



# Machine Learning With Observers Predicts Complex Spatiotemporal Behavior

George Neofotistos<sup>1,2\*</sup>, Marios Mattheakis<sup>1</sup>, Georgios D. Barmparis<sup>2</sup>, Johanne Hizanidis<sup>2</sup>, Giorgos P. Tsironis<sup>1,2</sup> and Efthimios Kaxiras<sup>1,3</sup>

<sup>1</sup> School of Engineering and Applied Sciences, Harvard University, Cambridge, MA, United States, <sup>2</sup> Department of Physics, University of Crete, Heraklion, Greece, <sup>3</sup> Department of Physics, Harvard University, Cambridge, MA, United States

## OPEN ACCESS

### Edited by:

Joseph Betouras,  
Loughborough University,  
United Kingdom

### Reviewed by:

Pantelis Rafail Vlachas,  
ETH Zürich, Switzerland  
Alexandre M. Zagouk,  
Loughborough University,  
United Kingdom

### \*Correspondence:

George Neofotistos  
neofotistos@physics.uoc.gr

### Specialty section:

This article was submitted to  
Quantum Computing,  
a section of the journal  
Frontiers in Physics

**Received:** 10 December 2018

**Accepted:** 12 February 2019

**Published:** 05 March 2019

### Citation:

Neofotistos G, Mattheakis M,  
Barmparis GD, Hizanidis J,  
Tsironis GP and Kaxiras E (2019)  
Machine Learning With Observers  
Predicts Complex Spatiotemporal  
Behavior. *Front. Phys.* 7:24.  
doi: 10.3389/fphy.2019.00024

Chimeras and branching are two archetypical complex phenomena that appear in many physical systems; because of their different intrinsic dynamics, they delineate opposite non-trivial limits in the complexity of wave motion and present severe challenges in predicting chaotic and singular behavior in extended physical systems. We report on the long-term forecasting capability of Long Short-Term Memory (LSTM) and reservoir computing (RC) recurrent neural networks, when they are applied to the spatiotemporal evolution of turbulent chimeras in simulated arrays of coupled superconducting quantum interference devices (SQUIDs) or lasers, and branching in the electronic flow of two-dimensional graphene with random potential. We propose a new method in which we assign one LSTM network to each system node except for “observer” nodes which provide continual “ground truth” measurements as input; we refer to this method as “Observer LSTM” (OLSTM). We demonstrate that even a small number of observers greatly improves the data-driven (model-free) long-term forecasting capability of the LSTM networks and provide the framework for a consistent comparison between the RC and LSTM methods. We find that RC requires smaller training datasets than OLSTMs, but the latter require fewer observers. Both methods are benchmarked against Feed-Forward neural networks (FNNs), also trained to make predictions with observers (OFNNs).

**Keywords:** prediction, machine learning, chimera state, graphene, long short-term memory, reservoir computing, branched flow

## INTRODUCTION

Predicting the state of complex, non-linear dynamical systems as a function of time is an important problem of great practical utility. Recent advances of artificial neural networks and machine learning (ML) methods have made possible significant applications in science, industry, and technology [1–15], with reliable prediction comprising one of the most promising areas of research. In this report, we investigate the long-term forecasting capability of two widely used ML methods, the Long Short-Term Memory (LSTM) and the reservoir-computing (RC) recurrent neural network architectures, to predict the spatiotemporal evolution of two distinct complex dynamical phenomena: (i) multi-clustered turbulent chimera states, that is collective, self-organized patterns of coexisting coherence and incoherence in coupled oscillators systems; and (ii) the onset of branching singularities in electronic flow in two dimensional disordered materials. These two phenomena represent opposite ends in complex spatiotemporal evolution with chimeras relating to dynamic self-organization and branching relating to the stochastic onset of singular motion.

LSTM networks [4] have proven to be successful in predicting sequence-involving tasks such as speech recognition [5], machine translation [6], and human dynamics [7] among others. While they have the ability to learn and reproduce long, isolated sequences, they do not seem to provide robust predictions in complex physical systems that involve chaotic behavior or to capture dependencies between multiple correlated sequences [8–14]; in many cases they are outperformed by simpler methods, such as residual multilayer perceptrons [8], while in other cases their predictions diverge [10, 14]. Recently, hybrid LSTM models have been proposed to improve forecasting in applications ranging from chaotic systems [8–10], to pedestrian trajectories [11], particle tracking in high-energy physics [12], and anomaly detection in the post-mortem time series of superconducting magnets [13]. For example, in the method of predicting human trajectories by “social LSTMs” [14], one LSTM network is assigned to each pedestrian, but in order to improve the prediction, each LSTM network is not independent of all other LSTMs, but are connected to those corresponding to nearby pedestrians’ sequences by the introduction of a “social” pooling layer which allows the LSTMs of spatially proximal sequences to share their hidden-states with each other. In another approach, aiming at improving the forecasting of higher-dimensional chaotic systems, a hybrid architecture extends the global LSTM applied to the system with a mean stochastic model (MSM-LSTM) in order to ensure convergence to the invariant measure [10].

Reservoir computing (RC), first proposed by Jaeger and Haas [15], comprises a linear input layer, a recurrent non-linear reservoir network and a linear output layer. This approach has recently been applied to inference problems in chaotic systems [1, 16, 17]. Specifically, “observers” trained by RC (the reservoir observers), provide continual “ground truth” system state measurements as input to the prediction method, deducing the state of the chaotic dynamical system as a function of time from the limited number of the concurrent system state measurements. At each time step, RC estimates the desired unmeasured variables from the measured variables and predicts the evolution the physical system. Lu et al. [3] have recently demonstrated the effectiveness of the method by applying it to the Rössler system, the Lorenz system, and the spatiotemporally chaotic Kuramoto–Sivashinsky equation, carefully pointing out that the method addresses the inference of unmeasured state variables rather than their prediction.

## INTRODUCING A NEW METHOD: “OBSERVER LSTM” (OLSTM)

LSTM networks have proven successful in learning and generalizing sequential tasks from isolated sequences, such as handwriting and speech. Inspired by this success, we first considered a model with a single LSTM network assigned to each system node, which is independent of all other LSTMs. The prediction error for this approach turned out to be very large. This is not surprising since chimera states are collective phenomena and the simplistic use of one LSTM network per node, independent of all others, does not capture well the

interaction between the nodes; the sequences of different nodes are not isolated but correlated.

