Journal of Applied Nonlinear Dynamics 15(2) (2026) 277-290

Journal of Applied Nonlinear Dynamics

https://lhscientificpublishing.com/Journals/JAND-Default.aspx

Machine-learning Prediction of Type-III Instabilities and Spontaneous Energy Localization

Adrian Mittal[1], Fedya Grishanov[2], Anne Pham[3], Noah Lape[2], L.Q. English[2†]

[1] International Research Institute of North Carolina, Sanford, NC 27330
[2] Department of Physics and Astronomy, Dickinson College, Carlisle, PA 17013
[3] Department of Data Analytics, Dickinson College, Carlisle, PA 17013

Submission Info

Communicated by C. Skokos
Received 14 September 2024
Accepted 18 September 2025
Available online 1 April 2026

Keywords

Spontaneous pattern formation
Machine learning
Energy localization

Abstract

Modulational instability of a plane-wave mode is known to lead to spontaneous energy localization in nonlinear lattices without the presence of impurities. Where the energy dynamically localizes in the system is highly sensitive to initial conditions. Here we numerically investigate sine-Gordon-type lattices and first show that spatial smoothing of the observed dynamical variables on the lattice can already substantially improve predictions of the eventual localization site from early-time data traces alone, suggesting that the type-III instability that forms the energy hotspot is sensitive to energetic clusters. We then show that machine learning leads to additional dramatic improvement in prediction accuracy. Finally, we also examine the role of chain impurities in determining the localization site.

1 Introduction

Dynamic energy localization can be observed in many nonlinear oscillating systems [1]. Energy localization is a phenomenon whereby a significant portion of a system's total energy becomes concentrated on just a small segment of the total system. Examples of such spontaneous processes of energy localization can be found across the physical world. A good example was provided by the experimental work on driven micro-mechanical oscillator arrays [2], where dynamic instabilities provided a pathway to eventual energy localization. These paradigmatic findings suggest that such processes could also be at play in large mechanical structures, and indeed nonlinear instability and feedback have been identified as playing a role in, for instance, the 1940 catastrophic collapse of the Tacoma Narrows Bridge [3, 4].

Another large-scale example where nonlinear oscillator patterns and energy localization could pose a major issue is in high-voltage transmission tower-line systems driven by strong wind, as they are characterized by a geometry of closely spaced nodes at the towers and nonlinear interaction via the cables [5]. On a microscopic level, it has been proposed that the propagation of cracks in crystalline matter, and thus fracture, is aided by the activation of nonlinear, spatially localized modes [6, 7].

†Corresponding author.
Email address: englishl@dickinson.edu

ISSN 2164-6457, eISSN 2164-6473/$-see front materials
DOI:10.5890/JAND.2026.06.001

These are just a few examples showcasing the diversity of nonlinear physical systems that can exhibit nonlinear instability and spontaneous energy localization. What connects them is that the formation of "hotspots" often leads to some level of mechanical failure or structural damage. Preventing such instances of catastrophic energy localization, or at least predicting it, is therefore of acute interest.

Spontaneous energy localization in nonlinear connected systems often occurs as a consequence of dynamic instabilities. A normal (plane-wave) mode of a system, when driven into a nonlinear regime, becomes unstable against perturbations of a particular wavelength. In the presence of noise, which contains a broad spectrum of wavelengths, the instability will typically pick out and amplify the one with the largest growth rate. It has long been known that localization can ultimately be achieved via a modulational instability of a plane-wave mode [8,9]. Reference [10] delineates three classes of instabilities based on the wavenumber of the fastest-growing perturbation. It is known that the instability of type III, with its zero-wavenumber amplification, often leads to catastrophic localization.

In this study, we will use the discrete sine-Gordon model as a well-known and paradigmatic system in which to numerically explore pattern formation and energy localization. The sine-Gordon model has been employed in many physical contexts [11]. Along with the KdV and the nonlinear Schrödinger equations, it is a prominent integrable system that supports solitons and kinks [12, 13]. Furthermore, the driven sine-Gordon system displays the phenomenon of pattern formation [14]. It is characterized by an onsite sine-nonlinearity and by inter-site (coupling) linearity. While it has also been applied directly to Josephson junctions [11, 15], its simplest physical realization is the torsionally coupled pendulum chain [16].

Our numerical study employed a data-driven approach to conduct a statistical analysis of predicting the first instance of spontaneous energy localization in such coupled nonlinear oscillators. First, to gain some physical insight, we develop and test a procedure for predicting energy localization in translationally invariant pendulum chains governed by a sine-Gordon type model, one that relies on a data-smoothing operation using a numerical Box-car integrator. By applying a moving average to the raw data in this way, thus smoothing out rapid spatial fluctuations, we demonstrate that we can achieve earlier and more accurate predictions of the formation of dangerous "hotspots" along these pure chains. This suggests that while the initial instability that forms the energy hotspot is not very sensitive to energies associated with single pendulums, it is sensitive to energetic pendulum clusters.

We follow up with a full-fledged machine-learning approach using a combination of a convolutional neural network (CNN) and a LSTM transformer, running the numerical integrator to generate large sets of training data. We find that prediction accuracy is substantially improved. This would indicate that machine learning picks up on features beyond the mere spatial energy cluster within the chaotic pattern-formation phase. Other studies on fully chaotic dynamical systems proved that neural networks can be trained to predict chaotic patterns substantially beyond the Lyapunov time [17, 18].

Finally, we numerically quantify the effect of small impurities along the chain in "seeding" the dynamic localization process and attracting the energy to its site. Oftentimes, the location and strength of such impurities are not known *a priori* but still have a large influence on the eventual localization site.

We stress that our approach is data-driven in the sense that we attempt to use early-time data traces alone to predict the eventual localization sites. Thus, while we are not taking a theoretical model-based / statistical-mechanics route [19], our predictions do shed light on the type III instability inherent in the pendulum chain.

2 Methods

2.1 The system

We consider the discrete sine-Gordon system given by:

$$\frac{d^2\theta_n}{dt^2} + \omega_0^2 \sin\theta_n + \frac{\kappa}{I}(2\theta_n - \theta_{n+1} - \theta_{n-1}) = 0, \tag{1}$$

where κ represents the torsional spring constant, $I = ml^2$ denotes the moment of inertia of the pendulums of length l and end-mass m, and $\omega_0^2 = g/l$.

