

# Statistical Phase Transitions and Extreme Events in Shallow Water Waves with an Abrupt Depth Change

Andrew J. Majda <sup>a</sup> and Di Qi <sup>a</sup>

July 14, 2019

<sup>a</sup> Department of Mathematics and Center for Atmosphere and Ocean Science, Courant Institute of Mathematical Sciences, New York University, New York, NY 10012

## Abstract

Predicting extreme events in complex nonlinear systems is an extremely challenging and important area in climate science. Important examples include extreme events near coastal continental shelves for shallow water waves and deep internal waves. Recent laboratory experiments reveal a remarkable transition from near Gaussian to highly skewed anomalous statistics with extreme events by measuring the surface water wave displacements in shallow water across an abrupt depth change (ADC). A statistical dynamical model has been proposed and used to accurately predict the representative statistical transition using Gibbs invariant measures for the truncated KdV equation at low and high inverse temperatures representing flows before and after the ADC. In this paper, we use much lower-dimensional truncated paradigm models to understand the statistical phase transition and the creation of extreme events. Especially, a two-mode interacting model with exact integrable dynamics is adopted to characterize the core transition mechanism as the model parameter varies. The choice of the radically truncated two-mode model is motivated by the self-similar solution structures with reducing numbers of truncated modes. A clear separation of distinct dynamics in the phase space is discovered for the Gibbs ensembles sampled from different inverse temperatures. Direct numerical tests with various model truncation sizes are presented to illustrate the statistical transition in parameter regimes. The analysis here can also provide a theoretical guideline for a wider variety of models concerning the generation of extreme events and anomalous statistics.

## 1 Introduction

Extreme events and their anomalous statistics are universally observed in various complex nonlinear systems such as in the climate, material and neuroscience, as well as engineering design [13, 3, 6, 19, 21]. Understanding and accurate prediction of such phenomena is a grand challenge, and has become an active contemporary topic in applied mathematics. Qualitative and quantitative models [10, 15, 16, 13, 11] have been proposed with novel numerical algorithms to overcome the curse of dimension for extreme event prediction in large complex systems [5, 4, 15, 17]. One typical example with practical importance is the occurrence of Rogue waves with extreme events in different physical settings of shallow and deep water waves [18, 20, 21].

Recently, a statistical model [12] is proposed for predicting extreme events and anomalous statistical features in shallow water waves. This work is motivated by the controlled laboratory experiments in weakly turbulent shallow water waves going through an abrupt depth change (ADC). The water waves exhibit a remarkable transition from nearly Gaussian statistics in incoming wave trains with deeper water depth to outgoing waves trains after the ADC with extreme anomalous statistics and a large positive skewness of the surface height [3]. It is shown under various experimental control parameters that the anomalous statistical behaviors can be explained and quantitatively predicted using a simple statistical model developed in [12]. In particular, the typical statistical phase transition from a near-Gaussian distribution in the incoming state to a highly skewed non-Gaussian distribution in the outgoing state is effectively captured from the statistical formulation with a proper statistical matching condition.

The basic idea in the modeling procedure is illustrated by the diagram in Figure 1.1. The one-dimensional *truncated Korteweg-de Vries* (tKdV) equation is proposed as the governing equation modeling the incoming and outgoing flow states matched at the ADC location. The tKdV equation enjoys many desirable features with trackable dynamics in

a Hamiltonian structure. Then the conserved energy and Hamiltonian induce incoming and outgoing Gibbs invariant measures from the equilibrium statistical mechanics [9]. The statistical matching of the known near-Gaussian incoming Gibbs state at the ADC completely determines the anomalous outgoing Gibbs state to be predicted, with an analytic formula for the anomalous skewness to be described in this paper and verified by direct numerical simulations in [12]. The prediction model successfully captures key features of the experiment. The strategy described here should be useful for predicting extreme anomalous statistical behaviors in other dispersive media in different settings [19, 6]. One important example is the extreme events prediction in internal waves near continental shelves.

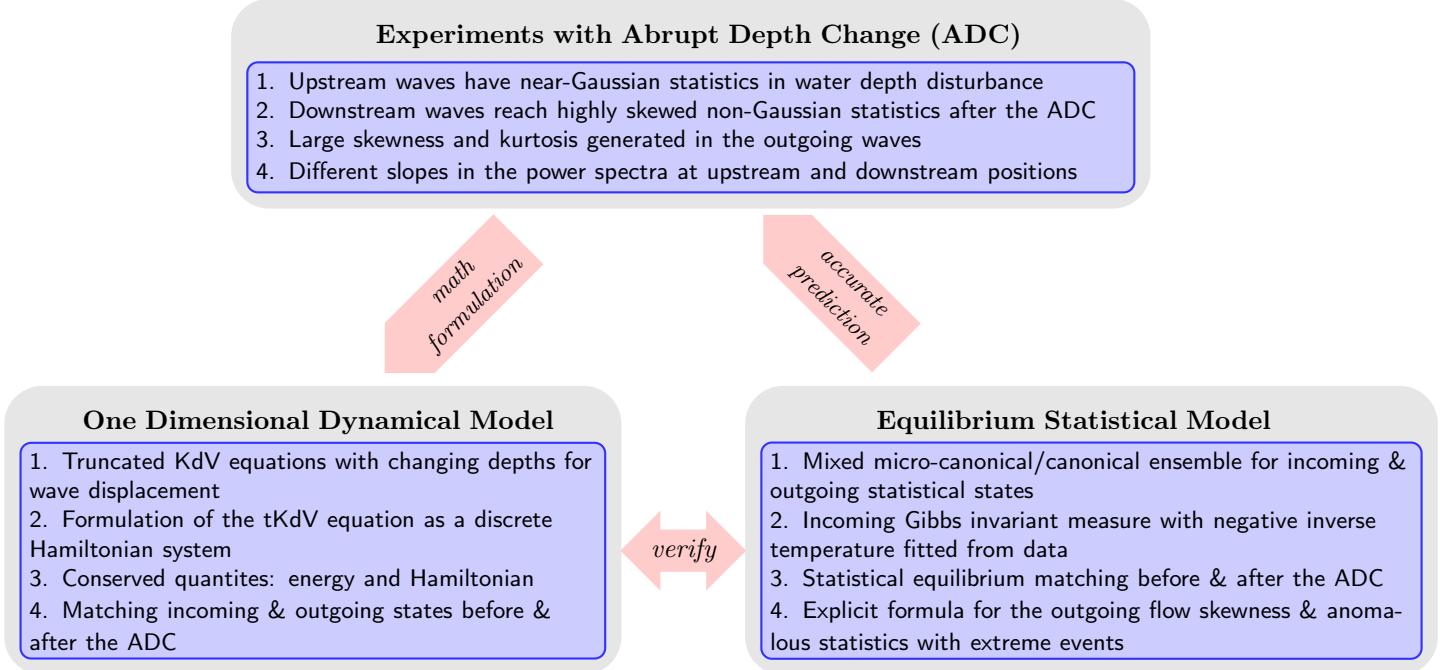


Figure 1.1: Diagram illustrating the strategies for modeling the anomalous statistical transition observed in water wave experiments going through an abrupt depth change. Dynamical model and equilibrium statistical theory are combined to offer theoretical explanation for the experimental observations.

In this paper, we revisit the mathematical formulation in the statistical phase transition and the generation of high skewness connected by both the dynamical and statistical models shown in Figure 1.1. The main goal is to link the complex multiscale interactions in the tKdV equation with low-order truncated models which can provide a simple but effective characterization of the typical phase transition for the Gibbs ensemble predictions before and after the ADC from low to high negative inverse temperature. Starting from the natural moderate truncation size provided in [12], the tKdV model solutions with decreasing numbers of truncated modes are compared down to only two interacting modes. The motivation of adopting a skeleton two-mode truncated model is from the observation of self-similar solution structures displayed by the models with reducing truncation sizes. Particularly, we seek the concise low-dimensional origin of this statistical phase transition mechanism, which plays the central role in the prediction of extreme events in the shallow water wave disturbances. In the two-mode interacting model, integrable dynamics provides an explicit phase portrait of the solution trajectory starting from different values of the Hamiltonian. The ensemble solutions are naturally separated in two parameter regimes based on the amplitude of the inverse temperature in the Gibbs measure. A direct connection between the low and high negative inverse temperature regimes with distinct statistics can be found from the simple trackable model. The structure found in the extreme two-mode model then can be generalized to models with higher wavenumber truncations, and is confirmed with direct numerical simulations.

In the structure of this paper, a review about the dynamical model and statistical equilibrium mechanics used for modeling the surface water waves before and after the ADC is shown in Section 2. The model statistics and mixing property with different truncation sizes are compared in Section 3. Then the two-mode tKdV model is proposed in Section 4 to offer a precise description for the phase transition mechanism. First an integrable dynamics is derived, and

then the two statistical phases get connected with different parameter regimes. The results are summarized in Section 5 as a conclusion.

## 2 Statistical Models for Surface Wave Turbulence and the Matching Conditions

In this section, we review the statistical model and matching condition developed in [12] with a natural moderate model truncation size to capture the experimental observations. The statistical scattering between the near-Gaussian incoming state and the highly skewed non-Gaussian outgoing state of the Gibbs measures sets up the phase transition from the low negative inverse temperature regime to the high negative inverse temperature regime.

### 2.1 Dynamical model by tKdV equation with water depth dependence

The surface wave turbulence can be modeled by a one-dimensional deterministic dynamical model, that is, the standard Korteweg-de Vries (KdV) equation [7]

$$u_t + uu_x + u_{xxx} = 0, \quad x \in [-\pi L, \pi L]. \quad (1)$$

Above the state variable  $u(x, t)$  is the leading-order surface wave disturbance. The KdV equation (1) can be also formulated as a *Hamiltonian system* as

$$\dot{u} = \mathcal{J} \frac{\delta \mathcal{H}}{\delta u}, \quad \mathcal{J} = -\partial_x, \quad \mathcal{H}(u) = \int_{-\pi L}^{\pi L} \left( \frac{1}{6} u^3 - \frac{1}{2} u_x^2 \right) dx. \quad (2)$$

The evolution of any functional  $\mathcal{F}(u)$  obeys the dynamical equation

$$\mathcal{F}_t = \{\mathcal{F}, \mathcal{H}\} = \int_{-\pi L}^{\pi L} \frac{\delta \mathcal{F}}{\delta u} \mathcal{J} \frac{\delta \mathcal{H}}{\delta u} dx,$$

through the Poisson bracket defined by the symplectic operator  $\mathcal{J}$ . Immediately, we have the conservation of the Hamiltonian in (2),  $\mathcal{H}_t = \{\mathcal{H}, \mathcal{H}\} = 0$ . Besides the Hamiltonian  $\mathcal{H}$ , the KdV equation also conserves the momentum  $\mathcal{M}$  and energy  $\mathcal{E}$  defined as

$$\mathcal{M}(u) = \int_{-\pi L}^{\pi L} u dx, \quad \mathcal{E}(u) = \frac{1}{2} \int_{-\pi L}^{\pi L} u^2 dx.$$

#### Truncated equation with normalized momentum and energy

In modeling water waves using the KdV equation, it is convenient to use a normalized version of the equation. The state variable  $u$  is normalized with zero mean and unit energy with the change of variables

$$t = \tilde{t}, \quad x = L\tilde{x} + M\tilde{t}, \quad u = E^{1/2}L^{-1/2}\tilde{u} + M,$$

where  $M$  is the conserved total momentum and  $E$  is the conserved total energy from the original system (1), and  $L$  defines the characteristic length scale of the system. In this way, the total momentum is normalized to zero  $\mathcal{M}(u) = 0$  without loss of generality due to the Galilean invariance and the total energy is rescaled to unity,  $\mathcal{E}(u) = 1$ , conserved during the evolution, while  $E$  characterizes the total energy injected in the system. The additional shift in time  $Mt$  in the new coordinate creates the Doppler shift from the non-zero mean momentum  $M$ . From now on in this paper, we use the normalized state variables and neglect the ‘tildes’ for simplicity in representation.

