The 70th Japan Society of Applied Physics Spring Meeting 2023
Highlighted Presentations Press Release

April 18, 2023

Reservoir computing utilizing spin waves: enhancement of computational performance through a practical approach for on-chip devices

Demonstrating the potential of spin-wave-generating devices for high-performance physical reservoir computing

Graduate School of Engineering, The Univ. of Tokyo[1], International Research Center for Neurointelligence, The Univ. of Tokyo[2], Graduate School of Information Science and Technology, The Univ. of Tokyo [3]
Ryosho Nakane[1], Akira Hirose[1], Gouhei Tanaka [1, 2, 3]

Summary

- **In spin-wave-based reservoir computing developed by the research group, reservoir output vectors with high dimensionality were successfully generated using a single reservoir device on a chip by manipulating physical dynamics under different external magnetic fields.**

- **Numerical experiments on a benchmark task of nonlinear time-series prediction demonstrated that the computational performance of the spin-wave-based reservoir computing system is comparable to that of other state-of-the-art physical reservoir computing systems.**

- **This strongly suggests that spin-wave-based reservoir computing on a chip is promising for purpose-specific edge computing.**

A research group led by Ryosho Nakane at the University of Tokyo established a methodology for improving computational performance without increasing the number of device output terminals. They employed spin-wave-based reservoir computing, which has been undergoing research and development, and performed a challenging benchmark prediction task using multilevel input/output data in a physics-based numerical simulation assuming a feasible device structure. Detailed analyses confirmed that device output signals in response to a time-series input signal change with an external magnetic field through the change in the spin dynamics. Reservoir output vectors were obtained by collecting output signals under three external magnetic field conditions and used for reservoir computing. In a prediction task with a 10th-order nonlinear auto-regressive moving average model (NARMA-10), it was found that the computational performance of the proposed system is comparable to the best one reported for other state-of-the-art physical reservoir computing systems. Thus, they succeeded in generating reservoir output vectors with high dimensionality using a single reservoir device for computational performance improvement. Their achievement brings reservoir computing hardware, which is currently the focus of accelerated research, one step closer to practical technology.

Details

Emerging Machine Learning Scheme: Reservoir Computing

Artificial intelligence (AI) has experienced remarkable development in recent years, while it is supported by large-scale computation through machine learning. In particular, deep learning is a computational scheme that attempts to emulate the performance of the human brain with leveraging learning process (also referred to as training process). It is generally demonstrated using mathematical models called artificial neural networks (ANNs) that are inspired by neural circuits of the human brain. Reservoir computing is currently attracting much attention as a unique computational scheme in AI-related technology.

Reservoir computing uses an information “reservoir”, which is a distinctive feature. A typical example of reservoir computing model is the echo state network (ESN) that was originally derived from widely-used ANNs called recurrent neural networks (RNNs). Like RNNs, the ESN has three parts: an input part, a reservoir part (intermediate part), and a readout part. An important note is that the intermediate part in the ESN does not have adjustable connection weights between neurons, namely, those connection weights are fixed.

In typical RNN systems, all the connection weights, including those in the intermediate network, are iteratively adjusted during the training phase to minimize the error between the system output and the desired output. Hence, their physical implementations require high cost, e.g., hardware assembly with rewritable memory storage for the weight values. Additionally, such RNN systems have difficulties in adjusting the training algorithm and suppressing power consumption attributed to enormous times changes in weight values. In ESN systems, on the other hand, connection weights are adjusted only for ones between the intermediate and output parts. Because the optimum weight values for a target task are determined by a simple training algorithm, ESN systems have an advantage of requiring less computational process and hardware resource, thereby potentially less power consumption than typical RNN systems.

Nevertheless, it does not straightforwardly mean that ESN systems for reservoir computing can fully replace RNN systems for deep learning. The superiority or inferiority of these two computational schemes depends mainly on priority specifications because of a trade-off between versatility and training cost.