In order to address the independent LSTMs’ limited ability to capture dependencies between multiple correlated sequences, we propose and demonstrate a new method, which we call “Observer LSTM (OLSTM),” based on the extension of the notion of “reservoir observers” to the LSTM networks. In the OLSTM method, we assign one LSTM to each (non-observer) system node but also assign “observer” status to certain system nodes (“LSTM observers,” taken at equidistant positions for simplicity) which provide continual “ground truth” measurements as input to the prediction method. We demonstrate that their presence even in small numbers (of order <10% of the total number of nodes) greatly improves the long-term data-driven forecasting capability of the LSTM networks and provides the framework for a consistent comparison between the RC and LSTM methods.

Time-series data are used to train each network while no knowledge of the underlying system equations is required. Each individual LSTM network is trained by taking as input a number of past values (denoted as  $N_p$ ) for the node at hand, plus the ground-truth values provided by all observers. Thus, the OLSTM method generates a generalized sequence, which combines the  $N_p$  past values and the time-varying systemic interaction. Using the trained networks, long-term predictions are made by iteratively predicting one step forward for each node, using as input the node’s previous values and the values provided by the relatively very small number of observers.

We study the networks’ long term forecasting capability as a function of the number of observers and as a function of training-set size, and we compare the OLSTM performance with that of “reservoir observers” trained by RC, which utilizes a single (“global”) network for the entire system. We benchmark both methods against a standard Feed-Forward neural network (FNN) method, with the same number of observers (OFNN). We compare quantitatively the networks’ performance by calculating the normalized root mean square error (RMSE) at each time step, for all system nodes, over the predicted time steps, as in [3], Equation (22), without counting observer nodes and training time steps, since they do not contribute to the prediction error.

## RESULTS AND DISCUSSION

We structure our study of predicting complex dynamics in extended physical systems with ML approaches in two parts: the first part concerns chimeras and the second concerns branching in flows. Together, these two cases capture extremes of complex spatiotemporal behavior that severely challenge any method aspiring to predict the long-term behavior. Chimera states challenge the ML methods to predict partial, self organized coherence while the stochastic onset of branching challenges the ML methods to predict stochastic yet singular events.

### Predicting Turbulent Chimeras in Coupled Arrays

Chimera states are collective, self-organized patterns of coexisting coherence, and incoherence in coupled oscillator systems. Following the first discovery of chimeras for

symmetrically coupled Kuramoto identical oscillators in 2002 [18], this counterintuitive symmetry breaking of partially coherent and partially incoherent behavior has received enormous attention. Many recent theoretical works have focused on the study of chimera states in a variety of physical systems such as superconducting metamaterials [19–21] quantum systems [22], and laser arrays [23, 24], to mention only a few. Chimeras have also been studied in models addressing neuron dynamics in hierarchical and modular networks [25, 26]. It has been suggested that chimera states may be related the phenomenon of unihemispheric sleep observed in mammals and birds [27], epileptic seizures [28], and blackouts in power-grids [29]. For finite systems, chimera states are known to be chaotic transients [30], can be stabilized by various newly developed control schemes [31–33]. Following the theoretical predictions, chimera states were experimentally verified for the first time in populations of coupled chemical oscillators [34] and in optical coupled map lattices realized by liquid-crystal light modulators [35] and, later on, in mechanical [36], electronic [37], and electrochemical oscillators [38]; for a recent review see [39].

Chimeras can be stationary or turbulent. Turbulent chimeras have been observed experimentally [35] and have been classified in numerical studies of large arrays of SQUIDs (Superconducting QUantum Interference Devices) and in arrays of lasers with various types of interactions [19–21, 23], among other physical systems. Their actual trajectories are highly non-linear and comprise an immense challenge to predicting their occurrence. In the following sections, we present the long-term prediction results of the ML methods applied on turbulent chimeras in simulated SQUID and semiconductor laser arrays.

SQUID metamaterials constitute a subclass of superconducting artificial media whose function relies both on the geometry and the extraordinary properties of superconductivity and the Josephson effect [40, 41]. Recent experiments on SQUID metamaterials in the superconducting state have demonstrated their wide-band tuneability, significantly reduced losses, and dynamic multistability [42, 43]. The simplest version of a SQUID consists of a superconducting ring interrupted by a Josephson junction [44]; the device is a highly non-linear resonator with a strong response to applied magnetic fields. SQUID metamaterials exhibit peculiar magnetic properties including negative diamagnetic permeability that were predicted both for the quantum [45] and the classical [46, 47] regime.

We investigate the long-term prediction capability of the ML methods under study on turbulent single-headed and double-headed chimeras (“head” stands for incoherent cluster) observed numerically in an array of  $N$  identical  $rf$  SQUIDs coupled together magnetically through dipole-dipole forces. In this system, the magnetic flux  $\Phi_n$  threading the  $n$ -th SQUID loop is given by:

$$\Phi_n = \Phi_{ext} + LI_n + L \sum_{m \neq n} \lambda_{|m-n|} I_m,$$

where the indices  $n$  and  $m$  run from 1 to  $N$ ,  $\Phi_{ext}$  is the external flux in each SQUID,  $\lambda_{|m-n|} = \frac{M_{|m-n|}}{L}$  is the dimensionless

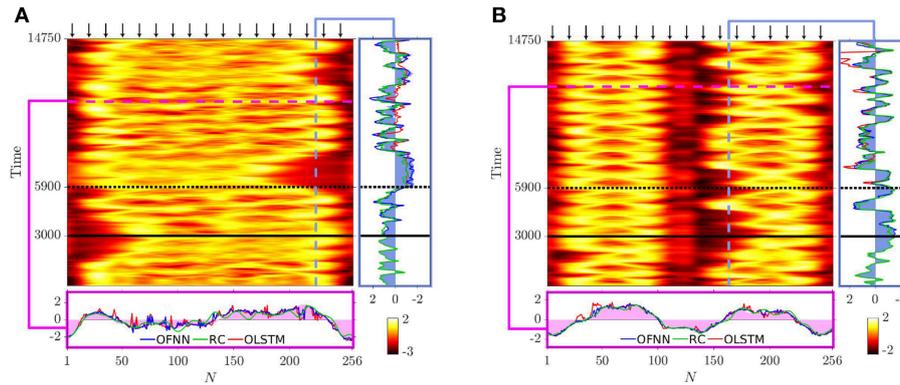
coupling coefficient between the SQUIDs at positions  $m$  and  $n$ ,  $M_{|m-n|}$  being their corresponding mutual inductance, and

$$-I_n = C \frac{d^2 \Phi_n}{dt^2} + \frac{1}{R} \frac{d\Phi_n}{dt} + I_c \sin \left( 2\pi \frac{\Phi_n}{\Phi_0} \right)$$