Since the sine-Gordon chain is a Hamiltonian system, these governing equations of motion, Eq. (1), can be derived from the following Hamiltonian (together with Hamilton's equations, letting the generalized coordinate q be θ_n):

$$\mathscr{H} = \frac{1}{2}I(\sum_n \dot{\theta}_n^2) + (I\omega_0^2)\sum_n(1-\cos\theta_n) + \frac{1}{2}\kappa\sum_n(\theta_n - \theta_{n-1})^2. \tag{2}$$

The first term represents the total kinetic energy, the second term is the gravitational potential energy, and the third term is the spring potential energy of the pendulum chain.

We now add one additional term to this system, namely a term that makes individual pendulums sensitive to the global average of all pendulums. Such a global - or mean-field - interaction could be practically achieved in a number of ways. One could envision placing the pendulum chain on a movable platform, whose motion would be dictated by the center of mass of all pendulums, which would in turn act back on all pendulums individually. Alternatively, one could envision a single spring running across the entire length of the chain whose orientation follows the mean position of all pendulums. Such an interaction would add one additional term to the Hamiltonian, namely:

$$\mathscr{H}_{global} = \frac{1}{2}h\sum_n(\theta_n - \bar{\theta})^2, \tag{3}$$

where $\bar{\theta}$ indicates the chain average over all of the pendulum angles, and h is the interaction strength (expressed again as a torsional spring constant). Our main reason for including such a term is that it affords us the flexibility of selecting the instability type, as will be explained shortly.

Equation (1) is often written in non-dimensional form by introducing dimensionless time, $\tau = \omega_0 t$. Then, after also including $\mathscr{H}_{global}$, our final system takes the form:

$$\frac{d^2\theta_n}{d\tau^2} = -\sin\theta_n + \tilde{g}(\theta_{n+1} + \theta_{n-1} - 2\theta_n) + \tilde{h}(\bar{\theta} - \theta_n), \tag{4}$$

where $\tilde{g} = \frac{\kappa}{I\omega_0^2}$ and $\tilde{h} = \frac{h}{I\omega_0^2}$. Furthermore, the energy at each site can be defined, in accordance with Eq. (2), as:

$$E_n = \frac{1}{2}(\theta')_n^2 + (1-\cos\theta_n) + \frac{\tilde{g}}{2}(\theta_n - \theta_{n-1})^2 + \frac{\tilde{h}}{2}(\theta_n - \bar{\theta})^2, \tag{5}$$

where the prime refers to differentiation with respect to τ.

The last term in Eq. (4) represents the global term in the Hamiltonian. The sign of $\tilde{h}$ will determine the type of instability we can expect. Its main effect is to shift the uniform mode of oscillation (k=0) relative to the rest of the dispersion curve in frequency. In our numerical simulations, we keep this parameter ($\tilde{h}$) positive, so as to initiate an instability of type III. This has been shown to lead to "catastrophic" localization - see also Ref. [20]. Such significant localization has been experimentally observed in a driven chain of pendulums, where a single intrinsic localized mode can form through a

modulational instability of a uniformly driven pendulum system [16, 21]. Furthermore, such instability also often features an amplitude threshold, above which it is activated. Incidentally, if $\tilde{h} \leq 0$, the instability (of type I, II) then produces a sinusoidal pattern on top of the uniform oscillation, and this modulation would then give rise to a train of localized features.

In subsequent simulation results, unless otherwise specified, the following values were chosen for the two parameters appearing in Eq. (4): $\tilde{g} = 0.75$, $\tilde{h} = 0.05$. The system was initialized with all pendulums at rest and displaced through an angle of $A = 1.75$ rad, with a superimposed uniform noise distribution of amplitude 0.025.

2.2 Python implementation and data smoothing

We used the Python programming language to simulate Eq. (4) for 100 pendulums with periodic boundary conditions. Like other numerical ODE integrators [22], in Python a second-order ordinary differential equation (ODE) must also first be transformed into a system of first-order ODEs, as can always be done. For a single pendulum, we get a system of two coupled ODEs, but for a system of 100 pendulums, there will be 200 coupled ODEs to solve, 100 of which will represent the angular accelerations of every pendulum, given by Eq. (4). With these calculated accelerations, Python's "odeint()" routine evolves the dynamical variables of the chain (all pendulum angles and velocities) forward in time, step by step. Using the data for the angular position and velocity of the entire pendulum system, we can then calculate the energy of each individual pendulum using Eq. (5).

Computer implementation of the torsionally coupled pendulum chain allows us to generate thousands of individual runs with different initial noise profiles. We can thus perform a statistical analysis to test predictors of where energy will localize in the chain. We also examine the average time it takes for a "hotspot" to emerge as a function of uniform mode amplitude and noise level. Here we compute a moving average (sometimes called box-smoothing) of the original energy graph and locate the node with the highest moving average energy. Mathematically, this operation on a raw dataset, $f[i]$, can be described as,

$$\bar{f}[i] = \frac{1}{2k+1} \sum_{-k}^{+k} f[i+k]. \tag{6}$$

The next step is to locate the actual (as opposed to predicted) first instance of enhanced localization of energy. We somewhat arbitrarily set this threshold corresponding to a single pendulum reaching 10% of the total system energy, or 10 times the equipartition value in this chain ($N = 100$). Because total system energy is conserved, we can locate this first instance by scanning through the array of energies at each timestamp until we find an energy value that is above our threshold. The node at which this value occurred is the actual instance of energy localization and the timestamp at which this value occurred is the localization time.

We can now calculate the difference between the predicted node and the actual node. We can do this by taking the absolute value of their difference. However, in a ring of $N = 100$ the actual distance cannot exceed 50. Differences, d, initially computed to be larger than 50 actually map to a true distance of $N - d$.

In a first physics-inspired attempt, we test the hypothesis that pendulum clusters of above-average energy "seed" the spontaneous localization process. Thus, one important parameter here will be the *kernel size*, i.e., the width of the moving-average window.

Finally, we investigate the role small impurities have on the energy localization site. Such on-site impurities can often occur when working with a physical realization of a coupled pendulum chain or any other macroscopic lattice. We want to determine how the strength of even a single impurity may affect both the location of energy localization and the localization time.