Finally, the KdV equation is further truncated in the first  $\Lambda$  spectral modes to generate turbulent dynamics. A Galerkin projection  $\mathcal{P}_\Lambda$  is applied to the original equation (1) with a high wavenumber truncation up to  $\Lambda$

$$u_\Lambda(x, t) \equiv \mathcal{P}_\Lambda u = \sum_{|k| \leq \Lambda} \hat{u}_k(t) e^{ikx}, \quad (3)$$

with in total  $J = 2\Lambda + 1$  grid points. The wavenumber truncation  $\Lambda$  is fixed in a moderate value for generating weakly turbulent dynamics. The Galerkin truncated state variable  $u_\Lambda$  is normalized with zero mean  $\hat{u}_0 = 0$  and unit energy

$2\pi \sum_{k=1}^{\Lambda} |\hat{u}_k|^2 = 1$ . Therefore, the water wave motion is described by the truncated KdV equation (tKdV) by projecting the continuous equation (1) to the truncated subspace with water depth  $D$  dependence [7, 12]

$$\frac{\partial u_{\Lambda}}{\partial t} + \frac{D^{-3/2}}{2} E^{1/2} L^{-3/2} \frac{\partial}{\partial x} \mathcal{P}_{\Lambda} (u_{\Lambda})^2 + D^{1/2} L^{-3} \frac{\partial^3 u_{\Lambda}}{\partial x^3} = 0, \quad x \in [-\pi, \pi]. \quad (4)$$

The model (4) is non-dimensionalized in the periodic domain  $[-\pi, \pi]$  with the three model parameters  $(E, L, D)$ . The additional projection in front of the quadratic term  $u_{\Lambda}^2$  is used to remove the aliasing modes that go beyond the range  $|k| > \Lambda$ , which are conserved quantities. The conserved Hamiltonian is discretized accordingly in the finite dimensional subspace decomposed into the difference of two components containing cubic and quadratic terms

$$\mathcal{H}_{\Lambda} = D^{-3/2} E^{1/2} L^{-3/2} H_3(u_{\Lambda}) - D^{1/2} L^{-3} H_2(u_{\Lambda}), \quad H_3(u) = \frac{1}{6} \int_{-\pi}^{\pi} u^3 dx, \quad H_2(u) = \frac{1}{2} \int_{-\pi}^{\pi} u_x^2 dx. \quad (5)$$

Above, the cubic term  $H_3$  describes the skewness of the state, while the quadratic term  $H_2$  characterizes the *slopes of the surface waves*,  $u_x$ . The amplitudes of the characterizing model parameters  $(E, L, D)$  can be discovered from a scale analysis from the experimental data. See [12] for the reference values of several representative regimes from a detailed scale analysis.

### Deterministic matching condition

In modeling the water waves going through an abrupt depth change, the water depth is normalized to the unit  $D_- = 1$  before the ADC and becomes smaller  $D_+ < 1$  after the ADC. Accordingly, the surface disturbance state is modeled by the state  $u_{\Lambda}^-$  for the incoming wave with water depth  $D_-$  and  $u_{\Lambda}^+$  for the outgoing waves with water depth  $D_+$ . The abrupt depth change from  $D_-$  to  $D_+$  is assumed to take place at  $t = T_{\text{ADC}}$ . A *deterministic matching condition* is given for the surface displacement  $u_{\Lambda}^{\pm}$  agreeing at the locations before and after the abrupt depth change  $T_{\text{ADC}}$

$$u_{\Lambda}^-(x, t) |_{t=T_{\text{ADC}}-} = u_{\Lambda}^+(x, t) |_{t=T_{\text{ADC}}+}, \quad (6)$$

assuming the abrupt depth change is met at  $t = T_{\text{ADC}}$ .

The intuition for the distinct model dynamics comes from the balance between the cubic and quadratic terms in the Hamiltonian  $H_{\Lambda}^{\pm}$ . After the depth change,  $D_+ < 1$ , more weight is added in the cubic term,  $H_3$ , for stronger nonlinearity and weaker dispersion for the third-order derivative term reflected by the smaller coefficient for  $H_2$  in the Hamiltonian. Since  $\frac{\partial u}{\partial x}$  is the slope of the wave height,  $H_2(u)$  measures the wave slope energy.

## 2.2 Equilibrium statistical mechanics for the stationary invariant measure

For a better characterization of the turbulent waves in incoming and outgoing flows, we introduce the statistical description of the tKdV model captured by ensemble simulations. First, the equilibrium probability distribution can be quantified by an invariant statistical measure. The equilibrium invariant measure is dictated by the conservation laws in the tKdV equation. There exist two important conserved functionals, the total energy  $\mathcal{E}_{\Lambda}$  and the Hamiltonian  $\mathcal{H}_{\Lambda}$ , in the tKdV equation (4). The choice is to pick a mixed Gibbs measure with microcanonical ensemble in the quadratic energy  $\mathcal{E}_{\Lambda}$  and canonical ensemble in the Hamiltonian  $\mathcal{H}_{\Lambda}$  [9, 2, 1]. The invariant Gibbs measure is then defined based on canonical Hamiltonian fixed on the isosurface with constant energy (normalized to unit)

$$\mathcal{G}_{\theta}(u_{\Lambda}; E) = C_{\theta} \exp(-\theta \mathcal{H}_{\Lambda}) \delta(\mathcal{E}_{\Lambda} - 1),$$

with  $\theta$  the *inverse temperature*. Summarizing the expressions for the truncated variables, the invariant Gibbs measure for the tKdV model (4) about the normalized state variable  $u_{\Lambda}$  with unit energy can be written explicitly as

$$\mathcal{G}_{\theta^{\pm}}(u_{\Lambda}) = C_{\theta^{\pm}} \exp \left( -\theta^{\pm} \left\{ h_3^{\pm} \int_{-\pi}^{\pi} u_{\Lambda}^3 dx - h_2^{\pm} \int_{-\pi}^{\pi} (\partial_x u_{\Lambda})^2 dx \right\} \right) \delta \left( \frac{1}{2} \int_{-\pi}^{\pi} u_{\Lambda}^2 dx - 1 \right), \quad (7)$$

with the coefficients  $h_3^{\pm} = \frac{1}{6} E^{1/2} L^{-3/2} D_{\pm}^{-3/2}$  and  $h_2^{\pm} = \frac{1}{2} L^{-3} D_{\pm}^{1/2}$  depending on the model parameters. A constant mean state will not alter the final invariant measure with a Doppler shift in the solution (see Section 3.1).

The distinct statistics in the upstream and downstream waves can be controlled by the inverse temperature parameter  $\theta$ . It is found [12, 2] that the negative temperature regime,  $\theta < 0$ , gives the correct energy spectra and PDFs

to predict the experiments. The Gibbs invariant measure (7) transfers from near-Gaussian to highly skewed distribution as the amplitude of the inverse temperature  $\theta$  increases. The expectation of any functional  $F(u)$  can be computed based on the above invariant measures (7) using proper sampling strategies

$$\langle F \rangle_{\mathcal{G}_\theta} \equiv \int F(u) \mathcal{G}_\theta(u) du.$$

Besides, the invariant measure has the advantage to predict an equilibrium energy spectrum without the requirement to run the tKdV equation directly. On the other hand, the time autocorrelation and transient statistics about the state variable  $u_\Lambda$  can only be recovered from the direct model simulations.

### Statistical matching condition in invariant measures before and after the abrupt depth change

In the incoming flow field, the inverse temperature  $\theta^-$  is chosen so that  $\mathcal{G}_\theta^-$  has Gaussian statistics consistent with the incoming flow which is assumed to be accurately measured. The downstream statistics need to be recovered from the matching condition at the ADC point. The idea for the statistical matching condition is to connect the invariant measures in incoming and outgoing flow statistics using different inverse temperatures  $\theta^\pm$  [12]. The Gibbs measures  $\mathcal{G}_\theta^\pm$  before and after the abrupt depth change are defined based on the different inverse temperatures  $\theta^\pm$  on the two sides of the solutions

$$\begin{aligned} \mu_t^-(u_\Lambda^-; D_-), \quad u_\Lambda |_{t=T_{\text{ADC}-}} &= u_0, \quad t < T_{\text{ADC}}; \\ \mu_t^+(u_\Lambda^+; D_+), \quad u_\Lambda |_{t=T_{\text{ADC}+}} &= u_0, \quad t > T_{\text{ADC}}, \end{aligned}$$

where  $u_0$  represents the deterministic matching condition (6) between the incoming and outgoing waves.  $\mu_t^\pm$  are the PDFs for the states  $u_\Lambda^\pm$  before and after the depth change. The invariant measures  $\mu_\infty^\pm(u; D^\pm)$  as  $t \rightarrow \infty$  is used to represent the equilibrium distributions at the steady state. The two distributions,  $\mu_t^-, \mu_t^+$  should also be matched at the depth change location  $T_{\text{ADC}}$ , so that

$$\mu_\infty^-(u_\Lambda) = \mu_{t=T_{\text{ADC}}}^-(u_\Lambda) = \mu_{t=T_{\text{ADC}}}^+(u_\Lambda).$$

In matching the flow statistics before and after the abrupt depth change, first we use the conservation of the deterministic Hamiltonian value  $H_\Lambda^+$  after the depth change. Then assuming ergodicity [1], the statistical expectation for the Hamiltonian  $\langle H_\Lambda^+ \rangle$  is conserved in time before and after the depth change at  $t = T_{\text{ADC}}$  and should stay in the same value as the system approaches equilibrium as  $t \rightarrow \infty$ . The final *statistical matching condition* for the outgoing flow statistics with the inverse temperature  $\theta^+$  can be found by

$$\langle H_\Lambda^+ \rangle_{\mathcal{G}_\theta^+} = \langle H_\Lambda^+ \rangle_{\mathcal{G}_\theta^-}, \quad (8)$$

with the outgoing flow Hamiltonian  $H_\Lambda^+$  and the Gibbs measures  $\mathcal{G}_\theta^\pm$  before and after the abrupt depth change. This statistical relation connects the near-Gaussian regime in the incoming flow statistics with the highly skewed regime in the outgoing flow field. Explicit formulas for the Gibbs measures  $\mathcal{G}_{\theta^\pm}$  before and after the abrupt depth change can be used in (7). Notice that in the statistical matching condition (8), the statistical information for the incoming Hamiltonian  $H_\Lambda^-$  is not required. The Hamiltonian before the depth change is only used to construct the invariant measure,  $\mathcal{G}_{\theta^-}$ .

### Analytic formula for the upstream skewness after the ADC

Furthermore, a statistical link between the upstream and downstream energy spectra is derived [12] for an analytical prediction of the skewness in the flow state  $u$  after the ADC. The skewness of the state variable  $u_j$  at one spatial grid point is defined as the ratio between the third and second moments

$$\kappa_3 = \langle u_j^3 \rangle_\mu / \langle u_j^2 \rangle_\mu^{\frac{3}{2}}.$$

With mild assumptions on the distribution functions of small inflow skewness and homogeneous outflow statistics [12], the skewness of the downstream state variable  $u_\Lambda^+$  after the ADC is given by the from the energy spectra in inflow and outflow statistics and the wave slope energy spectra

$$\kappa_3 = C \sum_{1 \leq |k| \leq \Lambda} k^2 (r_k^+ - r_k^-) = \frac{C}{2\pi} \int_{-\pi}^{\pi} \left[ \langle (u_x)^2 \rangle_{\mu+} - \langle (u_x)^2 \rangle_{\mu-} \right] dx, \quad C = 3\pi^{\frac{3}{2}} L^{-\frac{3}{2}} E^{-\frac{1}{2}} D_+^2, \quad (9)$$

with  $r_k^- = \left\langle |\hat{u}_k|^2 \right\rangle_{\mu_-}$  the variance in upstream statistics and  $r_k^+ = \left\langle |\hat{u}_k|^2 \right\rangle_{\mu_+}$  the statistics from downstream statistics, and the coefficient  $C$  only depending on the model setup. The upstream equilibrium measure is assumed Gaussian with zero skewness,  $\langle H_3 \rangle_{\mu_-} = \frac{1}{6} \int_{-\pi}^{\pi} \langle u^3 \rangle_{\mu_-} dx = 0$  (see the supporting information in [12] for the detailed derivation).