Deep learning allows subtle optimization thanks to a large number of adjustable connection weights. Whereas it requires considerable hardware resource, it can have versatility. On the other hand, reservoir computing has an extremely smaller number of adjustable connection weights just between the intermediate and output parts. Hence, it requires reasonable hardware resource but it has less versatility. Those features are important to find advantages of reservoir computing. “Less versatility does not mean reservoir computing is a poor-performance scheme. We are conducting research for implementing reservoir computing in society, not as a general-purpose computing, but as a purpose-specific and low-power-consumption computing with high performance, where hardware resource is limited," says Ryosho Nakane of the Graduate School of Engineering at the University of Tokyo.

To distinguish from reservoir computing on software, reservoir computing using physical dynamics (physical devices) in the reservoir part is called physical reservoir computing.

Physical reservoir computing has a technical merit that it can be implemented with a feasible on-chip device. In addition, it can efficiently perform information processing of unstructured time-series data, such as temporal sensing data. Hence, “We are working on a study to improve computational performance in physical reservoir computing to develop infrastructure for advanced Internet-of-Things (IoT) society while taking its advantages,” says Nakane.

Tackling the trade-off problem between device feasibility and computational performance

Dr. Nakane and his research group conducted physics-based simulations on a feasible on-chip device. One of their goals was to establish a methodology that can improve computational performance without losing robust feasibility of a device, and another was to achieve high computational performance in a challenging benchmark task using multilevel input/output data.

A successful reservoir computing requires "nonlinearity" (nonlinear input-output characteristics through the reservoir), "high dimensionality" (high dimensionality in a reservoir output vector extracted from diverse output signals), and "short term memory" (fading input information in output signals). To realize these three with a very simple device structure, the virtual node method has been widely used, because it can conveniently generate a reservoir output vector through sampling of a single time-series output signal. However, this method presumably has a limitation in improving computational performance, from a technical perspective.

This research group has been developing spin-wave-based reservoir computing schemes, with focusing on spin (magnetization) dynamics in a continuous magnetic film as a physical reservoir because of its rich nonlinear physical phenomena. In this study, external magnetic fields with different magnitudes were applied, aiming at achieving high dimensionality to enhance the computational performance.

“There is a general trade-off between device feasibility and computational performance,” says Nakane. The ideal physical reservoir device for high computational performance has ”many-output terminals (act as output nodes)”, to generate diverse output signals in response to a single time-series input sequence. However, the greater the number of output nodes, the greater the number of output terminals needed in a device. This reduces device feasibility.

The virtual node method is another root to avoid this trade-off. This method achieves “one-input node, many-output nodes” even using a “one-input terminal, one-output terminal” device, namely, it virtually increases the number of output nodes through special processing. However, as the sampling time of a time-series output signal is shortened to increase the number of output nodes, the time-domain waveforms at output nodes tend to be similar with each other, which does not lead to higher dimensionality in a reservoir output vector. This can be a limiting factor in computational performance.

“We focused on the physical dynamics that determines the characteristics of a reservoir device. By combining output signals generated under different physical conditions, we could obtain a reservoir output vector with high dimensionality from a single reservoir device,” says Nakane.

In numerical experiments, external magnetic fields with three different magnitudes were applied to their spin-wave-based device, and the corresponding output signals were obtained. Through detailed analyses, they successfully confirmed large variations in the spin dynamics using the magnetic field (Figure 1).

Highest computational performance among on-chip physical reservoir computing

Furthermore, Nakane and his colleagues evaluated whether a reservoir output vector obtained from the collected output signals actually enhances the computational performance. According to Nakane, anyone can guess that such procedure would increase the dimensionality, but the most important aspect is whether the computational performance actually improves.