2.3 Machine learning methods

Dataset generation

Moving beyond our physics-inspired hypothesis for predicting hot-spots in energy, we now turn to machine learning to ascertain how well a data-centric approach can perform in predicting the same. Here we generated a dataset of 100,000 simulations of the sine-Gordon model, in the same manner as described earlier. As before, in Eq. (4), the parameters $\tilde{g} = 0.75$ for the coupling strength, and $\tilde{h} = 0.5$ for the global interaction strength were used. Each simulation was initialized with a uniform angular displacement $\theta_i(0) = A + \eta_i$, where $A \sim \mathscr{U}(1.65, 1.85)$ is a random amplitude, $\eta_i \sim \mathscr{U}(-0.5, 0.5) \times 0.05$ is additive noise, and initial velocities $\dot{\theta}_i(0) = 0$. As explained before, the equations of motion were solved using the 'odeint' solver from SciPy over a time interval $[0, 150]$. To reduce storage requirements, only the first 10 timesteps were retained, capturing the early dynamics. Using $\Delta t = 0.5$, we made use of data from $t \in [0, 4.5]$.

Again, energy localization was identified by monitoring the per-pendulum energy, according to Eq. (5). A pendulum (i) was labeled as localized at time t if $E_i(t) > 0.1 \cdot E_{\text{total}}(0)$, where $E_{\text{total}}(0) = \sum_{i=1}^{N} E_i(0)$. The first such instance defined the 'localized_pendulum' (index i) and 'localization_time' (timestep t). Simulations were parallelized using Python's 'ThreadPoolExecutor' with two threads, and data were saved in an HDF5 file using the 'h5py' library , resulting in a compact dataset of approximately 3 GB.

Data preprocessing

For machine learning, we extracted the energy profile $E_i(t)$ for $t = 0$ to 4.5 (shape: 10×100) and the initial angular displacement $\theta_i(0)$ (shape: 100), repeating $\theta_i(0)$ across all 10 timesteps to align dimensions. These were stacked to form a feature tensor of shape $(N_{\text{sim}}, 10, 100, 2)$, where $N_{\text{sim}} = 100,000$ is the number of simulations. The features were normalized using scikit-learn's 'StandardScaler' , standardizing the flattened data (shape: $(N_{\text{sim}} \cdot 10 \cdot 100, 2)$) to zero mean and unit variance before reshaping. Targets were defined as: (a) *Localization*: Binary label (1 if 'localized_pendulum' $\neq -1$, 0 otherwise); and (b) *Location*: Pendulum index (0 to 99 or -1 if no localization).

The dataset was split into training (60%), validation (20%), and test (20%) sets using stratified sampling ('train_test_split' from scikit-learn [23]) to maintain the distribution of localization labels.

Machine learning model

We developed a convolutional neural network (CNN) combined with a bidirectional long short-term memory (LSTM) network to predict localization and location, implemented in TensorFlow. The model architecture consists of:

1. An input layer accepting tensors of shape $(10, 100, 2)$.

2. Four time-distributed 1D convolutional layers ('Conv1D') with 32, 64, 128, and 256 filters, respectively, each with ReLU activation and same padding.

3. A time-distributed 'Flatten' layer to reshape the output. 4. A bidirectional LSTM layer with 128 units, 20% dropout, and 20% recurrent dropout to capture temporal dependencies.

5. A dense layer with 64 units and ReLU activation.

6. Two output layers: a sigmoid-activated dense layer for binary localization (1 unit) and a softmax-activated dense layer for location (100 units).

This neural network was trained with the model compiled using the Adam optimizer and a composite loss function:

Localization Loss: Binary cross-entropy, weighted by 0.5.

Location Loss: Custom masked sparse categorical cross-entropy, ignoring -1 labels, weighted by 1.0.

Performance was evaluated using:

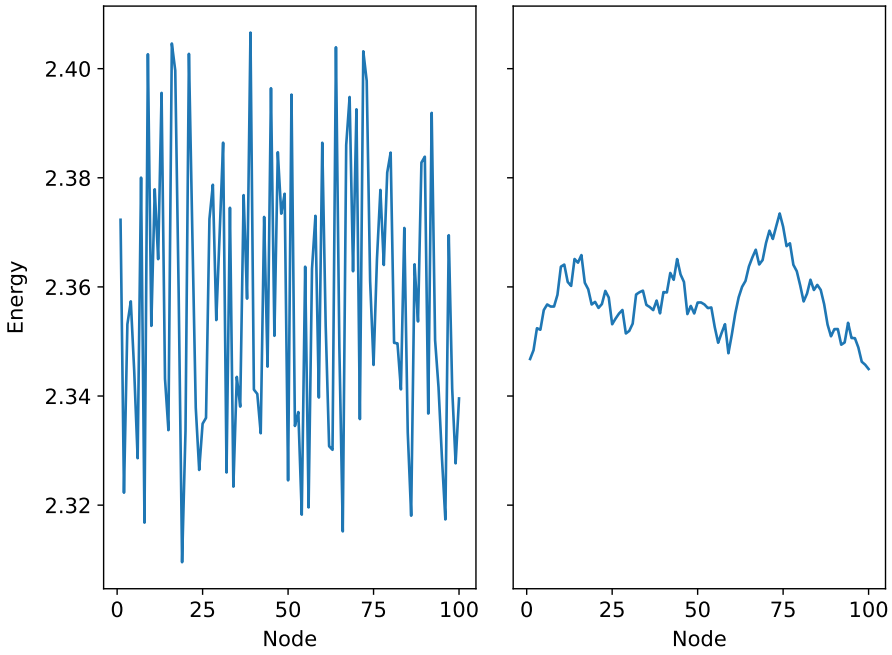


Fig. 1 (a) The raw data trace of the system's energy alongside (b) the smoothed version.

Localization Accuracy: Standard binary classification accuracy.

Location Accuracy (with tolerance): Custom metric counting predictions within $\pm n$ pendulums of the true location, excluding non-localized cases. At first we used $n = 4$, but, as our results improved, we started using $n = 1, 2$ as well.