is the current in the  $n$ -th SQUID given by the resistively and capacitively shunted junction (RCSJ) model [48], with  $\Phi_0$  the flux quantum and  $I_c$  the critical current of the Josephson junctions. Each individual SQUID is a highly non-linear oscillator exhibiting multistability in a certain parameter regime. This is crucial for the occurrence of the chimera states when considering the coupled system. The number of possible states in a SQUID metamaterial is not merely the sum of the combinations of individual SQUID states, since their collective behavior provides many more possibilities. Depending on the choice of initial conditions, various space-time flux patterns may be obtained. “Wild” turbulent chimeras are important for testing the prediction methods because they offer non-trivial dynamical evolution (trajectories) upon which we can test the long-term data-driven (model-free) forecasting of the ML methods. Such chimeras have been generated in Hizanidis et al. [20] (see their Figure 4d for a double-headed chimera with large size of incoherent clusters and Figure 4e for a single-headed chimera, in which the largest part of the metamaterial is occupied by an incoherent cluster with varying size and position in time), where the evolution of the individual SQUID fluxes  $\Phi_n$  is monitored.

We apply RC, OLSTM, and OFNN methods for long-term prediction of the dynamics of these single-headed and double-headed chimeras. Prediction snapshots for the single-headed chimera are presented in **Figure 1A** as insets on the actual (ground-truth) spatiotemporal evolution of the fluxes. In predicting the evolution of this chimera, 17 “observers” were used in the positions marked by the tips of the arrows. The predicted time series for the flux of SQUID #223 is being shown (**Figure 1A**, right vertical inset) as well as the predicted fluxes for all SQUIDs (entire metamaterial) at time step  $t_n = 11,000$  (**Figure 1A**, bottom inset). Similar results for the double-headed chimera are presented in **Figure 1B**, also for 17 observers; the predicted time series for the flux of SQUID #163 is being shown in **Figure 1B** (right vertical inset) along with the predicted fluxes for all SQUIDs at time step  $t_n = 12,000$  (**Figure 1B**, bottom inset). The specific SQUID nodes (#223 and #163) have been chosen in order to depict the long-term forecasting capability of the ML methods in challenging regimes that include *both* coherent and incoherent behavior. We emphasize that in both cases these are very long-time predictions, not just short-term prediction of just a few time steps beyond the training time; specifically, these predictions comprise more than twice (for OLSTMs and OFNNs) and three times (for RC) the training time. Nevertheless, the ML methods produce non-divergent predictions.

In our implementation of OLSTMs we found that a large value of the number of past steps  $N_p$  is not necessary, and in fact even with  $N_p = 1$  the results are quite satisfactory; this implies that the presence of observers more than compensates the need for a short memory to guide the predictions. Thus, our results demonstrate that the “neighborhood” of the node is important; it



**FIGURE 1** | Spatiotemporal plots of single- and double-headed chimeras and predicted time series and fluxes. **(A)** Spatiotemporal plot of a single-headed chimera state, generated in the 1-dimensional SQUID array of Hizanidis et al. [20] Figure 4e, depicting the evolution of the fluxes (values are color-coded), for the one-headed chimera observed in a large array of 256 coupled SQUIDs by dipole-dipole moments, which has been studied numerically for a nonlocal coupling scheme. In predicting the spatiotemporal evolution of this chimera, 17 “observers” have been placed in the positions marked by the tips of the arrows. The thick horizontal black line marks the end of the RC training time, while the dotted horizontal black line marks the end of the OLSTM (and OFNN) training time. Right, vertical inset: predicted time series for the flux of SQUID #223: shaded blue color depicts the actual (ground-truth) data, red line depicts prediction by OLSTM, green line depicts prediction by RC, blue line depicts prediction by OFNN. Bottom inset: predicted fluxes for all SQUIDs (entire metamaterial) at time step  $t_n = 11,000$ : symbols are the same as in the right inset, but with pink depicting the actual (ground-truth) data. **(B)** same as in part (a), but for a double-headed chimera state generated in the 1-dimensional linear SQUID array of Hizanidis et al. [20] Figure 4d. Right, vertical inset: the predicted time series for the flux of SQUID #163. Bottom inset: the predicted fluxes for all SQUIDs (entire metamaterial) at time step  $t_n = 12,000$ .

changes dynamically and the assistance of observers in predicting the future values is more appropriate to process the temporal variation of the “neighborhood” state. The same considerations apply to both SQUID-array and laser-array chimeras.

In laser systems, chimeras were first reported both theoretically and experimentally in a virtual space-time representation of a single laser system subject to long delayed feedback [34]; small networks of globally delay-coupled lasers have also been studied and chimera states were found for both small and large delays. In the present report, we apply the ML methods to predict the trajectories of the turbulent chimeras presented in Figure 4V of the recent work of Shena et al. [23] on multi-clustered chimeras in a large array of coupled semiconductor lasers with non-local coupling. This array is a ring of  $M = 200$  semiconductor lasers of class B. Each node  $j$  is symmetrically coupled with the same strength to its  $R$  neighbors on either side (non-local coupling). The evolution of the slowly varying complex amplitudes  $\mathcal{E}_j = E_j \exp(i\phi_j)$ , with  $E_j$  is the amplitude and  $\phi_j$  the phase of the electric field, and the corresponding population inversions  $N_j$  are given by equations

$$\frac{d\mathcal{E}_j}{dt} = (1 + ia) \mathcal{E}_j N_j + \frac{ke^{-i2C_p}}{2R} \sum_{l=j-R}^{j+R} \mathcal{E}_l,$$

$$\frac{dN_j}{dt} = \frac{1}{T} [p - N_j - (1 + 2N_j) |\mathcal{E}_j|^2], \quad j = 1, \dots, M,$$

where all indices have to be taken modulo  $M$ .  $T$  is the ratio of the lifetime of the electrons in the excited level and that of the photons in the laser cavity. Lasers are pumped electrically with the excess pump rate  $p = 0.23$  [23]. The linewidth enhancement factor  $a$  models the relation between the amplitude and the

phase of the electrical field; a value of  $a = 2.5$  was used, typical for semiconductor lasers. The coupling strength  $k$  and the phase  $C_p$  are the control parameters that are used to tune the collective dynamics of the system. As a measure for phase and amplitude synchronization, the characterization of the phase synchronization of the system can be calculated by means of the Kuramoto local order parameter [49]:

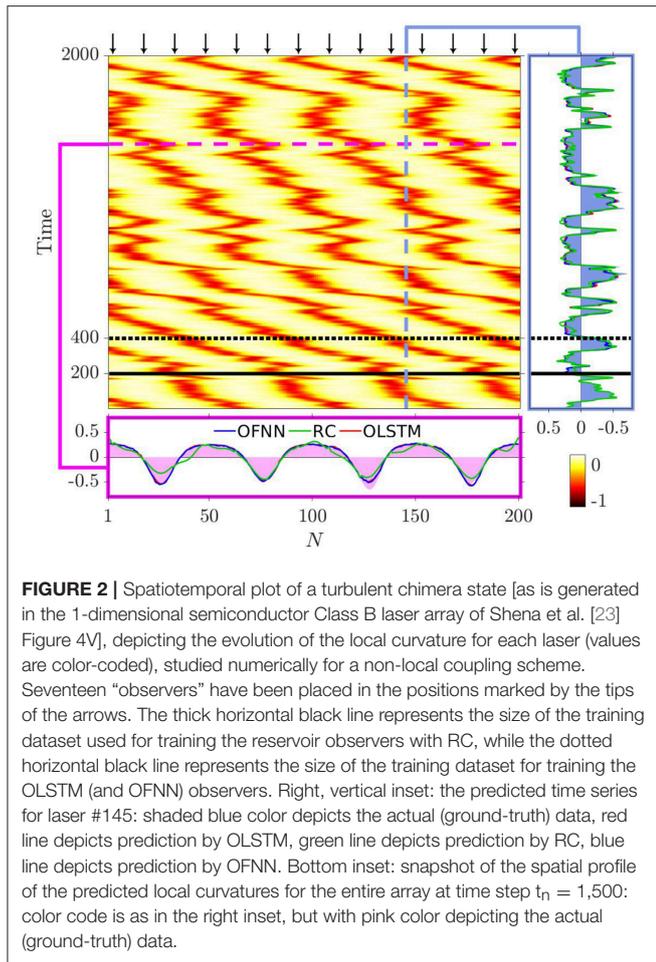
$$Z_j = \left| \frac{1}{2\zeta} \sum_{|l-j| \leq \zeta} e^{i\phi_l} \right|$$

Shena et al. [23] have used a spatial average with a window size of  $\zeta = 3$  elements (a  $Z_j$  value close to unity indicates that the  $j$ -th laser belongs to the coherent regime, whereas  $Z_j$  is closer to 0 in the incoherent part). This quantity can measure only the phase coherence and gives no information about the amplitude synchronization of the electric field, and for this reason the authors have used the classification scheme presented in Kemeth et al. [50] for spatial coherence, based on the calculation of the so called *local curvature* at each time instance, by applying the absolute value of the discrete Laplacian  $|DE|$  on the spatial data of the amplitude of the electric field:

$$|DE|_j(t) = |E_{j+1}(t) - 2E_j(t) + E_{j-1}(t)|, \quad j = 1, \dots, M.$$

In the synchronization regime the local curvature is close to zero while in the asynchronous regime it is finite and fluctuating.

Since our objective is to obtain “wild” turbulent chimeras in order to test the long-term predictions of the ML methods, we have chosen the turbulent chimera of Shena et al. [23] Figure 4V, generated with the choice of parameter values:  $R = 64$ ,  $C_p = 0.4\pi$ ,



$k = 0.225$ ,  $T = 392$ ,  $p = 0.23$ ,  $a = 2.5$ . Prediction snapshots are presented in **Figure 2**, as insets on the actual (ground-truth) spatiotemporal evolution of the local curvature for the flux of laser #145 and for all lasers, for similar choices as in **Figure 1**, that is, with 17 “observers” and at time step  $t_n = 1,500$ . These are extremely long-time predictions, with time horizons almost five times the training time (for OLSTMs and OFNNs) and ten times (for RC) the training time. In this case, as in that of the SQUID system, all ML methods under study produce non-divergent predictions.

In **Figure 3** we present the calculated values of the normalized root mean squared error for long-term prediction (top inset) of the RC, OLSTM, and the OFNN methods, and as a function of the number of observers and the size of the training dataset for a randomly selected time step for the chimera depicted in **Figure 2**; similar results are obtained for all systems. The RMSE values remain, on average, close to  $10^{-1}$ , for time horizons up to 2,000 time steps in the future. This figure also presents graphs of how the RMSE varies as a function of the number of observers (given as percentage of the number of the system nodes), and as a function of the size of the training datasets (given as percentage of the entire ground-truth time series). In all cases, low RMSE values are attained after a minimum number of observers have been included in the system, comprising

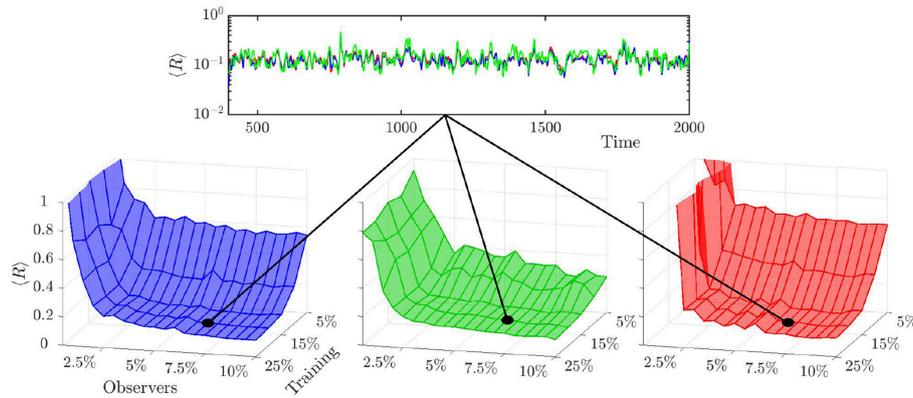
about 5% of the total number of oscillators, and the size of the training data reaches at least 15% of entire dataset. These results demonstrate that, even very small numbers of observers are very important for achieving non-diverging long-term predictions. Furthermore, they show that RC achieves low levels of RMSE with smaller size of training datasets, whereas OLSTM achieves low levels of RMSE with smaller number of observers.