The spectral difference between the incoming and outgoing flows calibrates the statistics in the ‘waves slopes’,  $u_x$ , before and after the depth change. In particular, the downstream skewness is positive if and only if the difference of the two wave slope energy spectra is positive. This means that there is more small scale wave slope energy in the outgoing state. We can see that a positive skewness can be generated from  $r_k^+ > r_k^-$  in small scale modes and  $r_k^+ < r_k^-$  in large scales, with the total energy of the system always normalized to unit,  $\pi \sum r_k^+ = \pi \sum r_k^- = 1$ . It implies that the outflow energy spectrum always has a slower decay rate than the inflow energy spectrum which possesses stronger energy in larger scales and weaker energy in the smaller scales.

### 3 Effective Truncation Sizes in Direct Numerical Simulations

One interesting issue in the tKdV equation (4) is the model performance as the truncation size  $\Lambda$  for the state variable  $u_\Lambda$  changes. The model mixing property and the statistical energy spectra as well as the equilibrium statistical measures vary with changing numbers of truncated modes, while the typical phase transition from low to high inverse temperature is maintained even with extremely low-dimensional truncation. This section investigates the trend of changes in the tKdV model statistics through direct numerical simulations according to flow states starting with different inverse temperatures.

#### 3.1 Spectral formulation of the tKdV equation and its autocorrelation functions

One advantage of adopting a smaller truncation size  $\Lambda$  is to achieve a clearer look at the linear and nonlinear coupling between different scale modes while still keep similar statistical invariant features. In the tKdV equation (4), the continuous state variable in fluctuation  $u(x, t)$  is projected to the spectral space  $u_\Lambda$  with a wavenumber truncation  $\Lambda$ . The projected modes (3) can be further separated with the linear wave frequency  $\omega_k$  of each wavenumber

$$u_\Lambda(x, t) = U + \sum_{1 \leq |k| \leq \Lambda} a_k(t) e^{i(kx - \omega_k t)}, \quad (10)$$

The frequency  $\omega_k$  will be defined next from the dispersion relation of the system. A constant background mean state,  $U$ , is also added to illustrate its effect on the wave speed. The energy spectrum of the system is not altered by this new decomposition, that is,  $E_k = \mathbb{E}|\hat{u}_k|^2 = \mathbb{E}|a_k|^2$ . The dispersion relation can be found from the linearized component of the tKdV equation (4)

$$\omega_k = Uk - L^{-3}D^{1/2}k^3 = kc_k, \quad (11)$$

with  $c_k = \omega_k/k$  the wave velocity for each mode and  $\omega_{-k} = -\omega_k$ . The mean state  $U$  creates right-moving linear waves on top of the nonlinear dispersion from the third-order derivative. Without the mean state  $U \equiv 0$ , all the waves are left-going with  $c_k < 0$ ; while with a non-zero mean  $U \neq 0$ , right-going waves are generated in large scales and left-going waves exist in the smaller scales after the critical wavenumber  $k_{\text{cr}} = L^{3/2}D^{-1/4}U^{1/2}$ .

The final equation for each the coefficient  $a_k$  can be simplified in the form concerning only the nonlinear coupling between different scales

$$\frac{da_k}{dt} = -D^{-3/2}E^{1/2}L^{-3/2}\frac{ik}{2} \sum_{|m| \leq \Lambda} a_m a_{m-k}^* \exp[i(\omega_k + \omega_{m-k} - \omega_m)t], \quad (12)$$

which defines the triad interactions between the modes  $a_k$ ,  $a_m$ , and  $a_{m-k}$ . The mean flow  $U$  in the dispersion relation makes no contribution in the triad wave interaction since its linear dependence on wavenumbers

$$\omega_k + \omega_{m-k} - \omega_m = -L^{-3}D^{1/2} \left[ k^3 + (m-k)^3 - m^3 \right] = 3L^{-3}D^{1/2}mk(m+k).$$

The resonant triad interactions happen at the modes with opposite wavenumbers  $m = -k$ .

Next, the autocorrelation functions characterizes the mixing properties of the turbulent system. It is usually useful to consider the autocorrelations in the spectral modes rather than at the physical grid points. The larger scales often get

longer correlation time than the smaller scale modes. Also from the dispersion relation (11), the high wavenumber mode gets a faster wave speed. The autocorrelation functions between two spectral modes can be computed as

$$\hat{\mathcal{R}}_{kl}(\tau) = \langle a_k(t + \tau) a_l^*(t) \rangle e^{-i\omega_k \tau} e^{-i(\omega_k - \omega_l)t} = \hat{\mathcal{C}}_{kl}(\tau) e^{-i\omega_k \tau} e^{-i(\omega_k - \omega_l)t}. \quad (13)$$

Above  $\hat{\mathcal{C}}_{kl}(\tau) = \langle a_k(t + \tau) a_l^*(t) \rangle$  is the autocorrelation matrix removing the dispersive waves from the original solution  $\hat{\mathcal{R}}(\tau) = \langle \hat{u}_k(t + \tau) \hat{u}_l^*(t) \rangle$  in the original spectral coefficients (3). The power spectrum of a stationary, ergodic process can be recovered by Fourier transform of the autocorrelation function [16]

$$\mathcal{R}(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S(\lambda) e^{i\lambda\tau} d\lambda, \quad S(\lambda) = \int_{-\infty}^{\infty} \mathcal{R}(\tau) e^{-i\lambda\tau} d\tau.$$

Especially for each spectral mode, the variance is the integration of the entire power spectrum by letting  $\tau = 0$

$$\hat{\mathcal{R}}_{kl}(0) = \left\langle |\hat{u}_k|^2 \right\rangle \delta_{kl} = \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{S}_{kl}(\lambda) d\lambda.$$

Usually, we assume the power spectra between different modes  $\hat{S}_{kl}, k \neq l$  to be relatively small compared with the diagonal terms  $\hat{S}_{kk}$  based on the homogeneous statistics.

### 3.2 Single trajectory simulations with different truncation sizes

The spectral equation (12) shows the interactions between modes in different scales. Obviously, a larger truncation size  $\Lambda$  will give a wider spectrum for the energy transfers between modes. Here, we first display the self-similarity in single trajectory solutions using different truncation sizes. We focus on the changes in the autocorrelation functions for the mixing scale of the system.

In the non-dissipative tKdV model (4), one important feature is the distinct solution structures from different initial states, representing different dynamical regimes in the phase space. We illustrate this issue by picking the initial states  $u_0(x) = u(x, 0)$  generated from the invariant Gibbs measures (7) with different inverse temperatures  $\theta$ . For the single trajectory simulation, only one sample is taken from the above Gibbs measure as the initial starting state. We compare the results starting from the initial states using large negative inverse temperature ( $\theta = -0.5$ ) and low negative inverse temperature ( $\theta = -0.1$ ) states from (7). More detailed statistical evolutions from ensemble simulations are shown next in Section 3.3.

To guarantee the conservation of the energy and Hamiltonian, a symplectic integrator [14, 12] is adopted for integrating the tKdV equation (4) in time. Different discrete model sizes  $J = 32, 16, 8$  (corresponding to wavenumber truncation  $\Lambda = 16, 8, 4$ ) are compared ( $J = 32$  is the standard case in the experiments in [12]). The detailed construction of the symplectic scheme for the tKdV model can be found in the supporting information in [12]. The other model parameters in the numerical simulations are fixed at  $E = 100, L = 6$ , consistent with the experimental data from the scale analysis in [12].

#### Time trajectories of the tKdV solutions with different truncation sizes

First, we compare the solution trajectories of the tKdV equation using different model truncation sizes. Figure 3.1 and 3.2 show typical realizations of the flow trajectories using initial states from both the Gibbs measures (7) with large  $\theta = -0.5$  and small  $\theta = -0.1$  inverse temperatures. The initial state determines the value of the conserved Hamiltonian in each solution. Comparing the two different initial cases, distinct performance in solutions can already be observed. The large negative inverse temperature case gives highly skewed solutions representing the frequent occurrence of the extreme events. With a small negative inverse temperature from the initial state, the solutions are not skewed and the dominant transporting peak along the waves can not be observed. The typical model phase transition can be discovered just from the deterministic single trajectory solution.

Then, by comparing different truncation sizes  $\Lambda = 16, 32, 4$ , the flow solutions gain more high frequency structures with a larger number of spectral modes. More smaller scale structures are observed in the larger truncation size case. By reducing the model truncation size, large scale spatial and temporal structures are maintained. The representative phase transition for the two initial state cases is unchanged in the process of reducing the model resolution. Even in the extreme truncation case with only  $\Lambda = 4$  interacting modes, the flow still shows relatively chaotic dynamics, and the phase transition from high skewed distribution to the near-Gaussian statistics is recovered.

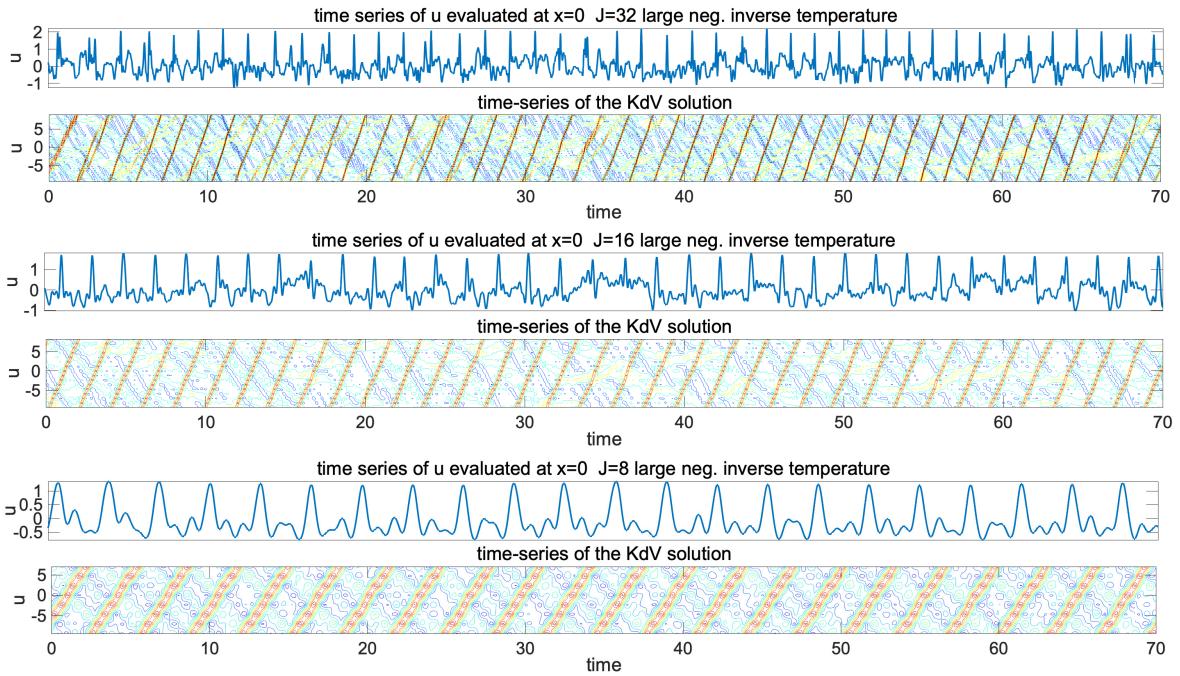


Figure 3.1: Trajectories from direct simulations of the tKdV equation with different truncation sizes  $\Lambda = 16, 8, 4$ . The initial state is taken from the Gibbs measure with a large negative inverse temperature.