Models were trained for up to 100 epochs with a batch size of 64, using early stopping (patience=10, monitoring validation loss) to prevent overfitting. The best model weights were restored based on validation loss. For more model architecture details, see the Appendix.

3 Results

3.1 Visualization

Before we systematically analyze the prediction accuracy of our methods, both physics-inspired and machine learning, let us illustrate the workings of the code using one particular run. Fig.1(a) depicts the initial energies of each of the 100 pendulums. Here the pendulum that contains the largest initial energy is located at the 40th node, but there is no clear maximum of the graph. However, the picture becomes significantly clearer just by smoothing the graph with a moving average window size of 15 - see Fig.1(b). The smoothed version of the raw energy graph offers a clearer maximum and shows that the largest concentration of energy occurs around the 70th node.

To determine which of these predictions (at $t = 0$) is better, we can examine the full-time evolution of the pendulum chain using these initial conditions, shown in Fig.2. Here the horizontal axis is spatial (i.e., the 100 pendulums), the vertical axis is computational time, and the color depicts the angle in panel (a), and the energy in panel (b). We see that eventually the energy sharply localizes on pendulums 70-72, and so it is clear that smoothing the energy resulted in a much better prediction.

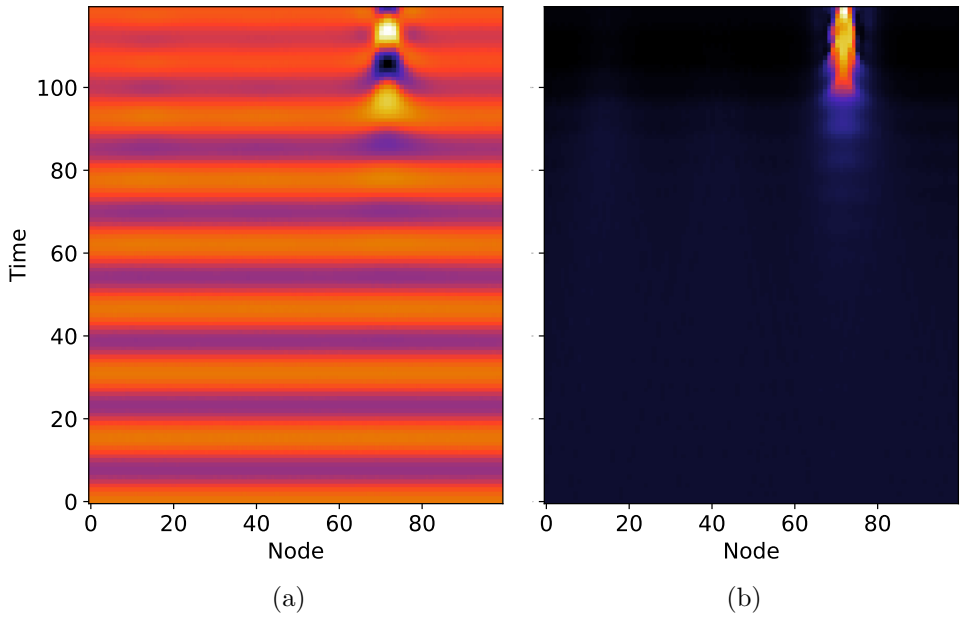


Fig. 2 Color-map of (a) angle and (b) energy oscillation for the entire simulation

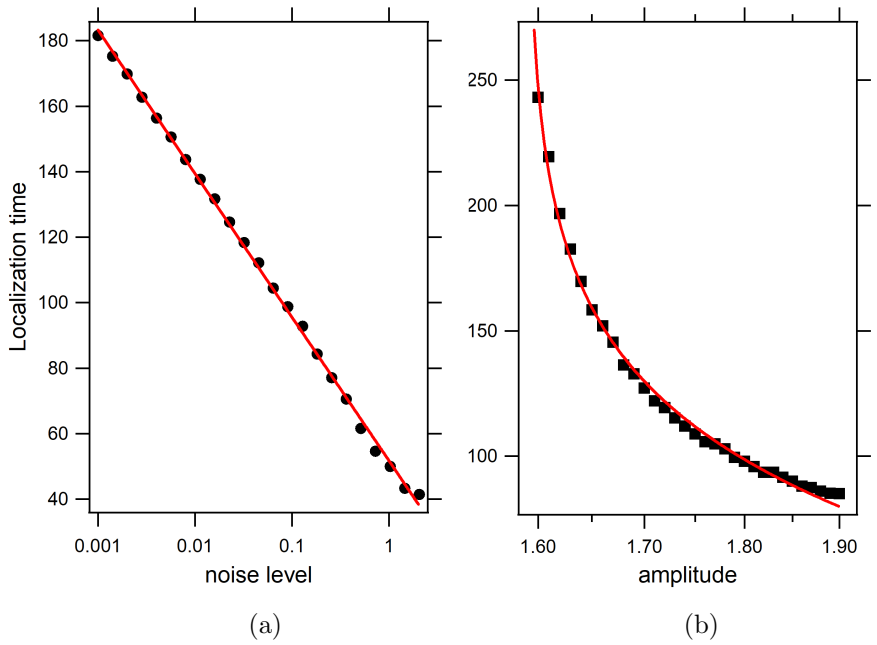


Fig. 3 Dependence of the average localized time on initial noise level and amplitude of the oscillation in the chain. (a) The dependence on noise level is seen as logarithmic. (b) The amplitude dependence features a vertical asymptote at around A=1.59 indicating the instability's amplitude threshold.

3.2 Localization time

By visual inspection of Fig.2, we see that sharp energy localization happens fairly abruptly around a timestamp of $t \approx 100$. We now briefly address the question of what determines this timescale. There are two adjustable parameters in the way the system is initialized, namely the noise level and the initial angular displacement (or amplitude).