## Predicting Singular Events

Wave focusing due to refractive index variation is a common occurrence in many physical systems, as, for example, in optical media where the index of refraction changes in a statistical way due to small imperfections or distributions of defects in the medium through which the wave propagates. Random spatial variability of the index leads to local focusing and defocusing of the waves and the formation of caustics (or wave “branches”) with substantially increased local wave intensity [51, 52]. Quantum particles like electrons also exhibit wave properties and as a result, electrons traveling in disordered media can form coalescing trajectories and exhibit phenomena similar to wave motion in optical random media. A case in point, where branching may have implications for technological applications, is the ultra-relativistic electronic flow in a two-dimensional (2D) random potential as manifested in graphene and other Dirac solids [52]. In these situations, branching arises in the flow of electrons through a region of inhomogeneous distribution of charge impurities in the substrate, which create a random potential for the electrons [53]. An additional bias voltage is introduced to induce the electronic propagation. Recently, it was shown that the onset of the electronic branched flow is determined by the statistical properties of the random substrate potential, which is quantified through a scaling-type relationship to capture the emergence of branches [54].

It is a formidable challenge to predict singular events like branching in wave propagation or electron flow because of the stochastic nature of the onset of such events. An important question is whether or not ML methods can dissect the stochastic nature of branching and “learn” the interactions that take place among trajectories, thereby providing an accurate detection mechanism for the caustics that mark the onset of branching. We attempt to resolve this issue with results from the RC, OLSTM, and OFNN methods, on singular branched flows in graphene with random potentials. Our results demonstrate that the ML methods we considered can adequately capture the stochastic temporal dependencies of the time series in this prototypical complex dynamical system.

Caustic event prediction in a 2D electron flow is facilitated if we consider one of the spatial dimensions (we refer to it as the “longitudinal” x-direction) as the “time-coordinate,” and therefore, map the stationary phenomenon of caustic formation onto a 1D spatio-temporal dynamical problem. In the framework of this approach, we model the motion of electrons as individual rays whose density matrix is transformed onto a vector of time series with dimension  $N$ , the number of mesh points



**FIGURE 3 |** Normalized RMSE ( $\langle R \rangle$ ) of RC, OLSTM, and the OFNN methods calculated for each time step over all system nodes at each predicted time step (top panel) and over the predicted time steps *and* all system nodes (bottom 3-dimensional plots) as a function of the number of observers (given as percentage of the entire number of system nodes) and the size of the training dataset (given as percentage of the entire ground-truth time series) for the chimera depicted in **Figure 2**. The black dots in each of the 3-dimensional plots represent the normalized RMSE for each ML method, respectively, at a randomly picked time step, calculated overall system nodes *and* previous time steps, but not counting observer nodes and training time steps. Red: OLSTM, green: RC, blue: OFNN. Axes, tick marks, tick labels, and grid lines are the same in all 3-dimensional plots.

of the remaining spatial dimension (the “transverse,” or  $y$ -direction). In the system we studied,  $N = 210$ , spanning a total length of 84 nm [54]. Our goal is to predict the onset of branching in time and its location in the electronic flow. **Figure 4** presents prediction snapshots of the onset of branching in electronic flows in graphene with random potentials, depicting the intensity of the flows, using 10 different “observers” along the “time” axis of the flows at different positions on the transverse axis. These results demonstrate that the RC, OLSTM, and OFNN methods are able to predict well the branching in the electronic flows in this system, even at very long prediction times. In these methods, observers are very important to achieve long-term forecasting capability; they act as real-time sensors that help predict the future values of the flows.

## CONCLUSIONS AND OUTLOOK

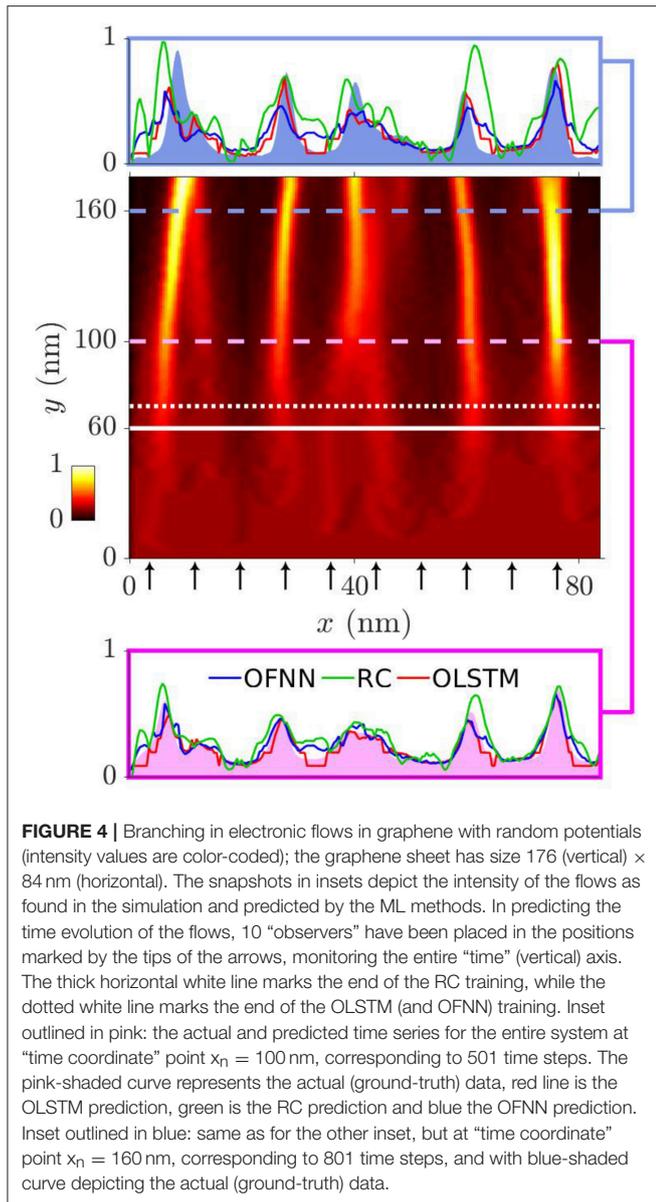
The issue of predicting complex spatiotemporal behavior using ML approaches is one of central importance for their potential applications in the physical sciences and beyond. Here, we attempted to address this issue by considering two distinct prototypical phenomena, viz. partially coherent chimera states and the stochastic onset of branching in 2D wave flows, as these are realized in coupled arrays of SQUIDs or lasers, and in the flow of electrons in graphene with random potentials, respectively. We find that ML approaches like LSTM and RC recurrent neural networks can perform well in predicting complex dynamics in extended physical systems when they involve “observers” that monitor the system evolution throughout its time dimension. The presence of observers is an inherent requirement of the second approach (RC), but not of the first (LSTM). Accordingly, we proposed a new method, which we call “Observer LSTM (OLSTM),” to address the limitations of single, independent

LSTM networks in capturing dependencies between multiple correlated sequences. We have also considered an observer-enhanced Feed-Forward network (OFNN) and tested the long-term prediction performance of the three approaches, OLSTM, OFNN, and reservoir observers trained by RC, on the two difficult problems of turbulent chimeras and 2D branching flows. Our results quantify how the prediction error (root mean squared error, RMSE, of the predicted values) varies as a function of the number of observers and of the size of the training datasets.