Nakane and his group performed spin-wave-based reservoir computing using a benchmark prediction task with a 10th order nonlinear auto-regressive moving average model (NARMA-10) (*1). NARMA-10 is a mathematical model generating a time-series nonlinear signal that is governed by inputs and response outputs (reservoir output node states) in the past 10 steps. The obtained computational performance (*2) is the highest among on-chip physical reservoir computing, and moreover, it is comparable to those of the other state-of-the-art demonstrations (Figure 2).

“The results strongly suggest that spin-wave-based reservoir computing can approximate complex dynamic behavior in the real world, such as those occur in nature and economics,” says Nakane.

Physical reservoir computing with an on-chip device has a high potential for practical applications in purpose-specific edge computing because of their strengths in high adaptability through real-time training, low power consumption, and high performance for time-series data processing. They are potentially useful to develop infrastructure for advanced IoT society in combination with sensors: highly-efficient traffic of information and telecommunication networks, sustainable social infrastructure, and smart factories.

Annotation

*1 **NARMA(Nonlinear Autoregressive Moving Average):** A nonlinear mathematical model that determines future outputs based on an input and outputs in the recent past. It is useful for approximate modeling of the dynamics of complex dynamical phenomena, such as ones occur in nature and economics. Furthermore, it has been widely used as a benchmark for performance evaluation in reservoir computing.

*2 Under the number of effective output nodes Ny = 360, normalized mean squared error NMSE = 0.042 in the prediction task with a 10th order nonlinear auto-regressive moving average model (NARMA-10).