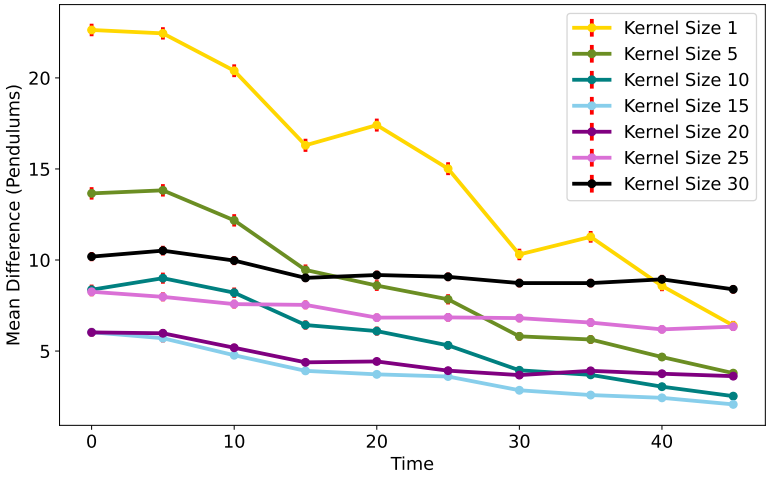


Fig. 4 Mean Difference vs Time for kernel sizes ranging from 1 to 30. Each series shows the absolute difference between a certain kernel size's prediction and the actual localization site, and how this difference changes as the time at which the prediction is made progresses.

To understand how changing these two parameters affects the localization time, we calculated the average localization time over 100 simulations for a series of different noise levels and amplitudes.

Let us first look at the role of noise. Fig.3(a) depicts the time it takes the system to first concentrate 10 percent of its total energy on a single pendulum, as a function of noise level (plotted on a logarithmic axis). There is an exponential relationship between these two quantities - the larger the noise the quicker the localization happens. Since the pattern-formation process is noise-activated, this relationship is expected; nonetheless, the precise exponential relationship (given by the straight line in the logarithmic plot) uncovered numerically is remarkable.

In Fig.3(b), the relationship between localization time and initial pendulum amplitude is illustrated. As expected, here we get a vertical asymptote at around $A = 1.59$ due to the amplitude threshold exhibited by this type of instability. (The solid line is a negative logarithmic function to guide the eye.)

3.3 Prediction quality using only data smoothing

We are now ready to explore the prediction quality of energy localization based on a hypothesis that the type-III pattern-formation process is guided by random clusters of higher energy. Here, our major question pertains to the optimal kernel size in the smoothing algorithm. Fig.4 summarizes our numerical findings. It plots the mean absolute difference between the predicted and actual location against the time of prediction (or "timestamp") for each of the 6 different kernel sizes. Note that the error bars are included in the graph but so small that they are not readily visible.

A few trends can immediately be observed from this graph. The clearest trend is that as time progresses the accuracy of the predictions increases for most of the kernel sizes. This makes sense because the closer the prediction occurs to the real event, the more accurate one would expect it to be. Note that the localization event occurred around a timestamp of 100.

Secondly, the kernel size with the worst accuracy at almost every timestamp is 1. A kernel size of 1 is equivalent to the raw (non-smoothed) energy data, which clearly demonstrates that using a moving average substantially increases the accuracy of predictions. Kernel sizes of about 12-18 seem to be the most accurate over a broad range of times.

To delve deeper into the effect of different kernel sizes, we examined the relationship between kernel size and mean difference at the timestamps of 5. This is effectively the same as examining the columns of data points corresponding to 5 on the horizontal axis in Fig.4.

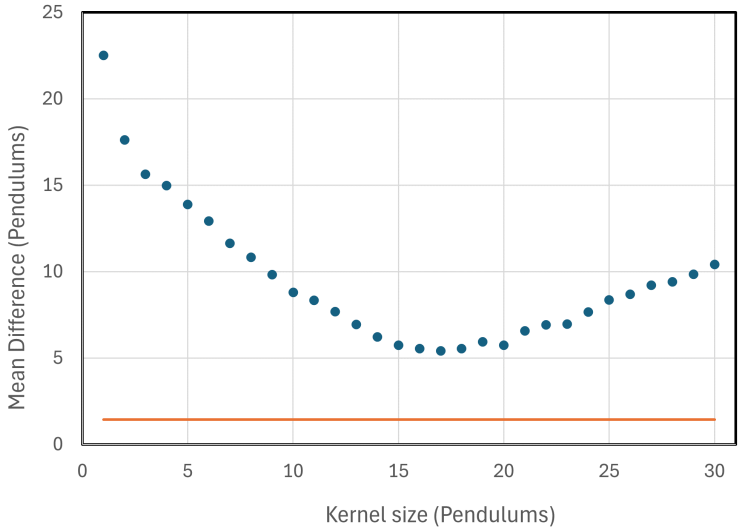


Fig. 5 Mean Difference vs Kernel Size at a timestamp of $t = 5$. A lower mean difference entails a higher prediction quality and vice versa. We circles represent the mean difference for each kernel size (smoothing window size). We see that the best prediction occurs for a kernel size in the range of 15-20. The solid line below the circles represents the mean difference achieved by ML, as discussed later. The x-axis represents different kernel sizes experimentation for box-cart smoothing only, which is irrelevant to the solid line.

Fig.5 plots the mean absolute difference as a function of kernel size (smoothing window size) at a prediction timestamp of $t = 5$, with each data point representing the average of 2000 simulations. Notice the sharp improvement in prediction as the kernel size is increased from 1 to about 10. A minimum is then reached just after 15, past which the prediction gets worse again. Thus, smaller kernel sizes are inaccurate due to the insufficient smoothing of the spatial noise, and the larger kernel sizes are unable to reach pinpoint accuracy due to an over-smoothing of important spatial features in the energy landscape. The optimal range of kernel sizes (to maximize the prediction accuracy) is from 15 to 20, with the most efficient kernel size occurring at 17.

We conclude that the noise-sensitive process whereby the initial instability of type III gives rise to an energy hotspot is sensitive to energetic clusters comprising 10-15 pendulums. In other words, an initial region of elevated energy over that length scale will likely produce an energy hotspot in that same region later in time. Interestingly, a single pendulum of large energy is not nearly as determinative of the eventual localization region.

3.4 Machine-learning results

To evaluate the efficacy of our machine-learning approach, we assessed the CNN-LSTM model's performance on a test set of 20,000 simulations, focusing on its ability to predict both the occurrence of energy localization (localization accuracy) and the specific pendulum where it occurs (location accuracy). The model was trained on early-time data (first 10 timesteps, $0 \leq t \leq 4.5$), using initial angular displacements and per-pendulum energy profiles as features. As before, localization was defined as a pendulum's energy exceeding 10% of the total system energy, with the first such instance determining the localization site and time.