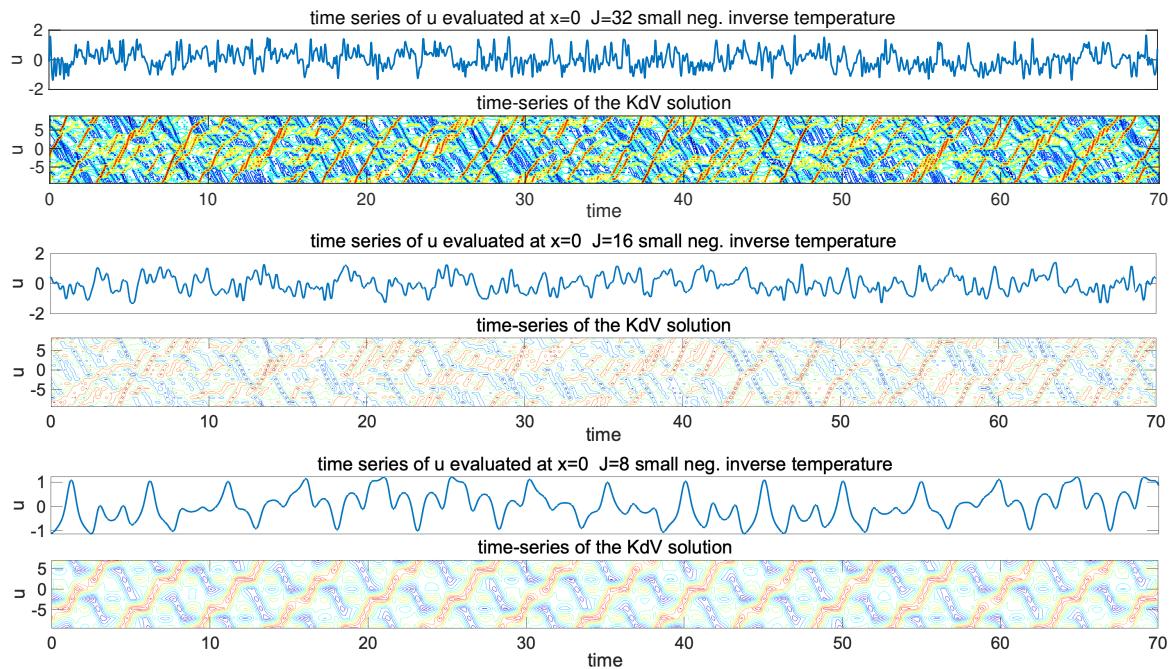


Figure 3.2: Trajectories from direct simulations of the tKdV equation with different truncation sizes  $\Lambda = 16, 8, 4$ . The initial state is taken from the Gibbs measure with a small negative inverse temperature.

## Autocorrelation functions and mixing rates

Next, we compare the autocorrelation functions and mixing properties in the models with different truncation sizes. In Figure 3.3, we compare the autocorrelation functions in the first three Fourier modes  $\hat{u}_k$ ,  $k = 1, 2, 3$  using both large and small negative inverse temperatures for initial data. In general, the states are correlated for longer time with smaller number of  $J$ . The small negative inverse temperature case generates much faster mixing rate. In comparison, the solutions with the skewed statistical state are correlated for a much longer time. This is related with the highly skewed transporting peak persistent in the solutions shown in Figure 3.1. The distinct performances in the autocorrelation functions with different inverse temperature offer a further illustration of the phase transition in two states. Especially for the  $J = 8$  case, we can observe that the solution is almost non-mixing with periodic trajectory in the small inverse temperature case, and becomes rapidly decaying in time when starts from a larger inverse temperature. In addition, it can be shown that the negative inverse temperature regime mixes in a faster rate than that of the positive inverse temperature [12].

For further illustration of the flow time series, we plot in the last row of Figure 3.3 the power spectra  $|\hat{S}(\omega)|$  by taking the Fourier transform of the time-series of the solutions  $u_\Lambda$ . Clearly with larger number of grid points  $J$  in the model, more high frequency modes are excited showing a wider spectra of small scale variability, while the low frequency stays relatively the same with different truncation sizes. Also in the intermediate frequency range, the decay slopes with different truncation sizes keep relatively the same, showing the invariant small scale fluctuation amplitude in all the models.

### 3.3 Model statistics from ensemble simulations and Gibbs invariant measures

Previously, single trajectory simulations of the tKdV equation are compared with different truncation sizes. For each single trajectory solution, the Hamiltonian stays in one conserved value. More precisely, the model statistics should be characterized by an ensemble simulation starting from different initial values of the Hamiltonian. The uncertainty of the model then comes from the initial distribution of the state variable.

In the ensemble simulations, we still compare the three truncation sizes  $J = 32, 16, 8$ . The ensemble size is picked as  $N = 10000$ . The same model parameters are used the same as Section 3.2. The initial ensemble is sampled from the Gibbs measure (7) using different inverse temperatures.

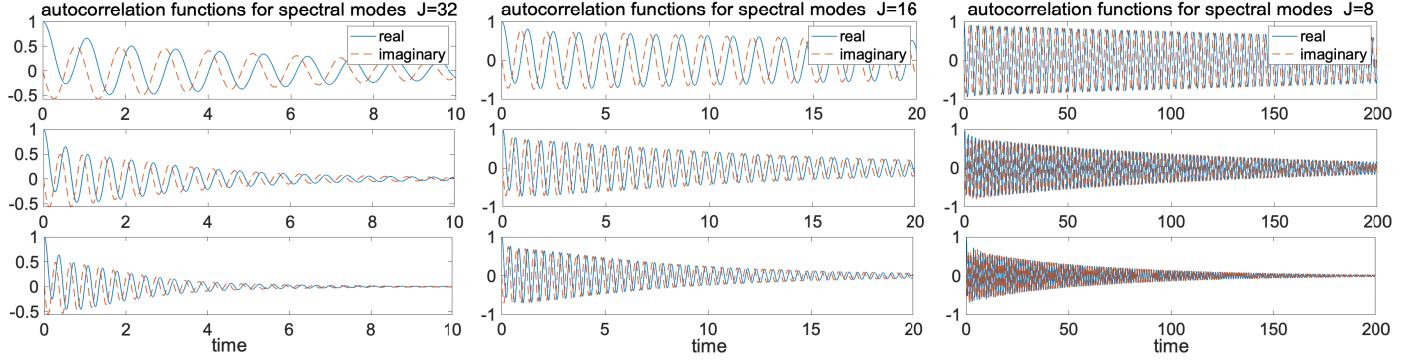
#### Equilibrium energy spectra from different model truncation sizes

First, we compare the equilibrium energy spectra for the energy in each mode in the truncated model with different model truncation sizes. The total energy of the system is always normalized to the constant value  $\mathcal{E} = 1$ . Since the total energy in each case is fixed, the energy in each mode increases with a smaller number of total modes. Figure 3.4 plots the equilibrium variance in each mode. Similar trends in the spectra are created in the three different truncation cases. With the large negative inverse temperature, a decaying spectrum is shown; while with the near-Gaussian small negative inverse temperature state, the energy spectra become almost flat with more energetic modes among the smallest scales. The low inverse temperature case with more active smaller scales confirm the time-series in Figure 3.2 for more fluctuations without a dominant large scale structure.

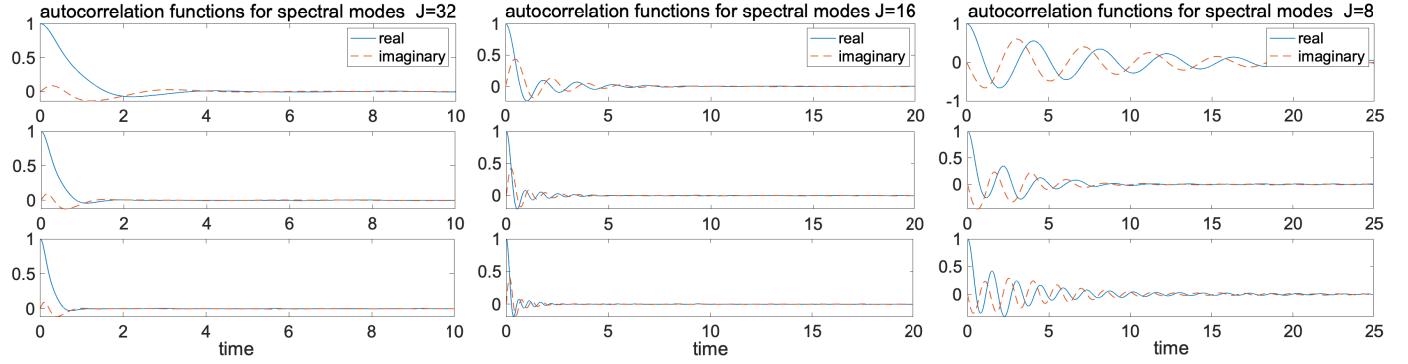
#### Probability density functions in the two cases

Next, we show the final equilibrium PDFs of the truncated state variable  $u_\Lambda$  from the direct ensemble simulations of the model. The final distribution from direct model simulation is also compared with the invariant Gibbs measure prediction directly sampled from the distribution (7). A proper Markov chain Monte Carlo strategy [12] is adopted for accurate sampling of the target distributions. Figure 3.5 gives the results with the different truncation sizes. It shows that the Gibbs measure as the initial state gets maintained by the tKdV model dynamics, and gives good approximation for the final state distributions.

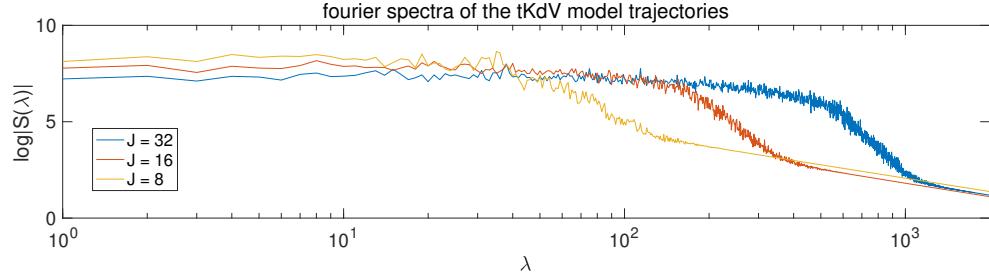
Starting from a smaller negative inverse temperature, the final equilibrium PDF becomes near Gaussian. This case can be compared with the time trajectories in Figure 3.2, where the solution is symmetric in positive and negative directions. On the other hand, a larger negative inverse temperature generates a highly skewed PDF of  $u_\Lambda$ . Stronger skewness can be induced with a smaller truncation size. A bimodal distribution is observed in the severe truncation case  $J = 8$ . This shows the typical statistical phase transition from the previous faster mixing near-Gaussian state according to the changing values in the inverse temperature. The three different truncation cases display similar trend in the statistical transition from near-Gaussian distribution to a highly skewed distribution to the positive side. This implies that a severe truncation



(a) autocorrelations from solutions with large negative inverse temperature



(b) autocorrelations from solutions with small negative inverse temperature



(c) power spectra  $\log \hat{S}(\lambda)$  from the tKdV solutions with different truncation sizes

Figure 3.3: Autocorrelation functions of the first three Fourier modes  $\hat{u}_1, \hat{u}_2, \hat{u}_3$  with different truncation sizes  $\Lambda = 16, 8, 4$ . Results from large negative inverse temperature (upper row) and small negative inverse temperature (middle row) are also compared. The power spectra from the time series with different truncation sizes are compared in the bottom row.

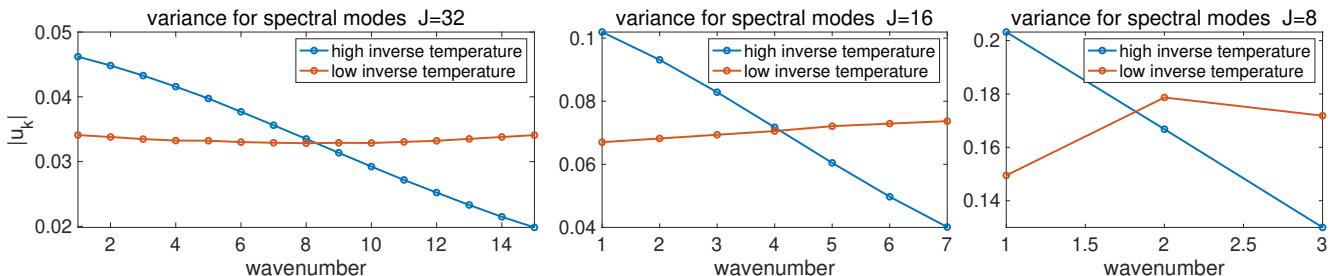


Figure 3.4: Energy spectra for the equilibrium variances in each mode  $\hat{u}_k$  from the tKdV ensemble simulations with different truncation sizes  $\Lambda = 16, 8, 4$ . Two cases with different negative inverse temperatures are compared.

