Version of 21 February 2012

Metastable energy strata in weakly nonlinear wave equations

Ludwig Gauckler[1,3], **Ernst Hairer**[2], **Christian Lubich**[1], **Daniel Weiss**[1]

[1] Mathematisches Institut, Universität Tübingen, Auf der Morgenstelle, D-72076 Tübingen, Germany. E-mail: {Gauckler,Lubich,Weiss}@na.uni-tuebingen.de
[2] Section de mathématiques, 2-4 rue du Lièvre, Université de Genève, CH-1211 Genève 4, Switzerland. E-mail: Ernst.Hairer@unige.ch
[3] present address: Institut für Mathematik, Technische Universität Berlin, Straße des 17. Juni 136, D-10623 Berlin, Germany.

Abstract: We consider the problem of the long-time stability of plane waves under nonlinear perturbations of linear Klein-Gordon equations. This problem reduces to studying the distribution of the mode energies along solutions of one-dimensional semilinear Klein–Gordon equations with periodic boundary conditions when the initial data are small and concentrated in one Fourier mode. It is shown that for all except finitely many values of the mass parameter, the energy remains essentially localized in the initial Fourier mode over time scales that are much longer than predicted by standard perturbation theory. The mode energies decay geometrically with the mode number with a rate that is proportional to the total energy. The result is proved using modulated Fourier expansions in time.

1. Introduction

Consider first the d-dimensional linear wave equation or Klein–Gordon equation $u_{tt} - \Delta u + \rho u = 0$ ($x \in \mathbb{R}^d$, $t \in \mathbb{R}$; with $\rho \geq 0$) with real initial data that are linear combinations of $\mathrm{e}^{\pm \mathrm{i} k\cdot x}$ for some wave vector $k \in \mathbb{R}^d, k \neq 0$. The solution is a linear combination of plane waves $\mathrm{e}^{\mathrm{i}(\pm k\cdot x \pm \omega t)}$ with all combinations of signs, with the frequency $\omega = \sqrt{|k|^2+\rho}$. This is evidently no longer the case when a small nonlinearity $g(u)$ (such as u^2 or u^3 for small initial data) is introduced on the right-hand side of the equation. The question then arises as to what is the long-time behaviour of solutions of the nonlinearly perturbed equation for such initial data. At any time $t > 0$ where a solution exists, it will have a Fourier series representation $u(x,t) = \sum_{j=-\infty}^{\infty} u_j(t)\, \mathrm{e}^{\mathrm{i} j k\cdot x}$. What can be said of the mode amplitudes $|u_j(t)|$, or better of the mode energies $E_j(t) = \frac{1}{2}|\omega_j u_j(t)|^2 + \frac{1}{2}|\dot u_j(t)|^2$ (with the frequencies $\omega_j = \sqrt{j^2|k|^2+\rho}$), for large t? Will the mode energies for $|j| \neq 1$, which vanish initially, remain small and decay geometrically with growing $|j|$ over long times? This stability problem of plane waves under nonlinear perturbations of the equation will be studied in the present paper, with a positive answer for almost all $\rho > 0$ over time scales that are much longer than standard perturbation theory would suggest.

Without loss of generality, we may assume that the wave vector has unit length (after changing ρ to $\rho|k|^2$) and, by rotational invariance, equals the first unit vector $k = (1, 0, \dots, 0)$. The problem then reduces to studying the one-dimensional semilinear wave equation

$$u_{tt} - u_{xx} + \rho u = g(u), \quad -\pi \le x \le \pi, \ t \ge 0, \tag{1}$$

with periodic boundary conditions, for small (real) initial data that are linear combinations of $\mathrm{e}^{\pm \mathrm{i}x}$. We assume that the nonlinearity g is at least quadratic at 0:

$$g \text{ is real-analytic near 0 and } g(0) = g'(0) = 0. \tag{2}$$

We expand the solution $u(x,t)$ into its Fourier series

$$u(x,t) = \sum_{j=-\infty}^{\infty} u_j(t)\, \mathrm{e}^{\mathrm{i}jx}$$

and consider the mode energies

$$E_j(t) = \frac{1}{2}|\omega_j u_j(t)|^2 + \frac{1}{2}|\dot{u}_j(t)|^2, \text{ with the frequencies } \omega_j = \sqrt{j^2 + \rho}. \tag{3}$$