Figures

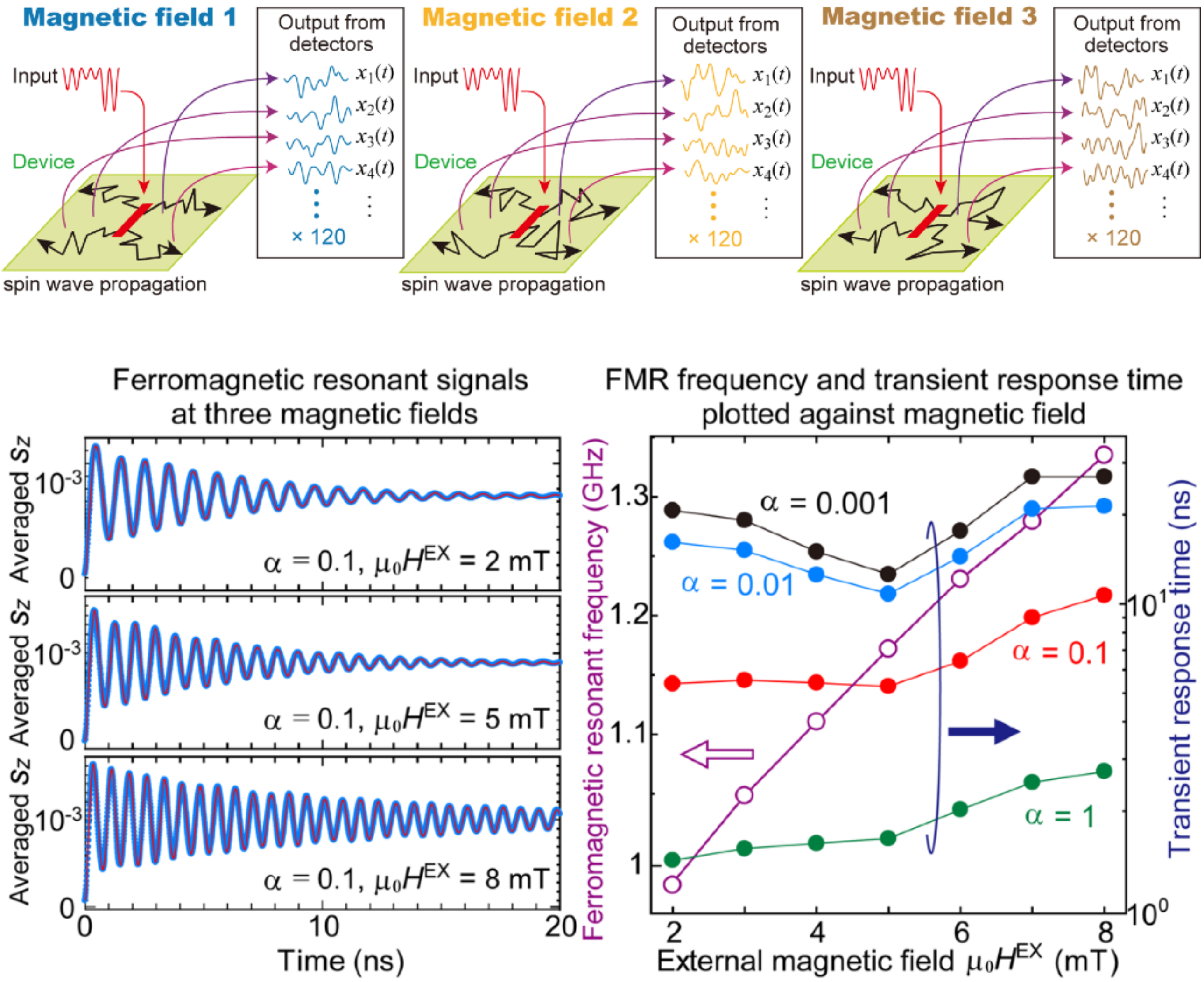

Figure 1 (Upper) Image of output waveforms under different magnetic fields. (Bottom) Characterization of the spin dynamics with an external magnetic field (2-8 mT), where ferromagnetic resonant frequency and relaxation time were estimated to be 0.98–1.36 GHz and 5–10 ns, respectively, under α = 0.1 condition.

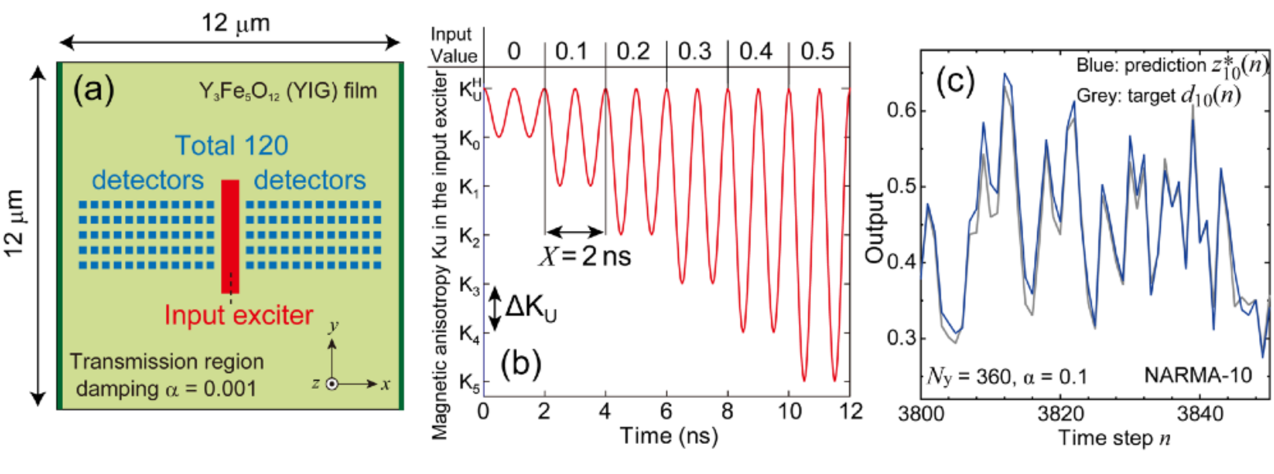

Figure 2 (a) Top view of spin-wave-based device used for numerical experiments. (b) Six-valued input representation as the amplitude of a 1-GHz carrier wave. (c) NARMA-10 signal in the testing phase: target signal (gray), predicted signal (blue).