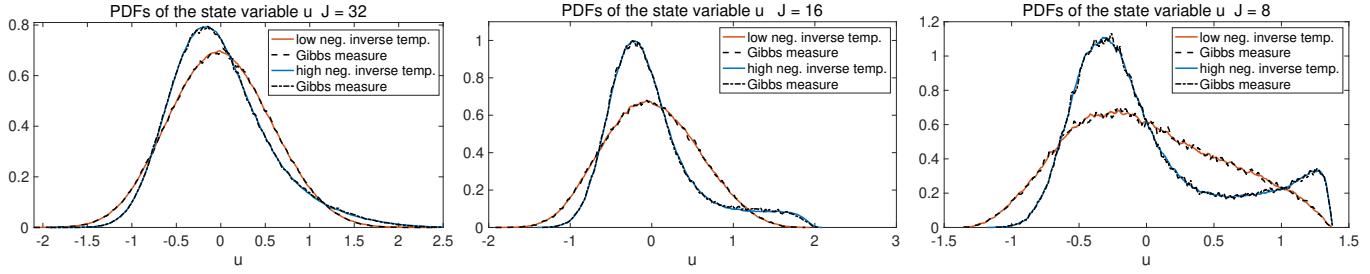


Figure 3.5: The PDFs of the state variable  $u_\Lambda$  from direct model ensemble simulations. The results using initial flow statistics from different negative inverse temperatures are compared in solid lines. The dashed lines show the corresponding PDFs predicted directly from the Gibbs invariant measures in the two cases.

model still serves as a desirable model for investigating the phase transition by maintaining the representative properties, while enjoys more trackable model dynamics.

## 4 Phase Transition in a Paradigm Two-Mode Interacting tKdV Model

The numerical results with different model truncation sizes in Section 3 illustrate the invariant trend of the phase transition as we alter the inverse temperature in the Gibbs measure. Especially even with a severe truncation with  $\Lambda = 4$  interacting modes, the typical statistical phase transition from near-Gaussian distribution to the highly skewed extreme event regime is still maintained. This implies the possibility of pushing the tKdV model to even further extreme with  $\Lambda = 2$  modes. Although the system with the radical truncation is no longer mixing, the integrable dynamics of this paradigm two-mode framework enables us to achieve a clear characterization about the phase transition process and gain a better understanding for the mechanism in the generation of extreme events.

### 4.1 A Two-Mode Interacting tKdV Model with Exact Integrable Dynamics

In the extreme case, the dynamics of the tKdV model (4) is constrained on a two-dimensional invariant subspace, that is, to consider the following two modes of wavenumber  $l$  and  $2l$

$$u_\Lambda = \hat{u}_1 e^{ilx} + \hat{u}_2 e^{i2lx} + c.c., \quad \text{with } 2l \leq \Lambda < 3l,$$

where  $|k| \leq \Lambda$  is the largest wavenumber truncation in the original tKdV model, and *c.c.* for the complex conjugate. All the other modes  $k \neq l, 2l$ ,  $|k| \leq \Lambda$  stay zero during the evolution of the two-mode model since the nonlinear coupling constrains the energy inside the subspace. The two modes  $(\hat{u}_1, \hat{u}_2)$  with wavenumbers  $l, 2l$  form a closed system. The corresponding explicit equations for the two modes  $(\hat{u}_1, \hat{u}_2)$  can be found as the two-mode case of the spectral equation (12) as

$$\begin{aligned} \frac{d\hat{u}_1}{dt} + ilC_1\hat{u}_1^*\hat{u}_2 - il^3C_2\hat{u}_1 &= 0, \\ \frac{d\hat{u}_2}{dt} + ilC_1\hat{u}_1^2 - i8l^3C_2\hat{u}_2 &= 0, \end{aligned} \tag{14}$$

with the coefficients  $C_1 = E^{1/2}L^{-3/2}D^{-3/2}$  and  $C_2 = L^{-3}D^{1/2}$  for simplicity in the representation. The conserved total energy of the system (14) is still normalized to 1, that is,  $\mathcal{E} = 2\pi(|\hat{u}_1|^2 + |\hat{u}_2|^2) = 1$ . The conserved Hamiltonian (5) can be also computed explicitly as

$$\mathcal{H} = C_1 H_3 - C_2 H_2 = 2\pi \left[ C_1 \operatorname{Re}(\hat{u}_1^2 \hat{u}_2^*) - C_2 l^2 (|\hat{u}_1|^2 + 4|\hat{u}_2|^2) \right].$$

Finally, the invariant Gibbs measure (7) for this two-mode system can be written in the following explicit form

$$\mathcal{G}_\theta(\hat{u}_1, \hat{u}_2) = C_\theta^{-1} \exp \left[ -2\pi\theta \left( C_1 \operatorname{Re}(\hat{u}_1^2 \hat{u}_2^*) - C_2 l^2 (|\hat{u}_1|^2 + 4|\hat{u}_2|^2) \right) \right] \delta \left( 2\pi (|\hat{u}_1|^2 + |\hat{u}_2|^2) - 1 \right),$$

given the inverse temperature  $\theta$ . Typical solutions of the two-mode model can be found next in Section 4.2.

## Special exact solution and its instability to perturbations

The two-mode tKdV model (14) can be organized into the conditional Gaussian framework [11]. For the second equation for  $\hat{u}_2$ , it can be determined directly from the realization of the first state  $\hat{u}_1$ , that is,

$$\hat{u}_2(t) = -ilC_1 \int_0^t e^{i8l^3C_2(t-s)} \hat{u}_1^2(s) ds, \quad (15)$$

with the initial value  $\hat{u}_2(0) = 0$ .

Another single plane wave solution can be found starting from the initial state  $\hat{u}_1(0) = 0, \hat{u}_2(0) = \hat{u}_{2,0}$ ,

$$\hat{u}_1(t) = 0, \quad \hat{u}_2(t) = \hat{u}_{2,0} [\cos(8l^3C_2t) + i \sin(8l^3C_2t)], \quad |\hat{u}_{2,0}| = 1. \quad (16)$$

Still, the general solution of the above equation (14) is not easy to solve directly. For the evolution of a small perturbation  $\hat{u}_1$  according to the steady state  $\hat{u}_2 = |\hat{u}_2| e^{i\theta_2}$ , it can be decomposed into the real and imaginary components  $\hat{u}_1(t) = a_1(t) + ib_1(t)$

$$\frac{d}{dt} \begin{bmatrix} a_1 \\ b_1 \end{bmatrix} = C_1 l |\hat{u}_2| \begin{bmatrix} -\sin \theta_2 & \cos \theta_2 \\ \cos \theta_2 & \sin \theta_2 \end{bmatrix} \begin{bmatrix} a_1 \\ b_1 \end{bmatrix} + C_2 l^3 \begin{bmatrix} 1 & -1 \end{bmatrix} \begin{bmatrix} a_1 \\ b_1 \end{bmatrix}.$$

The nonlinear coefficient can be always decomposed into an unstable direction with growth  $C_1 l |\hat{u}_2|$  and a stable direction with damping  $-C_1 l |\hat{u}_2|$ . Then the periodic solution above for  $\hat{u}_2$  is always unstable to perturbations in the first mode  $\hat{u}_1$  depending on the amplitude of the second mode  $|\hat{u}_2|$ . Considering the conservation of total energy, the instability due to the nonlinear coupling transfers energy from the higher wavenumber mode  $\hat{u}_2$  to larger scale  $\hat{u}_1$ .

## Exact integrable structure of the two-mode tKdV system

For a better illustration about the two-mode model structures, we can rescale the original system (14) to find a cleaner formulation. To focus on the nonlinear coupling effect, we can further introduce the mode decomposition  $(\hat{v}_1, \hat{v}_2)$  with phase shifts to the original state variables as shown in (10)

$$u_\Lambda(x, \tau) = \hat{u}_1 e^{ilx} + \hat{u}_2 e^{i2lx} + c.c. = \hat{v}_1 e^{i(lx + \omega_l \tau)} + \hat{v}_2 e^{i(2lx + 8\omega_l \tau)} + c.c.$$

with  $\hat{u}_1 = \hat{v}_1 e^{i\omega_l \tau}$  and  $\hat{u}_2 = \hat{v}_2 e^{i8\omega_l \tau}$ . We further rescale the time  $\tau$  and the dispersion effect  $\omega_l$  for the new system according to the model parameters

$$\tau = C_1 l t, \quad \omega_l = \frac{C_2}{C_1} l^2 = L_0^{-\frac{3}{2}} D_0^2 E_0^{-\frac{1}{2}} l^2. \quad (17)$$

Then the two-mode dynamics (14) is then modified to the form focusing on the nonlinear coupling term

$$\frac{d\hat{v}_1}{d\tau} + ie^{i6\omega_l \tau} \hat{v}_1^* \hat{v}_2 = 0, \quad \frac{d\hat{v}_2}{d\tau} + ie^{-i6\omega_l \tau} \hat{v}_1 \hat{v}_2 = 0.$$

With no dispersive effect  $\omega_l \equiv 0$ , the above system reduces to the exact *truncated Burgers-Hopf* (TBH) equation discussed extensively in [8].

To further simplify the model dynamics, we introduce a change of variables for the amplitude and phase

$$\hat{v}_1(\tau) = \frac{r_1(\tau)}{\sqrt{2\pi}} e^{i\vartheta_1(\tau)}, \quad \hat{v}_2(\tau) = \frac{r_2(\tau)}{\sqrt{2\pi}} e^{i\vartheta_2(\tau)}, \quad r_1^2 + r_2^2 = 1.$$

The scaling factor is introduced according to the conserved energy  $2\pi (|\hat{v}_1|^2 + |\hat{v}_2|^2) = 1$ . Expressing the equations under the new variables and separating the real and imaginary parts, we have the new set of equations

$$\begin{aligned} \frac{dr_1}{d\tau} + \frac{r_1 r_2}{\sqrt{2\pi}} \sin(2\vartheta_1 - \vartheta_2 - 6\omega_l) &= 0, & \frac{d\vartheta_1}{d\tau} + \frac{r_2}{\sqrt{2\pi}} \cos(2\vartheta_1 - \vartheta_2 - 6\omega_l) &= 0, \\ \frac{dr_2}{d\tau} - \frac{r_1^2}{\sqrt{2\pi}} \sin(2\vartheta_1 - \vartheta_2 - 6\omega_l) &= 0, & \frac{d\vartheta_2}{d\tau} + \frac{r_1^2 r_2^{-1}}{\sqrt{2\pi}} \cos(2\vartheta_1 - \vartheta_2 - 6\omega_l) &= 0, \end{aligned}$$

The corresponding conserved Hamiltonian can be rewritten as

$$\mathcal{H} = C_1 \left[ \frac{(1 - r_2^2) r_2}{\sqrt{2\pi}} \cos(2\vartheta_1 - \vartheta_2 - 6\omega_l \tau) - \omega_l (3r_2^2 + 1) \right].$$

Above the conservation of energy relation  $r_1^2 + r_2^2 = 1$  is used. Observing the above dynamical equations and the Hamiltonian, clearly the solution is determined by the amplitude of the second mode  $r_2$  and the phase difference  $2\vartheta_1 - \vartheta_2$ .

According to the observations in the structure of the two-mode model, we introduce the two new variables  $r \equiv r_2$  and  $\phi = 2\vartheta_1 - \vartheta_2 - 6\omega_l\tau$ . The two-mode tKdV model can be finally formulated in the clean form about the amplitude  $r$  and the phase difference  $\phi$

$$\begin{aligned} \frac{dr}{d\tau} - \frac{1-r^2}{\sqrt{2\pi}} \sin \phi &= 0, \\ \frac{d\phi}{d\tau} + \frac{r^{-1}(3r^2-1)}{\sqrt{2\pi}} \cos \phi &= -6\omega_l, \end{aligned} \quad (18)$$

with the frequency  $\omega_l = C_2 l^2 / C_1$  and scaled time  $\tau = C_1 l t$ . The amplitude of the first mode  $\hat{v}_1$  can be computed from  $r = r_2$  using the energy conservation  $r_1 = \sqrt{1-r^2}$ , and the separated phases for  $\vartheta_1, \vartheta_2$  can be discovered from  $\phi$  by integrating the separate equations. Therefore, the two-mode model can be determined by the integrable formulation (18) about the two variables  $(r, \phi)$ .