We conclude that “observers” comprise *sine qua non* elements for robust data-driven (model-free) long-term forecasting capability of the LSTM and FNN networks, and they provide the framework for a consistent comparison between the LSTM and RC methods. Thus, observer-enhanced ML methods, like OLSTM, acting as high-level “intelligent” interpolation schemes, are capable of successfully predicting the non-linear spatiotemporal evolution of complex dynamical systems. Many issues remain to be evaluated and questions to be answered in establishing the robustness and efficiency of the observer-enhanced approaches. An example of such issues is the extent to which the presence of observers, while enhancing long-term predictability, affects the chaotic behavior of the system. A thorough investigation of such issues is the focus of on-going and future research.

## METHODS

The RC network architecture used in predicting turbulent chimeras comprises of 1,000 reservoir nodes, with spectral radius  $\rho = 1.0$ , average degree  $D = 80$ , scale of inputs weights  $\sigma = 1.5$ , bias constant  $\xi = 0.0$ , leakage rate  $\alpha = 0.9$ , ridge regression parameter  $\beta = 0.5$ , and time interval  $\Delta t = 0.01$ . The RC network applies to the system as a whole (single network



**FIGURE 4 |** Branching in electronic flows in graphene with random potentials (intensity values are color-coded); the graphene sheet has size 176 (vertical)  $\times$  84 nm (horizontal). The snapshots in insets depict the intensity of the flows as found in the simulation and predicted by the ML methods. In predicting the time evolution of the flows, 10 “observers” have been placed in the positions marked by the tips of the arrows, monitoring the entire “time” (vertical) axis. The thick horizontal white line marks the end of the RC training, while the dotted white line marks the end of the OLSTM (and OFNN) training. Inset outlined in pink: the actual and predicted time series for the entire system at “time coordinate” point  $x_n = 100$  nm, corresponding to 501 time steps. The pink-shaded curve represents the actual (ground-truth) data, red line is the OLSTM prediction, green is the RC prediction and blue the OFNN prediction. Inset outlined in blue: same as for the other inset, but at “time coordinate” point  $x_n = 160$  nm, corresponding to 801 time steps, and with blue-shaded curve depicting the actual (ground-truth) data.

architecture). For branching, the RC network used comprises of 3,000 reservoir nodes, with spectral radius  $\rho = 0.9$ , average degree  $D = 50$ , scale of inputs weights  $\sigma = 1$ , bias constant  $\xi = -0.4$ , leakage rate  $\alpha = 0.5$ , ridge regression parameter  $\beta = 0.05$ , and  $\Delta t = 0.05$ . The network applies to the system as a whole (single network architecture).

The OLSTM network architecture used in this study constitutes 400 LSTM cells with RELU activation functions (one hidden layer). A single LSTM network is applied to each (non-observer) system node (SQUID or laser oscillator or electron flow). Similarly, a single fully-connected (dense) Feed Forward network (OFNN), with RELU activation functions, is applied to each (non-observer) system node (SQUID or laser oscillator or electron flow). Optimization for LSTMs during training is performed using the Adam stochastic

optimization method [55] with a learning rate of 0.001 as implemented in *Keras*<sup>1</sup>.

In case of the SQUIDs system (**Figure 1**), the OLSTM and OFNN models were trained with 5,900 time steps (40% of the total 14,750 time steps) and RC with 3,000 time steps (about 20% of the total number of time steps). In case of the Laser Chimeras system (**Figure 2**), the OLSTM and OFNN models were trained with 400 time steps (20% of the total number of time steps) and RC model with 200 time steps (10% of the total number of time steps). Finally, in case of branching (**Figure 4**), OLSTM and OFNN were trained with 352 time steps (40% of the total number of time steps) and RC with 300 time steps (about 34%).

The data of the time series used (chimeras and electronic flows in graphene) was preprocessed as in [3] Equation (15) and smoothed by means of Matlab’s 2D Gaussian Smoothing Kernel with standard deviation  $\sigma = 3.5$  for chimeras and  $\sigma = 2.5$  for the electronic flows in graphene. The time series data have been divided into two separate sets, the training dataset and the validation dataset. The data is stacked in batches (of size 50 data points) in order to form the training (and validation) input and output of the networks. These training batches are used to optimize the parameters of the networks (weights and biases). The training proceeds by optimizing the network weights iteratively for each batch. The training loss function is a weighted version of the root mean square error. When the network parameters have been optimized once for all training data batches, one epoch is completed. After every epoch the RMSE in the validation data set is computed. Training is stopped after 200 epochs and the OLSTM (OFNN) network with the smallest validation error is selected in order to avoid over-fitting.

Each trained OLSTM network is then used to forecast the node’s state in the next time step in an iterative fashion (getting also input from the observers). All the hyper-parameters in OLSTM and OFNN models including the number of training epochs, the number of training batches, the number of neurons and the number of hidden layers are optimized so that they are leading to the smallest RMSE. Similarly, for RC we optimized the parameters with criterion the smallest value for RMSE.

As a comparison measure for the networks’ performance, we use the (normalized) root mean square error calculated at each time step, for all system nodes and over the predicted time steps (unless otherwise specified).

## DATA ACCESSIBILITY

The machine learning library *Keras* in python 3.6 was used for the implementation of the OLSTM and OFNN architectures (utilizing the “Metropolis” Supercomputing Facility of the Center for Quantum Complexity and Nanotechnology of the Physics Department of the University of Crete) while our code written on Matlab was used for the RC architectures.

<sup>1</sup> Available online at: <https://keras.io> (Accessed February 15, 2018).

## DATA AVAILABILITY

The raw data supporting the conclusions of this manuscript will be made available by the authors, without undue reservation, to any qualified researcher.

## AUTHOR CONTRIBUTIONS

GN, MM, and GDB designed and performed research, analyzed data, and wrote the manuscript. JH provided data on chimeras, made contributions to the manuscript, and constructive comments. GPT and EK contributed to the design of research, interpretation of results, and to the writing of the manuscript.

## ACKNOWLEDGMENTS

GPT acknowledges useful discussions with Prof. Edward Ott. GN and GPT acknowledge support by the European Commission under project NHQWAVE (MSCA-RISE 691209). JH acknowledges support by the General Secretariat for Research and Technology (GSRT) and the Hellenic Foundation for Research and Innovation (HFRI) (Code: 203). MM and EK acknowledge partial support from EFRI 2-DARE NSF Grant No. 1542807 and from ARO MURI Award No. W911NF14-0247. We used computational resources on the Odyssey cluster of the FAS Research Computing Group at Harvard University.

