. Figure 9(b) shows the MCs depending on delay time step \( k \), as expressed in Eq. (9). The solid and dashed line show MC curve with NMSE at 0.24 and 0.68 in Fig. 9(a), respectively. Saturated values of the curve correspond to the MC indicated in Fig. 9. Though the solid line, which had a MC of more than 7, still continued to increase at \( k = 10 \), the MC curve shown with dashed line reached a saturation state. These results corroborate that the NARMA10 task requires a 10th order history, as shown in Eq. (5).

These results indicate the possibility of a parallel computation of an independent task with the same physical calculation resource as a recurrent network. Room still exists to improve the performance of those bench mark tasks, such as by modifying the mask process. An important approach for an accurate emulation should be to know the characteristics of the given data and target task [22].

4. Integrative odor analysis for identification

4.1 Olfaction data and integrative task

In this section, we report the processing of odor information as integrative analysis to recognize spatiotemporal and various kinds of chemicals [23] with RC. Successful olfaction is thought to be the most difficult because odor information involves a wide variety of chemical substances. Supposing that we set a clear target, for example, a specific one-gas toxic detection, an appropriate sensor and its performance can be simply selected or developed, focusing several characteristics such as sensitivity, power consumption, environmental tolerance, size, or cost. However, the human olfactory
sense does not focus on some predetermined molecules. The olfactory receptors of living things have comprehensive sensitivity, which means wide sensitivity against various chemical substances but low selectivity. In other words, the biological odor signals can be considered to contain ambiguity or irrelevant information. Humans reportedly have several hundred kinds of comprehensive receptors and can discriminate between more than 1 trillion kinds of odors in their brain [24]. Odor identification algorithms are mainly classified using two methods. One is statistical processing, and the other is neural network analysis including RC.

In this analysis, we used a published database [23] for odor signals. The task was component identification and quantification of gas mixtures. We utilized parallel processing by arranging two tasks, which involved quantifying the methane and ethylene concentration, respectively. The details of the experiment for data collection are described in Ref. [23]. The data were generated from a 16 metal-oxide (MOX) sensor array in response to dynamic gas mixtures in a flow chamber. Time series of two setting points for the gas-flow controller and 16 sensor outputs are shown in Figs. 10(a)–(f). Although all sensors responded to both methane and ethylene, the sensitivity and dynamics differed. For example, sensors #1–4 were sensitive to ethylene, and the response speed was relatively low, whereas sensors #5–8 showed lower sensitivity to methane rather than ethylene. Sensors #9–12 and sensors #13–16 had high sensitivity against ethylene and methane, but these properties were not the same. These are regarded as artificial olfactory with comprehensive sensitivity, so an integrative analysis is a reasonable approach.

The integrative analysis of the odor information was examined using same optical reservoir system described in the previous section. In this integrative odor analysis, reservoir input stream $U(t)$ was prepared by the control PC in advance. Figures 10(g)–(i) show enlarged example of signal input procedures. Each sensor signals $u_n(t)$ ($n = 1, 2, \cdots, 15, 16$) were modulated 100-step square wave $u_n(t)$ and then, the total signal stream $U(t)$ was prepared by the control computer based on the Eq. (3).

Output weight connections were tuned to minimize the differences between the reservoir outputs and the setting points of gas-flow meters in Figs. 10(a) and (b). Figure 11 shows time profiles of target values, corresponding optical reservoir outputs and ESNs outputs for comparison. Outlines of the experimental time profiles were similar to those of ESN. However, accuracy of the experimental result was degraded compared with those of ESNs. Especially false-positive results were sometimes found, for example methane concentration at around the time of 21000 sec. It is considered that the reason for the large error in the experiment was derived from trade-off between output dynamic range and noise in the system. Training and calculation results were evaluated with the mean absolute error (MAE) shown in Eq. (11). $y_{target}$ is setting points of gas-flow meter. In this experiment, the
number of calculation points was around 4000 and was limited by the memory size of the measuring equipment, such as the AWG and the readout oscilloscope shown in Fig. 2(a).