The conserved Hamiltonian functional then can be simplified based on the final model variables as

$$\tilde{\mathcal{H}} = C_1^{-1} \mathcal{H} + \omega_l = \frac{(1-r^2)r}{\sqrt{2\pi}} \cos \phi - 3\omega_l r^2. \quad (19)$$

The new Hamiltonian is normalized as  $\tilde{\mathcal{H}} = C_1^{-1} \mathcal{H} + \omega_l$  by the two constants  $C_1$  and  $\omega_l$ . The conservation of the quantity  $\tilde{\mathcal{H}}$  can be checked by directly putting (19) back to the dynamical equation (18). The corresponding Gibbs measure becomes

$$\tilde{\mathcal{G}}_\theta(r, \phi) = C_\theta^{-1} \exp \left[ \theta \left( \frac{r^3 - r}{\sqrt{2\pi}} \cos \phi + 3\omega_l r^2 \right) \right], \quad (20)$$

with  $0 < r < 1$ . Thus the distribution of the system can be entirely determined by the two variables  $(r, \phi)$ .

**Fixed point solutions of the integrable system** The periodic structure in the solutions of the two-mode model can be seen more clearly from the above formulation represented by the new states  $(r, \phi)$ . Observing from the structures of the dynamical equations (18), there exist two fixed point solutions:

- i) Fixed point solution with phase shift  $\phi = 2\vartheta_1 - \vartheta_2 - 6\omega_l\tau = 0$ : the amplitude of the second mode stays in a constant as  $r_*^+ = r_2 = \sqrt{2\pi\omega_l^2 + \frac{1}{3}} - \sqrt{2\pi}\omega_l$ . The Hamiltonian reaches its maximum value  $\tilde{\mathcal{H}}_+ = \frac{r_*^+ - (r_*^+)^3}{\sqrt{2\pi}} - 3\omega_l (r_*^+)^2$ ;
- ii) Fixed point solution with phase shift  $\phi = 2\vartheta_1 - \vartheta_2 - 6\omega_l\tau = \pm\pi$ : the amplitude of the second mode stays in a constant as  $r_*^- = r_2 = \sqrt{2\pi\omega_l^2 + \frac{1}{3}} + \sqrt{2\pi}\omega_l$ . This fixed point is only possible with the permitted value  $r_*^- < 1$ . This is possible only when  $\omega_l = \frac{C_2}{C_1} l^2 < \frac{1}{3\sqrt{2\pi}}$ . In this case, the Hamiltonian shrinks to its minimum value  $\tilde{\mathcal{H}}_- = \frac{r_*^- - (r_*^-)^3}{\sqrt{2\pi}} - 3\omega_l (r_*^-)^2$ .

Since the above two fixed points are extrema of the Hamiltonian, they are both stable if exist. The solutions are consistent with previous exact solutions in (15). In addition, we can find another set of solutions acting as a limit cycle

- iii) Limiting ring solution with  $r = r_2 = 1$ : then  $r_1 = 0$  and  $\phi$  follows the equation,  $\dot{\phi} = -\sqrt{\frac{2}{\pi}} \cos \phi - 6\omega_l$ . In the case  $\omega_l < \frac{1}{3\sqrt{2\pi}}$ , the solution  $r \equiv 1$  is unstable to perturbations. On the other hand with  $\omega_l > \frac{1}{3\sqrt{2\pi}}$ , the solution  $r \equiv 1$  becomes stable (see Figure 4.2 and 4.4 for the phase plots).

The above limiting ring solution is from the exact solution example in (16). With the cleaner formulation here, it is more direct to see the unstable dynamics in the solutions.

**Integrable structure with the conserved system** Finally, combining the dynamics for  $r$  and the conserved Hamiltonian with the value  $\tilde{H}$ , we have the equality for the velocity  $\dot{r}$

$$\frac{1}{2} \left( \frac{dr}{d\tau} \right)^2 = \frac{(1-r^2)^2}{4\pi} \sin^2 \phi = \frac{(1-r^2)^2}{4\pi} \left[ 1 - \frac{2\pi}{(1-r^2)^2 r^2} (\tilde{H} + 3\omega_l r^2)^2 \right] = \frac{(1-r^2)^2}{4\pi} - \frac{1}{2} (\tilde{H} r^{-1} + 3\omega_l r)^2.$$

$\theta$	-1	-0.5	-0.2	-0.1	0	0.1	0.2	0.5	1
skewness	0.2345	0.1177	0.0466	0.0216	0.0005	-0.0232	-0.0519	-0.1181	-0.2276
kurtosis	2.0151	2.0051	1.9992	2.0007	1.9984	2.0006	2.0006	2.0033	2.0127

Table 1: Skewness and kurtosis of the two-mode state  $u_\Lambda$  from the sampled Gibbs measures with different inverse temperatures.

By rearranging the above relation, we find the analog particle energy in kinetic and potential energy parts for  $r \in [0, 1]$

$$I(\dot{r}, r) = \frac{1}{2}\dot{r}^2 + V(r) = \frac{1}{4\pi} - 3\omega_l \tilde{H}, \quad V(r) = -\frac{r^4}{4\pi} + \left(\frac{1}{2\pi} + 3\omega_l^2\right)r^2 + \frac{\tilde{H}^2}{2r^2}, \quad (21)$$

with the constant  $\tilde{H} = C_1^{-1}H + \omega_l$ . The two-mode system is then determined by the initial states of  $\dot{r}(0)$  and  $r(0)$ . Again in the non-dispersive case  $C_2 = 0$  then  $\omega_l \equiv 0$ , the above first integral (21) reduces to the TBH equation case derived in [8].

Especially, we see that the energy  $\frac{1}{2}\dot{r}^2$  must stay in positive values. This sets a constraint for the range of values of the normalized Hamiltonian  $\tilde{H}$ , that is,  $\min_r V(r) \leq \frac{1}{4\pi} - 3\omega_l \tilde{H}$ . Let  $r_* = \arg \min_r V(r)$ . The values of the Hamiltonian must change under the following range

$$\frac{\tilde{H}^2}{2r_*^2} - 3\omega_l \tilde{H} \leq \frac{r_*^4}{4\pi} - \left(\frac{1}{2\pi} + 3\omega_l^2\right)r_*^2 + \frac{1}{4\pi}.$$

The above inequality gives the range of permitted values for the Hamiltonian  $\tilde{H} \in [H_-, H_+]$ . As the equality in the above relation is reached, the only permitted value is  $\dot{r} \equiv 0$ . Then  $r_2 = \sqrt{2\pi\omega_l^2 + \frac{1}{3}} - \sqrt{2\pi\omega_l}$  and  $r_1 = \sqrt{1 - r_2^2}$ . This is the fixed point solution with the maximum Hamiltonian. The two modes are just moving in time with the constant amplitude. This forms the bimodal distribution in the physical state  $u_\Lambda$ .

## 4.2 Statistical phase transition in the Gibbs measure with different inverse temperatures

With the integrable structures of the two-state model, we are able to investigate the mechanism in the phase transition observed universally in Section 3 for various truncation sizes. For simplicity we pick  $l = 1$ . The initial state is constrained on the normalized energy surface  $2\pi(|\hat{u}_1|^2 + |\hat{u}_2|^2) = 1$ . We can control the value of the Hamiltonian  $H$  by changing the inverse temperature  $\theta$  in the Gibbs measure,  $\mathcal{G}_\theta \sim \exp(-\theta\tilde{H})$ . Usually, a negative inverse temperature  $\theta < 0$  emphasizes the positive value of the Hamiltonian, implying a positive skewness; while the positive inverse temperature adds more weight on the negative value of the Hamiltonian, generating a negative skewness. For convenience, we focus on the negative values of  $\theta < 0$  which represents the realistic case from the experiments [3, 12].

By sampling from the invariant Gibbs measure (20), the PDFs of the two-mode state  $u_\Lambda$  display the clear statistical transition from a symmetric distribution (near small  $\theta \sim 0$ ) to a bimodal structure (large  $|\theta|$ ). Figure 4.1 shows the distribution functions and two-mode spectra sampled from the Gibbs measure using the two-mode model with different values of the inverse temperature  $\theta$ . The distributions are symmetric in both the PDFs of the Hamiltonian and the state variable with  $\theta = 0$ . Highly skewed statistics are generated using large amplitudes of the inverse temperature. The slope of the energy spectrum (in this case, the difference between the two Fourier modes) also becomes steeper with a high skewness in larger values of  $\theta$ . The skewness and kurtosis in the tested cases are shown in Table 1. This offers a clear presentation of the statistical phase transition process to created highly skewed statistics by controlling the inverse temperature  $\theta$  from this simple two-mode system.

### Phase transition based on the conserved quantities

This phase transition from the symmetric distribution to the bimodal distribution can be explained from checking the contours of the normalized Hamiltonian. From the original definition (5), larger values of  $H$  shows stronger skewness in third moments, while smaller value of  $H$  shows less skewed near-Gaussian statistics. To display the different dynamical performance with different conserved Hamiltonian, Figure 4.2 plots the contours of the normalized Hamiltonian  $\tilde{H}$  in (19)

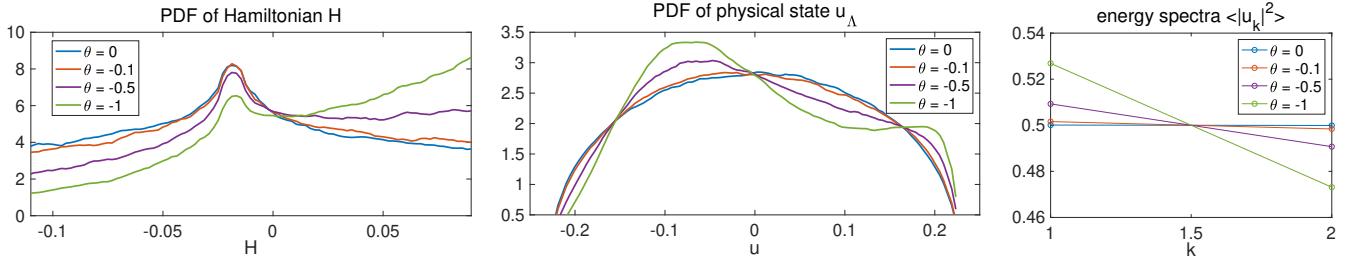


Figure 4.1: Model statistics and PDFs sampled from the Gibbs invariant measures. Different values of the inverse temperature  $\theta$  are compared.

as a function of  $r$  and  $\phi$  and the contours of the particle energy  $I$  in (21), for both the incoming flow ( $D = 1$ ) and outgoing flow ( $D = 0.24$ ) cases.

With small amplitudes of the inverse temperature  $|\theta| \sim 0$  in the Gibbs measure, the samples starting with small Hamiltonian  $\tilde{\mathcal{H}}$  have higher chance to be sampled. Among small values of the Hamiltonian  $|\tilde{\mathcal{H}}| \sim 0$ , the particle trajectory experiences abrupt change from  $r = 0$  to  $r = 1$ . As in the dashed lines in Figure 4.2a, the particle trajectory changes through the entire range from 0 to 1. Thus the two modes  $\hat{u}_1, \hat{u}_2$  exchange energy rapidly between each other, leading to a balanced symmetric structure in the physical state  $u_\Lambda$ . Accordingly, the potential energy  $V(r)$  and the kinetic energy  $\frac{1}{2}r^2$  in the particle energy  $I$  can also vary among a wide range. This corresponds to the changes of amplitudes in the time evolution of  $\hat{u}_2$  shown next in the time trajectory of Figure 4.3c.

With large amplitude of the negative inverse temperature  $\theta < 0$ , the Gibbs measure gets strong emphasize on the smallest value of the Hamiltonian. In this case, among the deep blue contours for  $H$  in Figure 4.2a (corresponding to the center localized small contours of  $I$  in Figure 4.2b), the value  $r$  (the amplitude of  $\hat{u}_2$ ) is constrained inside a narrow range. In the extreme case, The second mode  $\hat{u}_2$  reaches the fixed point  $r \equiv r_*$  (that is, the fixed point solution discussed in Section 4.1). This assigns one dominant scale with unchanged fixed amplitude  $r_*$  in one mode  $\hat{u}_2$ , and the other mode  $\hat{u}_1$  is fixed at another amplitude  $\sqrt{1 - r_*^2}$ . It leads to the bimodal structure in the distribution function in Figure 4.1. The corresponding time-series is shown in the first row of Figure 4.3b. Note that the center circles (and finally a single point at value  $r = r_*$  with zero velocity  $v = \dot{r} = 0$ ) in the contours for  $I(v, r)$  shown in Figure 4.2b correspond to the stable fixed trajectories with large amplitudes of the Hamiltonian  $\tilde{\mathcal{H}}$ . This creates the highly skewed bimodal structure in the distribution functions.