## REFERENCES

- Hinton G, Deng L, Yu D, Dahl GE, Mohamed A-R, Jaitly N, et al. Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups. *IEEE Signal Proc Mag.* (2012) **29**:82–97. doi: 10.1109/MSP.2012.2205597
- Silver D, Schrittwieser J, Simonyan K, Antonoglou I, Huang A, Guez A, et al. Mastering the game of go without human knowledge. *Nature.* (2017) **550**:354. doi: 10.1038/nature24270
- Lu Z, Pathak J, Hunt B, Girvan M, Brockett R, Ott E. Reservoir observers: model-free inference of unmeasured variables in chaotic systems. *Chaos.* (2017) **27**:041102. doi: 10.1063/1.4979665
- Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput.* (1997) **9**:1735–80. doi: 10.1162/neco.1997.9.8.1735
- Graves A, Mohamed AR, Hinton G. Speech recognition with deep recurrent neural networks. In: *Acoustics, Speech and Signal Processing (ICASSP) IEEE International Conference.* Vancouver, BC (2013). p. 6645–9.
- Wu Y, Schuster M, Chen Z, Le QV, Norouzi M, Macherey W, et al. Google's neural machine translation system: bridging the gap between human and machine translation. *arXiv.* (2016) arXiv:1609.08144.
- Fragkiadaki K, Levine S, Felsen P, Malik J. Recurrent Network Models for Human Dynamics. International Conference on Computer Vision (ICCV) 2015. arXiv: 1508.00271v2. Preprint, posted September 29, 2015. (2015)
- Maathuis H, Boulogne L, Wiering M, Sterk A. Predicting chaotic time series using machine learning techniques. In: Verheij B, Wiering M, editors. *Preproceedings of the 29th Benelux Conference on Artificial Intelligence (BNAIC 2017).* Groningen (2017). p. 326–40.
- Wan ZY, Vlachas P, Koumoutsakos P, Sapsis T. Data-assisted reduced-order modeling of extreme events in complex dynamical systems. *PLoS ONE.* (2018) **13**:e0197704. doi: 10.1371/journal.pone.0197704
- Vlachas PR, Byeon W, Wan ZY, Sapsis TP, Koumoutsakos P. Data-driven forecasting of high-dimensional chaotic systems with long short-term memory networks. *Proc R Soc A.* (2018) **474**:20170844. doi: 10.1098/rspa.2017.0844
- Hug R, Becker R, Hubner W, Arens M. Particle-based pedestrian path prediction using LSTM-MDL models. In: *2018 21st International Conference on Intelligent Transportation Systems (ITSC).* Hawaii (2018). p. 2684–91.
- Tsaris A, Anderson D, Bendavid J, Calafiura P, Cerati G, Esseiva J, et al. The HEP.TrkX project: deep learning for particle tracking. *ACAT.* (2017) **1085**:042023. doi: 10.1088/1742-6596/1085/4/042023
- Wielgosz M, Skoczen A, Mertik M. Recurrent neural networks for anomaly detection in the post-mortem time series of LHC superconducting magnets. *arXiv.* (2017) arXiv:1702.00833v1.
- Alahi A, Goel K, Ramanathan V, Robicquet A, Fei-Fei L, Savarese S. Learning to predict human behaviour in crowded spaces. In: Murino V, Cristani M, Shah S, Savarese S, editors. *Group and Crowd Behavior for Computer Vision*, 1st ed. Cambridge, MA: Elsevier, Academic Press (2017). p. 183–204.
- Jaeger H, Haas H. Harnessing nonlinearity: predicting chaotic systems and saving energy in wireless communications. *Science.* (2004) **304**:78–80. doi: 10.1126/science.1091277
- Pathak J, Hunt B, Girvan M, Lu Z, Ott E. Model-free prediction of large spatiotemporally chaotic systems from data: a reservoir computing approach. *Phys Rev Lett.* (2017) **120**:024102. doi: 10.1103/PhysRevLett.120.024102
- Pathak J, Lu Z, Hunt B, Girvan M, Ott E. Using machine learning to replicate chaotic attractors and calculate Lyapunov exponents from data. *Chaos.* (2017) **27**:121102. doi: 10.1063/1.5010300
- Kuramoto Y, Battogtokh D. Coexistence of coherence and incoherence in nonlocally coupled phase oscillators. *Nonlinear Phenom Complex Syst.* (2002) **5**:380. Available online at: <http://www.j-npcs.org/abstracts/vol2002/v5no4/v5no4p380.html>
- Lazarides N, Neofotistos G, Tsironis GP. Chimeras in SQUID metamaterials. *Phys Rev B.* (2015) **91**:054303. doi: 10.1103/PhysRevB.91.054303
- Hizanidis J, Lazarides N, Neofotistos G, Tsironis GP. Chimera states and synchronization in magnetically driven SQUID metamaterials. *Eur Phys J Special Topics.* (2016) **225**:1231–43. doi: 10.1140/epjst/e2016-02668-9
- Hizanidis J, Lazarides N, Tsironis GP. Robust chimera states in SQUID metamaterials with local interactions. *Phys Rev E.* (2016) **94**:032219. doi: 10.1103/PhysRevE.94.032219
- Bastidas V, Omelchenko I, Zakharova A, Schöll E, Brandes T. Quantum signatures of chimera states. *Phys Rev E.* (2015) **92**:062924. doi: 10.1103/PhysRevE.92.062924
- Shena J, Hizanidis J, Hovel P, Tsironis GP. Multiclustered chimeras in large semiconductor laser arrays with nonlinear interactions. *Phys Rev E.* (2017) **96**:032215. doi: 10.1103/PhysRevE.96.032215
- Shena J, Hizanidis J, Kovanis V, Tsironis GP. Turbulent chimeras in large semiconductor laser arrays. *Sci Rep.* (2017) **7**:42116. doi: 10.1038/srep42116
- Omelchenko I, Provata A, Hizanidis J, Schöll E, Hövel P. Robustness of chimera states for coupled FitzHugh-Nagumo oscillators. *Phys Rev E.* (2015) **91**:022917. doi: 10.1103/PhysRevE.91.022917
- Hizanidis J, Kouvaris NE, Zamora-López G, Díaz-Guilera A, Antonopoulos C. Chimera-like states in modular neural networks. *Sci Rep.* (2016) **6**:19845. doi: 10.1038/srep19845
- Rattenborg NC, Amlaner CJ, Lima SL. Behavioral, neurophysiological and evolutionary perspectives on unihemispheric sleep. *Neurosci Biobehav Rev.* (2000) **24**:817–42. doi: 10.1016/S0149-7634(00)00039-7
- Andrzejak RG, Rummel C, Mormann F, Schindler K. All together now: analogies between chimera state collapses and epileptic seizures. *Sci Rep.* (2016) **6**:23000. doi: 10.1038/srep23000
- Motter AE, Myers SA, Anghel M, Nishikawa T. Spontaneous synchrony in power-grid networks. *Nat Phys.* (2013) **9**:191–7. doi: 10.1038/nphys2535
- Wolfrum M, Omelchenko OE. Chimera states are chaotic transients. *Phys Rev E.* (2011) **84**:015201. doi: 10.1103/PhysRevE.84.015201
- Sieber J, Omelchenko OE, Wolfrum M. Controlling unstable chaos: stabilizing chimera states by feedback. *Phys Rev Lett.* (2014) **112**:054102. doi: 10.1103/PhysRevLett.112.054102