For location accuracy, we adopted a tolerance of ± 1 pendulum, reflecting the spatial extent of energy localization, which typically spans multiple pendulums. The model predicted the localization site within ± 1 pendulum of the true location in 94.38% of cases where localization occurred, with an average absolute error of 1.44 pendulums across 20,000 test simulations. Notably, 94.38% of predictions had an error of ≤ 1 pendulum, with rare outliers skewing the average error slightly higher - see Fig.6.

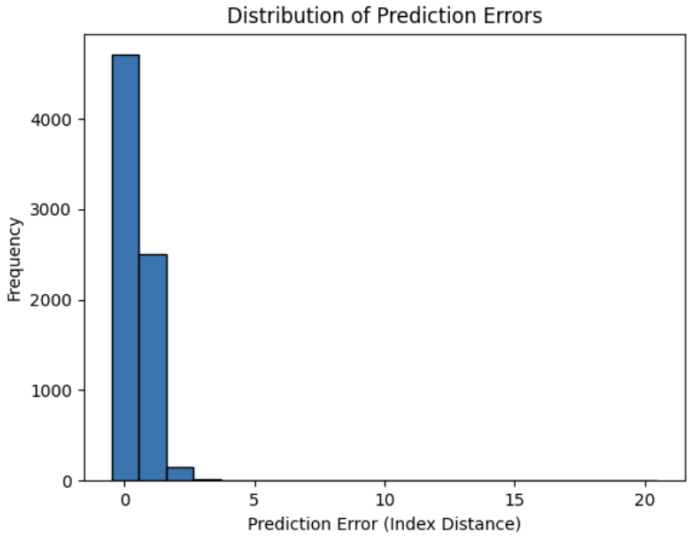


Fig. 6 Distribution of Prediction Errors

To contextualize these results, we compare them to the physics-inspired smoothing approach (Section III.C). At a timestamp of $t = 5$, the optimal smoothing kernel size of 17 yielded a mean absolute difference of approximately 5.4 pendulums (Fig.5).

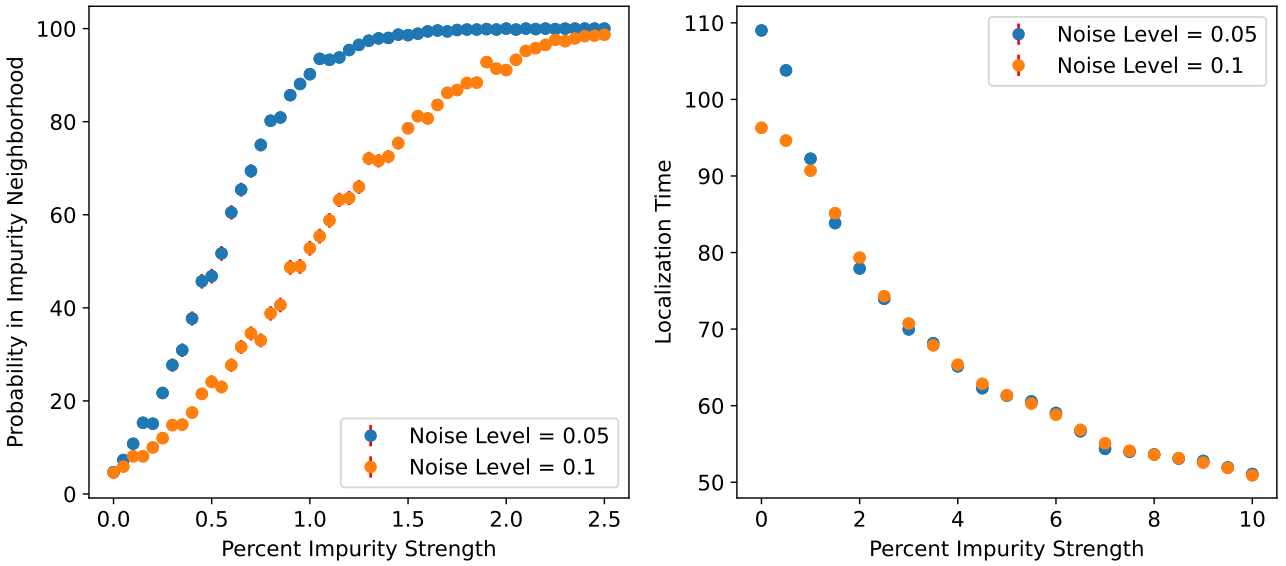
In contrast, our machine-learning model, using only $t \leq 4.5$ data, achieved a mean error of 1.44 pendula, representing a 3.75-times improvement over the best smoothing results at early times. This improvement over the physics-inspired method is illustrated graphically in Fig.5 by the solid (orange) line, which lies significantly below markers from the smoothing-procedure prediction.

These results underscore the machine-learning model's ability to capture complex spatial and temporal patterns in the sine-Gordon system's early chaotic dynamics (see also Ref. [17, 18]), beyond the energy clusters identified by smoothing. The CNN layers likely extract localized spatial features, while the bidirectional LSTM captures temporal evolution, enabling robust predictions despite the system's chaotic sensitivity to initial conditions. The high accuracy at early times ($t \leq 4.5$, well before the typical localization time of $t \approx 100$) suggests that the model identifies subtle precursors to type-III instabilities, potentially including higher-order correlations or nonlinear interactions not captured by smoothing alone.

The dramatic improvement over smoothing and traditional methods highlights the potential of data-driven approaches for predicting chaotic dynamics. The superior performance of machine learning shows that our neural network has discovered some hidden temporal and spatial dependencies that traditional physics methods did not uncover. By leveraging only early-time data, our model offers a practical tool for forecasting energy localization, with applications in mitigating structural instabilities in engineering systems. Future analyses will aim to interpret the learned features, potentially revealing new physical insights into type-III instabilities and their precursors.