Since we consider only real solutions, we have $\overline{u_j} = u_{-j}$ and therefore $E_{-j} = E_j$. We consider small initial data concentrated in the first mode:

$$\begin{aligned} &E_1(0) \le \epsilon, \qquad 0 < \epsilon \ll 1, \\ &E_j(0) = 0 \quad \text{for } |j| \ne 1. \end{aligned} \tag{4}$$

Our main result, Theorem 1, shows that for all mass parameters $\rho > 0$, with the exception of only finitely many in any bounded interval, and for sufficiently small ϵ, the energy remains essentially localized in the first mode over long times $t \le \epsilon^{-N}$, for arbitrarily fixed $N > 1$. The mode energies decay geometrically with growing j, with a decay rate that is proportional to ϵ for the first modes, and at least with a smaller power of ϵ close to 1 for all remaining modes.

We are not aware that the problem studied in this paper has been considered previously in the literature. Our result fits, however, into a series of results that have been obtained recently on the long-time behaviour of weakly nonlinear Hamiltonian partial differential equations by various authors [1,2,4–10,13–15]. The problem considered here is also closely related to the Fermi–Pasta–Ulam (FPU) problem [11,12], which in the small-energy regime deals with a nonlinearly perturbed system of near-resonant harmonic oscillators and for which metastability phenomena have been analyzed in [3,18].

The long-time behaviour of such weakly nonlinear problems has been analyzed rigorously by two different techniques: Birkhoff normal forms and modulated Fourier expansions in time. Here we follow the latter approach, which originally came up in the long-time analysis of numerical methods for oscillatory differential equations [16,17] and was later applied to Hamiltonian partial differential equations in [8,13,14]. It does not seem directly possible to obtain the results of the present paper using normal form techniques. While normal forms use coordinate transforms to take the system into a form from which the desired long-time properties can be read off, modulated Fourier expansions embed the given system into a larger modulation system that has a Hamiltonian structure from which the long-time behaviour can be inferred.

Our approach is close to that taken for the FPU problem in [18], but the problem considered here does not encounter the technical difficulties due to almost-resonances present in the FPU problem. On the other hand, the non-resonant situation considered here allows us to obtain stronger estimates over much longer time scales.

2. Main result

In this paper we prove the following theorem.

Theorem 1. *Fix an integer $K \geq 2$ and real numbers $s > \frac{1}{2}$ and $\rho_0 > 0$. For all except finitely many $0 < \rho < \rho_0$ the following holds: There exist $\delta_0 > 0$ and positive c and C such that for $0 < \theta \leq 1$ the mode energies (3) of solutions to the nonlinear wave equation (1)-(2) for initial data (4) with $0 < \epsilon^{\theta/2} \leq \delta_0$ satisfy, over long times*

$$0 \leq t \leq c\,\epsilon^{-\theta K/2},$$

the bounds

$$\begin{aligned} E_0(t) &\leq C\epsilon^2, \\ E_j(t) &\leq C\epsilon^j, \quad 0 < j < K, \\ \sum_{j=K}^{\infty} \epsilon^{-(j-K)(1-\theta)} j^{2s} E_j(t) &\leq C\epsilon^K \end{aligned} \tag{5}$$

and $|E_1(t) - E_1(0)| \leq C\epsilon^2$.

The values of ρ considered in the theorem are those for which the frequencies $\omega_j = \sqrt{j^2 + \rho}$ satisfy a certain non-resonance condition, which will be specified below. We note that for $\rho = 0$ the frequencies are fully resonant, and the theorem does not apply in this case. One encounters blow-up phenomena for $\rho = 0$, for instance for the nonlinearities $g(u) = u^2$ or $g(u) = u^3$.