In addition, the incoming flow with larger water depth  $D = 1$  shows a wider transition regime (the dashed lines in Figure 4.2a) compared with the sharp transition regime in the outgoing flow with shallower water depth  $D = 0.24$ . This also guarantees the faster transition to the highly skewed statistics after the flow goes through the ADC to the downstream.

The typical realizations of the solution structures from direct numerical simulation of the two-mode model are shown in Figure 4.3. The typical structures from different inverse temperature  $\theta$  are realized in the three trajectories according to different values of the Hamiltonian consistent with our previous discussions based on the phase contours:

- In a large negative value of the Hamiltonian  $H$  (in Figure 4.3a corresponding to a large positive value of  $\theta > 0$ ), the mode amplitude  $r$  varies inside a narrow range near a constant, and the phase difference  $\phi$  jumps between  $\pm\pi$ . This leads to a skewed physical solution  $u_\Lambda$  toward the negative value. This corresponds to the blue contours in Figure 4.2;
- In a large positive value of the Hamiltonian  $H$  (in Figure 4.3b corresponding to a large negative value of  $\theta < 0$ ), there are again two different scales in  $\hat{u}_1, \hat{u}_2$  changing inside a narrow range according to the center localized circle in Figure 4.2b for  $I$ . The amplitude  $r$  stays in a nearly constant value, and the phase difference  $\phi$  between the two scales stays near 0. The physical solution  $u_\Lambda$  shows clear separation of two scales and a bimodal distribution skewed on the positive side. This corresponds to the red contours in Figure 4.2;
- In a small value of the Hamiltonian  $H$  (in Figure 4.3c corresponding to a small value of  $\theta \sim 0$ ), the amplitudes in  $r_1, r_2$  and phase  $\phi$  change rapidly between the entire range. This corresponds to the black dashed line in phase plot in Figure 4.2. In this way, the two different scales are rapidly mixing in time. The physical solution distribution transits to the symmetric single peak structure.

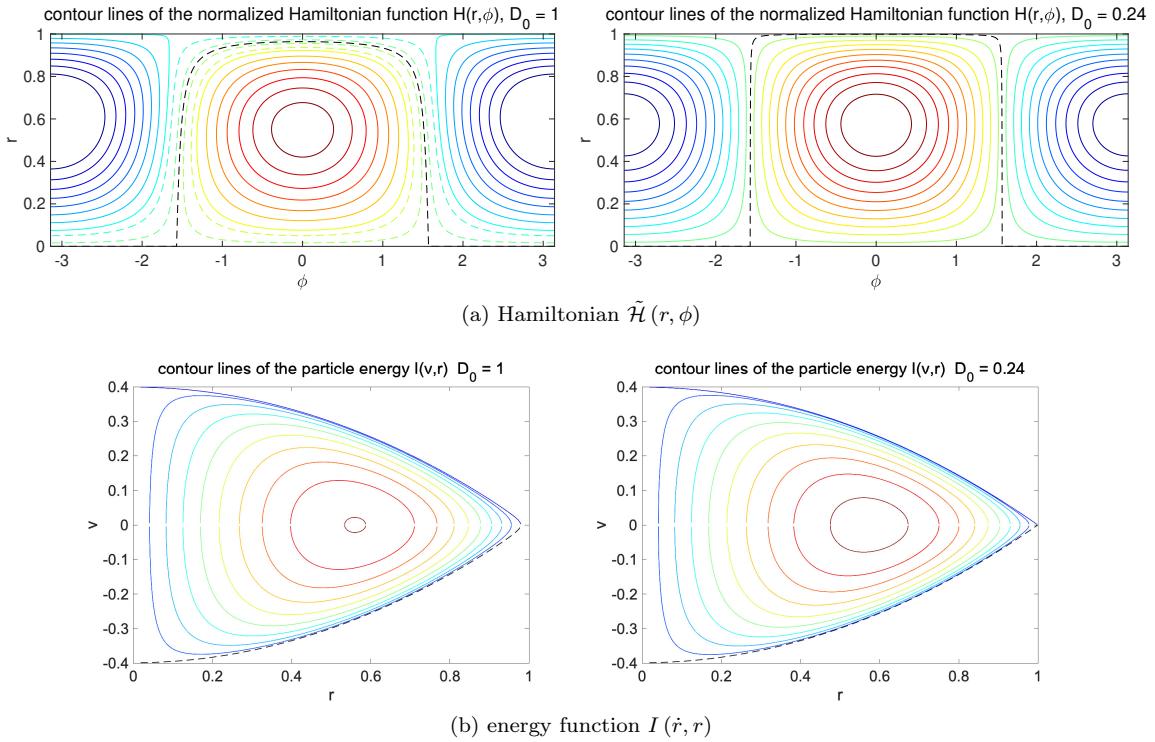


Figure 4.2: Contours of the normalized Hamiltonian  $\tilde{H}$  and the total particle energy  $I(\dot{r}, r)$  from the two-mode model. The incoming flow ( $D_0 = 1$ , left) and outgoing flow ( $D_0 = 0.24$ , right) cases are compared with the base mode as  $l = 1$ . In the contour lines for the Hamiltonian, dashed lines are for small values  $\tilde{H} \in [-0.2, 0.2]$ , the solid lines are for  $|\tilde{H}| > 0.2$ . The black dashed line is for the contour of  $\tilde{H} = 0$ .

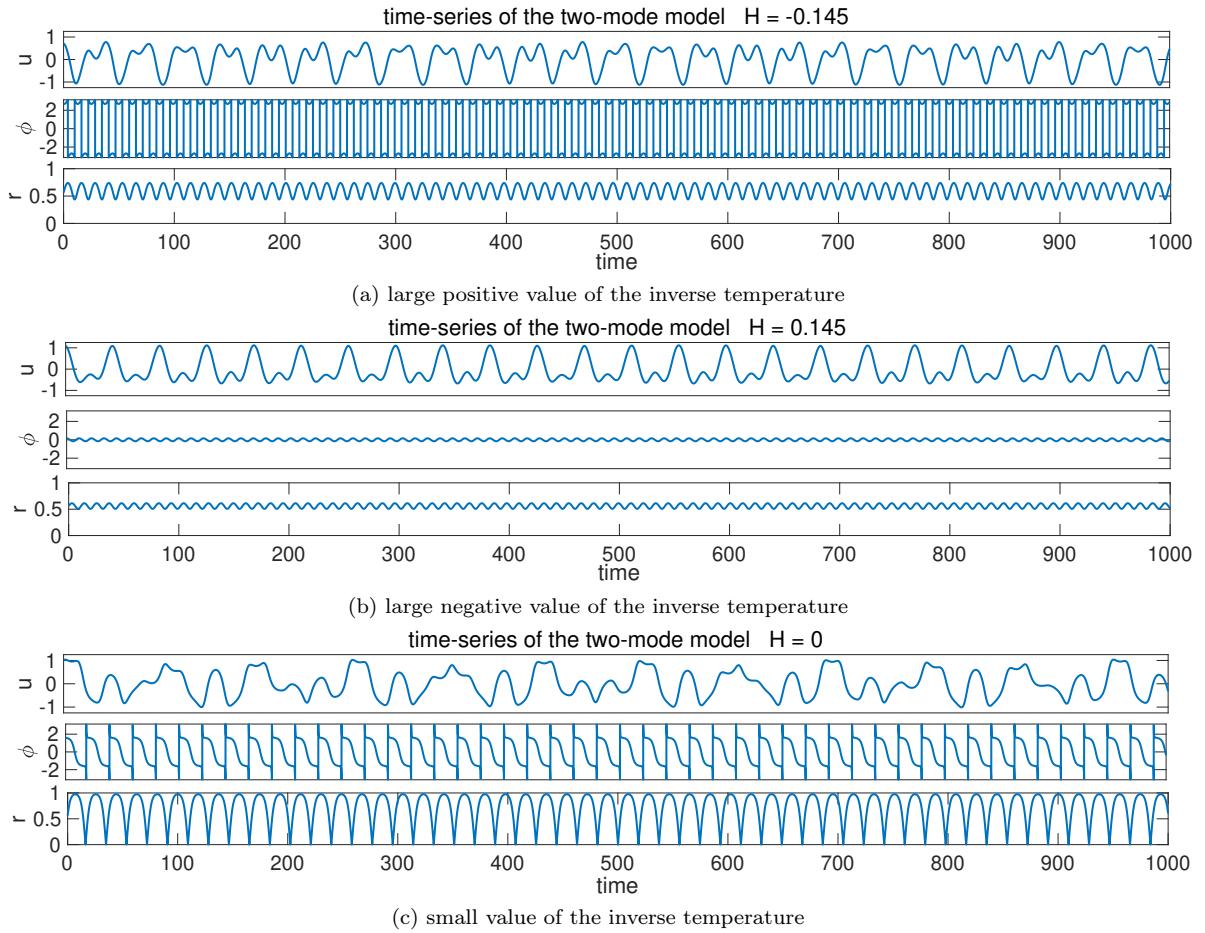


Figure 4.3: Time trajectories of the two-mode model solutions with three typical values of the Hamiltonian  $\tilde{H}$ . The solutions in the physical domain  $u_\Lambda$ , as well as the amplitude  $r$  and phase difference  $\phi$  of the two modes  $\hat{u}_1, \hat{u}_2$  in (18) are compared.

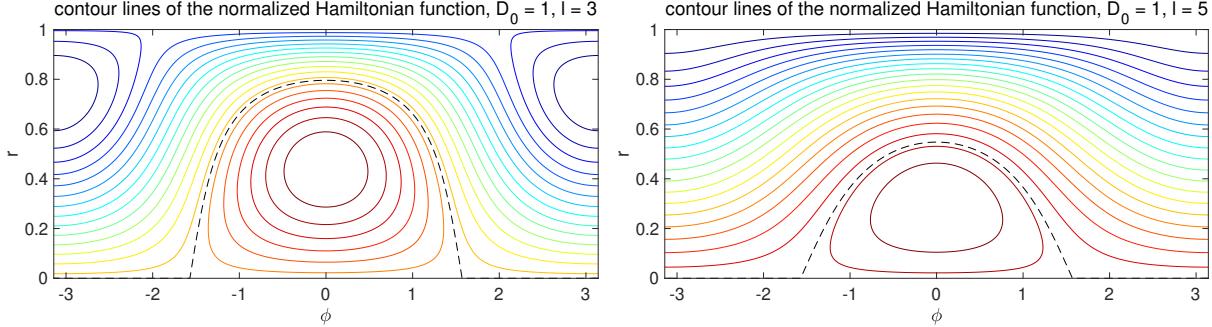


Figure 4.4: Contours of the normalized Hamiltonian  $\tilde{H}$  with the base mode wavenumber  $l = 3$  and  $l = 5$ . The black dashed line marks the location of the contour where  $\tilde{H} = 0$ .

### Two-mode coupling with higher wavenumber

The previous discussions are all based on the first two modes  $l = 1$  for the state (14) in the subspace. As a generalization, we can change the base mode to larger wavenumber  $l$  to check the coupling between the two-mode subspace  $(l, 2l)$ . Figure 4.4 plots the contours of the normalized Hamiltonian with larger wavenumber in the base mode  $l = 3$  and  $l = 5$ . The change is introduced to a larger value of the frequency  $\omega_l = \frac{C_2}{C_1}l^2$ . Consistent with the analytic results in Section 4.1, the fixed point for the maximum value of the Hamiltonian  $\tilde{H}^+$  persists, while the position of the minimum value of  $\tilde{H}^-$  is gradually shifted upward and finally vanishes to a stable limiting ring periodic solution. A final comment is that this change in shift space is typical for the incoming flow with  $D_0 = 1$ . With the outgoing flow  $D_0 = 0.24$ , the ratio  $C_2/C_1$  always stays in small value. Thus this phase space transition is less easier to observe.

### 4.3 Instabilities of the two-mode model solutions in higher dimensional phase space

In this last part, we check the instability of the two-mode state inside the invariant subspace. To show this, we consider the solution evolution starting from the non-zero initial state in the two-dimensional subspace inside a larger space of dimensionality  $\Lambda$

$$\hat{u}_k(0) = \begin{cases} \sqrt{E/2}, & k = l \text{ and } 2l, \\ 0, & k \neq l, 2l, \end{cases} \quad 2l \leq \Lambda < 3l.$$

With perturbations introduced from the numerical scheme, energy gradually leaks to the other dimensions of the system, and instability will drive the system away from the initial two-state integrable dynamics. This leads to the more complicated multiscale structure of the turbulent system.