3.5 The role of impurities

An unavoidable issue in any macroscopic physical lattice is spatial inhomogeneity. One form of inhomogeneity is the appearance of an impurity somewhere in the lattice, where one lattice element is slightly different from the others. We now examine the role of such an impurity in the pattern formation process. In particular, we study how changing the impurity strength affects where energy localizes and how long it takes to form in time. In our simulations, we set pendulum number 50 as the only pendulum



(a) (b)

Fig. 7 (a) As the impurity strength is increased, the probability of energy localizing near the impurity site also increases and (b) the localization time decreases.

with an impurity in order to showcase the effect of even one impurity in the chain. The impurity was introduced in the on-site term of the sine-Gordon model. Instead of the second term of Eq. (1) starting with ω_0^2, for $n = 50$ it starts with $(1-\epsilon)\omega_0^2$, where $0 < \epsilon \ll 1$. In other words, the pendulum at the impurity is slightly lengthened relative to the others. Impurities of the opposite sign, repel energy in this model.

When determining how impurities affect the location of energy we first notice that impurities of this kind (for $\epsilon > 0$) significantly increase the probability that energy localizes around pendulum 50. In Fig.7(a), we graph the impurity strength against the probability that energy would localize in proximity to pendulum 50. For this data, the acceptable proximity was from pendulum 48 to pendulum 52, a range of 5. Fig.7(a) clearly shows a positive correlation between the percent impurity strength and energy localization in the impurity neighborhood. The two datasets (blue and orange) are for different values of the spatial noise in the initialization. Blue (orange) dots represent an initial noise level (relative to the oscillation amplitude) of 1.4 percent (2.8 percent).

For the lower spatial noise (blue markers), we see that from 0 percent to around 1 percent the relation is relatively linear, and from 1 percent onward, the probability of localization at the impurity site slowly approaches 100 percent. For larger spatial noise in the initialization, it takes a larger impurity strength to achieve the same probability. This makes intuitive sense, given that the spatial noise and the spatial impurity are competing effects in the pattern-formation process that leads to energy localization.

The impurity strength also has a significant impact on localization time. We can see that in Fig.7(b), as the impurity strength increases the localization time decreases. However, the rate at which the localization time decreases also lessens with greater impurity strength, leading the localization time to asymptotically approach a value of around 50. Here we see that the noise level plays a less prominent role.

We thus find that oscillators with an impurity strength as low as 1 percent of an ideal oscillator, or $\epsilon = 0.01$, can significantly alter the result of a simulation by skewing the location and time of energy localization.

4 Conclusions

In this study, we investigated the noise-induced, type III instability that leads to a sharp localization of energy in a sine-Gordon model. Characteristic of a choatic process, this energy-localization dynamics is highly sensitive to slight variations in initial conditions. Because of this, an important question emerged: can the eventual outcome of energy localization be predicted at early times, and what should be known/computed to improve such an early prediction? We started with the physics-informed hypothesis of energy-clusters, and found that smoothing the relevant variables such as the angles or energies (using a numerical box-car integrator) significantly improved the accuracy of the prediction. An optimal kernel size in that smoothing was around 15, suggesting that the instability process is most sensitive to spatial energy clusters of that size.

We then turned to machine learning to ascertain whether further improvements in prediction accuracy were possible based on only early-time data. We found that training a CNN based on energy data at early times (from $0 \leq t \leq 5$), together with knowledge of the initial conditions for angles (at $t = 0$) yielded a dramatic improvement in prediction quality, relative even to the best-kernel smoothing procedure. This strongly suggest that there are additional features, beyond above-average energy clusters, that influence the instability dynamics. A future study will aim at teasing out relevant features uncovered via machine learning.

Finally, we found that even slight impurities in the chain significantly affect energy localization, in time and space. Spatial impurity and spatial noise are thus two competing forces in the pattern-forming process that lead to energy localization.

Acknowledgements

We would like to acknowledge helpful discussions with Hana Zwick, Dr. Robert Malkin, and Prof. Lulu Wang. We also thank Anya English for help with an early version of the code.

Appendix

Model architecture details

The machine learning model integrates convolutional and recurrent neural network components to predict energy localization in the sine-Gordon system. Here, we elucidate the function of each layer, the role of the `TimeDistributed` wrapper, the bidirectional nature of the LSTM, and the choice of activation functions.

The input layer accepts a 4D tensor of shape $(N_{\text{sim}}, 10, 100, 2)$, where N_{sim} is the number of simulations, 10 is the number of timesteps, 100 is the number of pendulums, and 2 represents the energy $E_i(t)$ and initial angular displacement $\theta_i(0)$ (repeated across timesteps). This tensor encapsulates the spatiotemporal evolution of the system, with energy profiles capturing dynamic localization patterns and initial angles providing static context.

Four time-distributed 1D convolutional layers (`Conv1D`), with 32, 64, 128, and 256 filters, respectively, process the input. Each `Conv1D` layer applies a sliding window (kernel size 3) to extract spatial features, such as local energy peaks or angular correlations, across the 100 pendulums at each timestep. The `TimeDistributed` wrapper ensures that the convolutional operation is applied independently to each of the 10 timesteps, preserving temporal structure and producing a sequence of feature maps (e.g., shape $(N_{\text{sim}}, 10, 100, 32)$ for the first layer). The "same" padding maintains the spatial dimension (100 pendulums), enabling consistent feature extraction across layers. The rectified linear unit (ReLU)

activation function, $f(x) = \max(0,x)$, is applied element-wise to introduce nonlinearity, enhancing the model's ability to learn complex patterns while mitigating vanishing gradient issues during training [24].

A time-distributed `Flatten` layer follows, reshaping each timestep's feature map (e.g., $(100,256)$ from the final `Conv1D`) into a 1D vector (e.g., 25,600 elements), yielding a tensor of shape $(N_{\text{sim}}, 10, 25600)$. This preserves the temporal sequence for recurrent processing while consolidating spatial features.

The bidirectional long short-term memory (LSTM) layer, with 128 units, processes the sequence of flattened feature vectors to capture temporal dependencies in the energy and angle dynamics. An LSTM unit maintains a cell state and gates (input, forget, output) to selectively remember or forget information across timesteps, making it suitable for modeling the sequential evolution of localization [25]. Bidirectionality involves two LSTM layers: one processing the sequence forward (timesteps 0 to 9) and another backward (9 to 0), concatenating their outputs. This enables the model to learn both past and future contexts, crucial for identifying localization events that may depend on early and late timestep patterns. Dropout (20%) and recurrent dropout (20%) are applied to prevent overfitting by randomly masking inputs and recurrent connections during training.