It is not difficult to extend the proof of Theorem 1 to initial data concentrated in a collection of low modes, $E_j(0) \leq \epsilon$ for $|j| \leq B$ and $E_j(0) = 0$ for $|j| > B$. For example for $B = 1$ the estimate (5) still holds if the first estimate is replaced by $E_0(t) \leq C\epsilon$. For larger values of B the range of modes of energy ϵ^j is stretched in (5),

$$E_l(t) \leq C\epsilon^j \quad \text{for} \quad 0 < j < K \quad \text{and} \quad (j-1)B \leq l \leq jB$$

and a corresponding modification of the last estimate in (5).

The proof of Theorem 1 also shows that the statement of the theorem still holds true for more general initial data that satisfy the bounds (5) with a given constant C_0 in place of C.

Stronger estimates can be obtained when further derivatives of g at 0 are zero. In particular, for odd functions g, we have $E_j(t) = 0$ for all t for all even j. For odd functions g we obtain the above estimates for odd j if $g'''(0) \neq 0$, e.g., for $g(u) = -u^3$ or $g(u) = \rho(\sin u - u)$ (the sine-Gordon equation), and slightly stronger estimates if $g'''(0) = 0$.

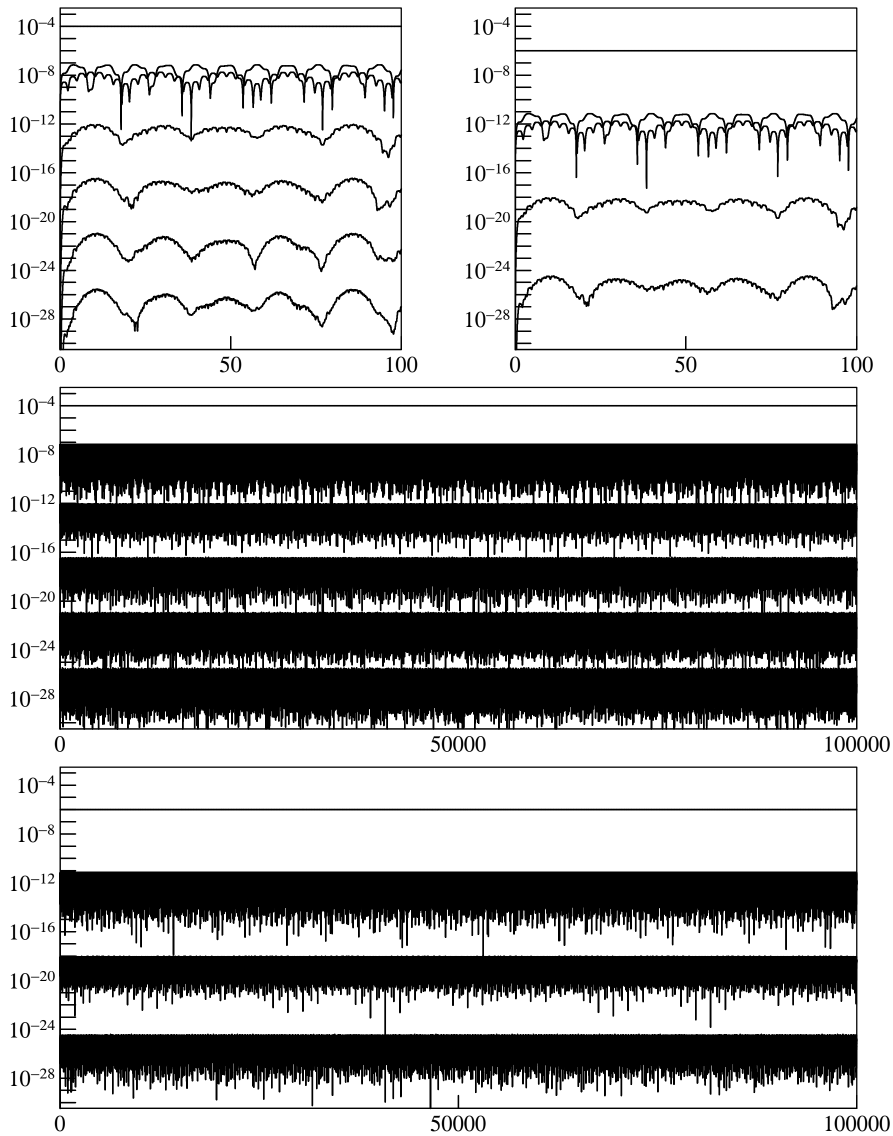

Fig. 1. Mode energies versus time.

We give a numerical illustration of the above result for $g(u) = u^2$, $\rho = 1/2$ and initial data

$$u(x,0) = \frac{2\sqrt{2\epsilon}}{\omega_1}\cos(x), \quad u_t(x,0) = 0,$$

so that

$$E_1(0) = \epsilon, \quad E_j(0) = 0 \quad \text{for } j \neq 1.$$

In Fig. 1, the mode energies $E_j(t)$ for $\epsilon = 10^{-4}$ (first and third figure) and $\epsilon = 10^{-6}$ (second and last figure) are plotted versus time on the intervals $[0, 100]$

(first two figures) and $[0, 10^5]$ (lower two figures). Lower lying curves correspond to higher modes j. The differential equation was discretized in space by a spectral collocation method with 2^7 grid points and in time by a trigonometric integrator [17, Ch. XIII] with step size 10^{-2}. Nearly identical results were obtained with other choices of the discretization parameters.