\[
MAE = \text{ave}\left[\sqrt{(y_{\text{target}}(t) - wx(t))^2}\right]
\]  

(11)

4.2 Feedback signal effect

Figure 12(a) shows the MAE depending on the normalized feedback signal intensity at a test phase in the total long time span. In a case without the feedback, the electrical connection between the tunable attenuator and power-divider in Fig. 2(a) was terminated. No significant difference occurred in the MAE with respect to the feedback ratio. However, Fig. 12(b) shows an enlarged illustration of the identification results of methane and ethylene with and without feedback signals. The output signal of ethylene without signal feedback had a large delay (70–80 sec) compared with the target value because of the response delay of the gas sensors shown in Fig. 10, especially sensor #1–4. With signal feedback, the delay of the output signal was reduced to around 20 sec. Figure 13 shows examples of internal node activities. Dashed and solid lines compare the temporal response in terms of the feedback signal. By constructing the feedback loop shown in Fig. 2(a), the response time shown in orange and green was increased, while the node response with red was delayed. With the feedback signal, the time-series diversity can be enhanced in the network, contributing to odor information analysis in a real environment. These results indicate that the delayed feedback optical system also has a slow-response compensation effect [23] of RC, which calculates the signal change of the short duration in a recurrent network. For example, real-time and sequential analysis will be powerful tools
Fig. 13. Internal node dynamics of the optical reservoir with and without a feedback loop.

Fig. 14. (a) Normalized electrical/optical transform function of a MZ modulator at various bias voltages, (b) identification errors depending on feedback coefficient.

Fig. 15. Time transients of the virtual node state at various bias voltage condition.

for fast detection of a moving odor source and enable quick escape from dangerous gases.

4.3 Nonlinearity
The dependence of identification on the nonlinearity of the node was investigated. The nonlinear properties of the MZ modulator were tuned by changing the operation voltage. In this section, bias voltage $\phi$ in Eq. (2) was shifted as shown in Fig. 14(a). The input signal amplitude was fixed to $V_{pp} = 1.1\, \text{V} \quad (V_{in} = 0.55\, \text{V})$ in each instance. The electrical-to-optical conversion property of the MZ modulator was almost linear at a bias voltage of $0\, \text{V}$. However, at a bias voltage of $-1\, \text{V} \quad (+1\, \text{V})$, the lower (higher) voltage side of the conversion characteristics acts as nonlinear conversion processing on the input signal. Figure 14(b) shows the calculation error depending on the bias voltage. At a bias of $0\, \text{V}$, values for the error are higher than those with other bias conditions. Figure 15 shows raster plots of node activity. These results indicate that the nonlinear conversion property of the
single node reservoir with delayed feedback is an important factor for highly accurate odor analysis from multiple sensor arrays with comprehensive sensitivity. The nonlinearity should improve the identification outputs because it increased the diversity of the 100 virtual node conditions. As a result, accurate emulation should be improved, enabling the identification errors to be reduced with a MZ modulator bias condition of $\pm 1 \text{V}$.

5. Conclusion

We implemented a parallel task into an optical delayed feedback system as a RC framework and simultaneously calculated an independent task and an integrative task with the same configuration. One task that processed random values was not influenced by another task during an independent two-task test. However, the calculation results tended to be affected by other signals during a specific task receiving deterministic time-series signals. For the integrative identification task, recurrent signals and a nonlinearity transformation were found to improve the identification performance. The reason for the improvement was induced diversity of internal node activity. These results suggest the flexibility of RC with delayed feedback systems for various tasks. This parallel computation technology can be applied for multi-modal sensing and will be helpful in solving complex problems in the real world.

Acknowledgments

We would like to thank Prof. Hirokazu Takahashi and Prof. Kohei Nakajima of the University of Tokyo for their helpful advice.

References