A dense layer with 64 units and ReLU activation aggregates the LSTM output, reducing dimensionality and learning higher-level features for classification. Two output layers follow: - The localization output, a dense layer with 1 unit and sigmoid activation, $f(x) = 1/(1+e^{-x})$, produces a probability (0 to 1) for binary classification (localization vs. no localization). Sigmoid is chosen for its bounded output, suitable for binary tasks. - The location output, a dense layer with 100 units and softmax activation, $f(x)_i = e^{x_i}/\sum_{j=1}^{100} e^{x_j}$, yields probabilities over the 100 pendulums, representing the likelihood of localization at each site. Softmax ensures normalized probabilities, ideal for multi-class classification.

The ReLU activation in convolutional and dense layers facilitates efficient gradient propagation and sparsity, while sigmoid and softmax activations align with the binary and multi-class nature of the outputs, respectively. The model's architecture, combining spatial feature extraction (CNN), temporal modeling (bidirectional LSTM), and robust classification (dense layers), effectively captures the complex dynamics of energy localization in the sine-Gordon system.

Model performance was assessed on the test set using the total loss (weighted sum of localization and location losses), localization loss, location loss, localization accuracy , and location accuracy.

References

[1] Campbell, D.K., Flach, S., and Kivshar, Y.S. (2004), Localizing energy through nonlinearity and discreteness, *Physics Today*, **57**, 43-49.

[2] Sato, M., Hubbard, B., and Sievers, A. (2006), Colloquium: Nonlinear energy localization and its manipulation in micromechanical oscillator arrays, *Reviews of Modern Physics*, **78**, 137-157.

[3] Green, D. and Unruh, W.G. (2006), The failure of the tacoma bridge: A physical model, *American Journal of Physics*, **74**, 706-716.

[4] Arioli, G. and Gazzola, F. (2015), A new mathematical explanation of what triggered the catastrophic torsional mode of the tacoma narrows bridge, *Applied Mathematical Modelling*, **39**, 901-912.

[5] Zhang, M., Zhao, G., Wang, L., and Li, J. (2017), Wind-induced coupling vibration effects of high-voltage transmission tower-line systems, *Shock and Vibration*, **2017**.

[6] Sharon, E., Cohen, G., and Fineberg, J. (2001), Propagating solitary waves along a rapidly moving crack front, *Nature*, **410**, 68-71.

[7] Zhou, S., Grønbech-Jensen, N., Bishop, A., Lomdahl, P., and Holian, B. (1997), A nonlinear-discrete model of dynamic fracture instability, *Physics Letters A*, **232**, 183-188.

[8] Kivshar, Y.S. (1993), Creation of nonlinear localized modes in discrete lattices, *Physical Review E*, **48**, 4132.

[9] He, Y., Witt, A., Trillo, S., Chabchoub, A., and Hoffmann, N. (2022), Extreme wave excitation from localized phase-shift perturbations, *Physical Review E*, **106**, L043101.

[10] Cross, M.C. and Hohenberg, P.C. (1993), Pattern formation outside of equilibrium, *Reviews of Modern Physics*, **65**, 851.

[11] Cuevas-Maraver, J., Kevrekidis, P.G., and Williams, F. (2014), The sine-gordon model and its applications, *Nonlinear Systems and Complexity*, **10**.

[12] Helal, M. (2002), Soliton solution of some nonlinear partial differential equations and its applications in fluid mechanics, *Chaos, Solitons & Fractals*, **13**, 1917-1929.
[13] Kivshar, Y.S. and Malomed, B.A. (1989), Dynamics of solitons in nearly integrable systems, *Reviews of Modern Physics*, **61**, 763.
[14] Ferré, M.A., Clerc, M.G., Coulibally, S., Rojas, R.G., and Tlidi, M. (2017), Localized structures and spatiotemporal chaos: comparison between the driven damped sine-gordon and the lugiato-lefever model, *The European Physical Journal D*, **71**, 1-8.
[15] Lomdahl, P. (1985), Solitons in josephson junctions: An overview, *Journal of Statistical Physics*, **39**, 551-561.
[16] Thakur, R.B., English, L.Q., and Sievers, A. (2007), Driven intrinsic localized modes in a coupled pendulum array, *Journal of Physics D: Applied Physics*, **41**, 015503.
[17] Pathak, J., Lu, Z., Hunt, B.R., Girvan, M., and Ott, E. (2017), Using machine learning to replicate chaotic attractors and calculate lyapunov exponents from data, *Chaos: An Interdisciplinary Journal of Nonlinear Science*, **27**, 121102.
[18] Pathak, J., Hunt, B., Girvan, M., Lu, Z., and Ott, E. (2018), Model-free prediction of large spatiotemporally chaotic systems from data: A reservoir computing approach, *Physical Review Letters*, **120**, 024102.
[19] Rumpf, B. (2004), Simple statistical explanation for the localization of energy in nonlinear lattices with two conserved quantities, *Physical Review E*, **69**, 016618.
[20] English, L.Q., Sato, M., and Sievers, A. (2003), Modulational instability of nonlinear spin waves in easy-axis antiferromagnetic chains. II. influence of sample shape on intrinsic localized modes and dynamic spin defects, *Physical Review B*, **67**, 024403.
[21] Cuevas, J., English, L.Q., Kevrekidis, P.G., and Anderson, M. (2009), Discrete breathers in a forced-damped array of coupled pendula: modeling, computation, and experiment, *Physical Review Letters*, **102**, 224101.
[22] Press, W.H., Teukolsky, S.A., Vetterling, W.T., and Flannery, B.P. (1992), *Numerical Recipes in C*, Cambridge University Press, New York, NY.
[23] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., and Vanderplas, J. (2011), Scikit-learn: Machine learning in python, *The Journal of Machine Learning Research*, **12**, 2825-2830.
[24] Goodfellow, I., Bengio, Y., and Courville, A. (2016), *Deep Learning*, MIT Press, http://www.deeplearningbook.org.
[25] Hochreiter, S. and Schmidhuber, J. (1997), Long short-term memory, *Neural Computation*, **9**, 1735-1780.