#### Solutions in a 4-dimensional space

In the simplest test case, we consider the full model space with four interacting modes  $\Lambda = 4$ . The invariant subspace constitutes of the two modes from  $l = 2$  and  $2l = 4$ . Figure 4.5 shows the simulation results in this typical regime. Especially, the ratio of energy inside the invariant two-mode subspace is plotted together with the time-series of the two modes  $\hat{u}_l$  and  $\hat{u}_{2l}$ . The system starts with non-mixing integrable two-state dynamics (14) for a period of time. All the energy is contained in the two-state subspace and the solution is exactly the same as the previous two-state model cases in Figure 4.3. As time goes on, the effect of instability takes place and the perturbations among the other mode begin to grow. Then chaotic structures gradually get developed with obvious mixing features. The energy inside the invariant subspace begins to fluctuate in time, representing the rapid transfer of energy between the subspaces.

#### Solutions in the 32-state model

In this second test case, we consider the standard model setup with the truncation size  $\Lambda = 16$  using initial state in the two-mode invariant subspace. The permitted invariant subspace is consist of the two modes  $\hat{u}_6, \hat{u}_{12}$  with  $l = 6$ . Figure 4.6 shows the results from direct simulations. Consistent with the previous case, we achieve the similar observation in the model evolution in time. The solution starts with the regular periodic solution as observed in the two-mode model case for

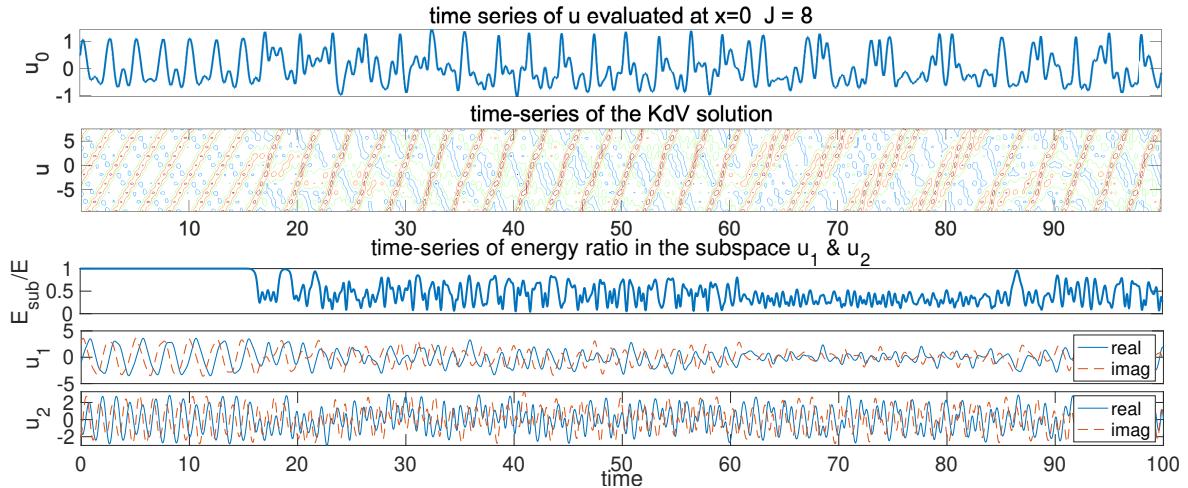


Figure 4.5: Solutions of tKdV model with truncation size  $\Lambda = 4$  starting from the two non-zero initial modes  $\hat{u}_2, \hat{u}_4$  in the highly skewed regime. The second part shows the percentage of energy in the invariant subspace from the initial state, together with the time-series of the two modes.

some period of time. Then instability takes over and turbulent dynamics are developed in time. The energy slowly leaks to the other modes to create mixing and turbulent features. The portion energy inside the two-mode invariant subspace becomes much smaller in this case due to the much higher dimensionality and nonlinear coupling between different scales.

## 5 Summarizing Discussion

This paper studies the mathematical formulation for extreme events and anomalous statistical features observed in water surface waves going across an abrupt depth change from laboratory experiments [3]. The phenomena are modeled using a statistical truncated KdV equation as a Hamiltonian system to match the incoming and outgoing state Gibbs measures before and after the abrupt depth change [12]. The invariant Gibbs measures for the statistical equilibrium solutions of the truncated model are characterized by the change of key statistics from a near-Gaussian PDF in the low inverse temperature regime to a highly skewed PDF in the high inverse temperature case. We investigate this representative statistical phase transition under the framework of low-order truncation models. A series of different model truncation sizes for the tKdV equation are compared in both single trajectory solutions and the final equilibrium distributions. The core features in the phase transition between two distinct statistical regimes stay invariant among the various experiments with reducing number of resolved modes. We finally reach the simplest model setup with only two interacting spectral modes.

Even though the model is no longer mixing with a severe two spectral mode truncation, the two-mode model enjoys many benefits from a clean integrable dynamics. The two-mode model is first analyzed in the phase space based on the explicit formulas for the conserved quantities corresponding to different values of the inverse temperature in the Gibbs measure. The large positive and negative inverse temperature regimes are separated from each other with localized amplitudes in the two modes. This leads to the highly skewed bimodal distributions in the invariant measure. The two positive and negative inverse temperature regimes are connected with another regime with a small value of the inverse temperature, where the amplitude of the two modes can vary among a wide range. This intermediate regime creates near-Gaussian statistics in contrast to the other two extreme regimes with high skewness of different directions. The results are confirmed with detailed direct numerical simulations of the truncated models. The analysis here also set a theoretical foundation for a wider variety of models in different settings showing transition to extreme events.

## Acknowledgement

This research of A. J. M. is partially supported by the Office of Naval Research N00014-19-S-B001. D. Q. is supported as a postdoctoral fellow on the grant.

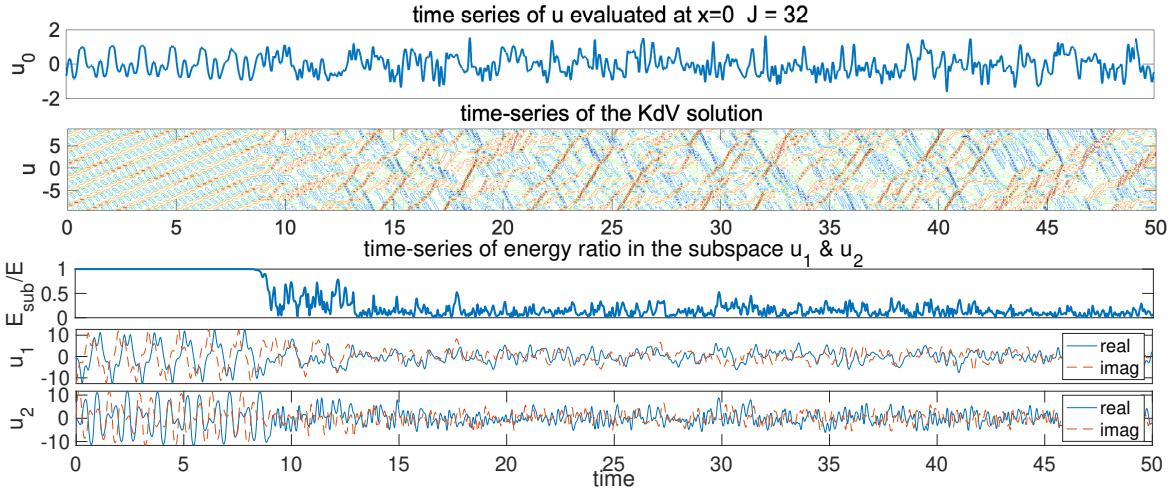


Figure 4.6: Solutions of tKdV model with truncation size  $\Lambda = 16$  starting from the two non-zero initial modes  $\hat{u}_6, \hat{u}_{12}$ . The second part shows the percentage of energy in the invariant subspace from the initial state, together with the time-series of the two modes.

## References

- [1] Rafail V Abramov, Gregor Kovačič, and Andrew J Majda. Hamiltonian structure and statistically relevant conserved quantities for the truncated Burgers-Hopf equation. *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*, 56(1):1–46, 2003.
- [2] J Bajars, JE Frank, and BJ Leimkuhler. Weakly coupled heat bath models for Gibbs-like invariant states in nonlinear wave equations. *Nonlinearity*, 26(7):1945–1973, 2013.
- [3] C Tyler Bolles, Kevin Speer, and MNJ Moore. Anomalous wave statistics induced by abrupt depth change. *Physical Review Fluids*, 4(1):011801, 2019.
- [4] Nan Chen and Andrew J Majda. Beating the curse of dimension with accurate statistics for the Fokker–Planck equation in complex turbulent systems. *Proceedings of the National Academy of Sciences*, 114(49):12864–12869, 2017.
- [5] Nan Chen and Andrew J Majda. Efficient statistically accurate algorithms for the Fokker–Planck equation in large dimensions. *Journal of Computational Physics*, 354:242–268, 2018.
- [6] R Höhmann, U Kuhl, H-J Stöckmann, L Kaplan, and EJ Heller. Freak waves in the linear regime: A microwave study. *Physical review letters*, 104(9):093901, 2010.
- [7] Robin Stanley Johnson. *A modern introduction to the mathematical theory of water waves*, volume 19. Cambridge university press, 1997.
- [8] A Majda and I Tomofeyev. Statistical mechanics for truncations of the Burgers-Hopf equation: a model for intrinsic stochastic behavior with scaling. *Milan Journal of Mathematics*, 70(1):39–96, 2002.
- [9] Andrew Majda and Xiaoming Wang. *Nonlinear dynamics and statistical theories for basic geophysical flows*. Cambridge University Press, 2006.
- [10] Andrew J Majda and Michal Branicki. Lessons in uncertainty quantification for turbulent dynamical systems. *Discrete & Continuous Dynamical Systems-A*, 32(9):3133–3221, 2012.
- [11] Andrew J Majda and Nan Chen. Model error, information barriers, state estimation and prediction in complex multiscale systems. *Entropy*, 20(9):644, 2018.
- [12] Andrew J Majda, MNJ Moore, and Di Qi. Statistical dynamical model to predict extreme events and anomalous features in shallow water waves with abrupt depth change. *Proceedings of the National Academy of Sciences*, 116(10):3982–3987, 2019.

- [13] Andrew J Majda and Xin T Tong. Intermittency in turbulent diffusion models with a mean gradient. *Nonlinearity*, 28(11):4171–4208, 2015.
- [14] Robert McLachlan. Symplectic integration of hamiltonian wave equations. *Numerische Mathematik*, 66(1):465–492, 1993.
- [15] Mustafa A Mohamad and Themistoklis P Sapsis. A sequential sampling strategy for extreme event statistics in nonlinear dynamical systems. *Proceedings of the National Academy of Sciences*, 115(44):11138–11143, 2018.
- [16] Di Qi and Andrew J Majda. Predicting fat-tailed intermittent probability distributions in passive scalar turbulence with imperfect models through empirical information theory. *Communications in Mathematical Sciences*, 14(6):1687–1722, 2016.
- [17] Di Qi and Andrew J Majda. Predicting extreme events for passive scalar turbulence in two-layer baroclinic flows through reduced-order stochastic models. *Communications in Mathematical Sciences*, 16(1):17–51, 2018.
- [18] A Sergeeva, E Pelinovsky, and T Talipova. Nonlinear random wave field in shallow water: variable Korteweg-de Vries framework. *Natural Hazards and Earth System Sciences*, 11(2):323–330, 2011.
- [19] DR Solli, C Ropers, P Koonath, and B Jalali. Optical rogue waves. *Nature*, 450(7172):1054, 2007.
- [20] Karsten Trulsen, Huiming Zeng, and Odin Gramstad. Laboratory evidence of freak waves provoked by non-uniform bathymetry. *Physics of Fluids*, 24(9):097101, 2012.
- [21] Claudio Viotti and Frédéric Dias. Extreme waves induced by strong depth transitions: Fully nonlinear results. *Physics of Fluids*, 26(5):051705, 2014.