32. Bick C, Martens EA. Controlling chimeras. *New J Phys.* (2015) **17**:033030. doi: 10.1088/1367-2630/17/3/033030
33. Omelchenko I, Omel'chenko OE, Zakharova A, Wolfrum M, Schöll E. Tweezers for chimeras in small networks. *Phys Rev Lett.* (2016) **116**:114101. doi: 10.1103/PhysRevLett.116.114101
34. Schmidt L, Schonleber K, Krischer K, Garcia-Morales V. Coexistence of synchrony and incoherence in oscillatory media under nonlinear global coupling. *Chaos* (2014) **24**:013102. doi: 10.1063/1.4858996
35. Larger L, Penkovsky B, Maistrenko Y. Laser chimeras as a paradigm for multistable patterns in complex systems. *Nat Commun.* (2015) **6**:7752. doi: 10.1038/ncomms8752
36. Martens EA, Thutupalli S, Fourriere A, Hallatschek O. Chimera states in mechanical oscillator networks. *Proc Natl Acad Sci USA.* (2013) **110**:10563–7. doi: 10.1073/pnas.1302880110
37. Gambuzza LV, Buscarino A, Chessari S, Fortuna L, Meucci R, Frasca M. Experimental investigation of chimera states with quiescent and synchronous domains in coupled electronic oscillators. *Phys Rev E.* (2014) **90**:032905. doi: 10.1103/PhysRevE.90.032905
38. Wickramasinghe M, Kiss IZ. Spatially organized dynamical states in chemical oscillator networks: synchronization, dynamical differentiation, and chimera patterns. *PLoS ONE.* (2013) **8**:e80586. doi: 10.1371/journal.pone.0080586
39. Pannagio MJ, Abrams D. Chimera states: coexistence of coherence and incoherence in networks of coupled oscillators. *Nonlinearity.* (2015) **28**:R67–87. doi: 10.1088/0951-7715/28/3/R67
40. Anlage SM. The physics and applications of superconducting metamaterials. *J Opt.* (2011) **13**:024001. doi: 10.1088/2040-8978/13/2/024001
41. Jung P, Ustinov AV, Anlage SM. Progress in superconducting metamaterials. *Superconduct Sci Technol.* (2014) **27**:073001. doi: 10.1088/0953-2048/27/7/073001
42. Butz S, Jung P, Filippenko LV, Koshelets VP, Ustinov AV. A one-dimensional tunable magnetic metamaterial. *Opt Express.* (2013) **29**:22540–8. doi: 10.1364/OE.21.022540
43. Trepanier M, Zhang D, Mukhanov O, Anlage SM. Realization and modeling of metamaterials made of rf superconducting quantum-interference devices. *Phys Rev X.* (2013) **3**:041029. doi: 10.1103/PhysRevX.3.041029
44. Josephson B. Possible new effects in superconductive tunnelling. *Phys Lett A.* (1962) **1**:251–3. doi: 10.1016/0031-9163(62)91369-0
45. Du C, Chen H, Li S. Quantum left-handed metamaterial from superconducting quantum-interference devices. *Phys Rev B.* (2006) **74**:113105. doi: 10.1103/PhysRevB.74.113105
46. Lazarides N, Tsironis GP. rf superconducting quantum interference device metamaterials. *Appl Phys Lett.* (2007) **90**:163501. doi: 10.1063/1.2722682
47. Strogatz SH. Exploring complex networks. *Nature.* (2001) **410**:268. doi: 10.1038/35065725
48. Likharev KK. *Dynamics of Josephson Junctions and Circuits.* Philadelphia, PA: Gordon and Breach (1986).
49. Omelchenko I, Omelchenko OE, Hövel P, Schöll E. When nonlocal coupling between oscillators becomes stronger: patched synchrony or multichimera states. *Phys Rev Lett.* (2013) **110**:224101. doi: 10.1103/PhysRevLett.110.224101
50. Kemeth FP, Haugland SW, Schmidt L, Kevrekidis IG, Krischer K. A classification scheme for chimera states. *Chaos.* (2016) **26**:094815. doi: 10.1063/1.4959804
51. Metzger JJ, Fleischmann R, Geisel T. Universal statistics of branched flows. *Phys Rev Lett.* (2010) **105**:020601. doi: 10.1103/PhysRevLett.105.020601
52. Mattheakis M, Pitsios IJ, Tsironis GP, Tzortzakis S. Rogue events in complex linear and nonlinear photonic media. *Chaos Solitons Fractals.* (2016) **84**:73–80. doi: 10.1016/j.chaos.2016.01.008
53. Topinka MA, LeRoy BJ, Westervelt RM, Shaw SEJ, Fleischmann R, Heller EJ, et al. Coherent branched flow in a two-dimensional electron gas. *Nature.* (2001) **410**:183–6. doi: 10.1038/35065553
54. Mattheakis M, Tsironis GP, Kaxiras E. Emergence and dynamical properties of stochastic branching in the electronic flows of disordered Dirac solids. *EPL.* (2018) **122**:27003. doi: 10.1209/0295-5075/122/27003
55. Kingma DP, Ba J. Adam: a method for stochastic optimization. In: *Proceedings of the 3rd International Conference on Learning Representations (ICLR 2015).* San Diego, CA (2015).

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer AZ declared a past co-authorship with one of the authors GT to the handling editor.

Copyright © 2019 Neofotistos, Mattheakis, Barmparis, Hizanidis, Tsironis and Kaxiras. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.