The starting point for the proof of Theorem 1 is to rewrite the partial differential equation (1) as a system of second-order differential equations for the Fourier coefficients $u_j(t)$:

$$\ddot{u}_j + \omega_j^2 u_j = -\nabla_{-j} U(\mathbf{u}) \tag{6}$$

with the potential

$$U(\mathbf{u}) = -\sum_{m\geq 3} \frac{g^{(m-1)}(0)}{m!} \sum_{j_1+\ldots+j_m=0} u_{j_1} \ldots u_{j_m}.$$

Here, ∇_{-j} denotes the partial derivative with respect to u_{-j} and $\mathbf{u} = (u_j)_{j\in\mathbb{Z}}$. Since $u(x,t)$ is real, we have $\overline{u_j} = u_{-j}$. In particular, u_0 is real. The proof of Theorem 1 via modulated Fourier expansions in time is outlined in Section 3. The proof proceeds by a sequence of auxiliary results of independent interest, which are proved in Sections 4 and 5.

3. Modulated Fourier expansion

3.1. Approximation ansatz. We will approximate the jth Fourier component of the solution by a sum of products of polynomials and exponentials,

$$u_j(t) \approx \sum_{\mathbf{k}\in\mathcal{K}_j} z_j^{\mathbf{k}}(t)\, \mathrm{e}^{\mathrm{i}(\mathbf{k}\cdot\boldsymbol{\omega})t}, \tag{7}$$

where $\mathcal{K}_j$ is a finite set of sequences $\mathbf{k} = (k_l)_{l\geq 0}$ with only finitely many nonzero integers k_l, where $\boldsymbol{\omega} = (\omega_l)_{l\geq 0}$ is the sequence of frequencies (3), and $\mathbf{k}\cdot\boldsymbol{\omega} = \sum_{l\geq 0} k_l\omega_l$. A subtlety that sets the present paper apart from previous work on modulated Fourier expansions, e.g., in [8], is the fact that here the expansion is not a multiscale expansion, but instead we use just one time-scale and ensure the uniqueness of the modulation functions $z_j^{\mathbf{k}}$ by requiring that they are *polynomials.* They will be constructed by inserting the ansatz (7) into (6), collecting the coefficients of $\mathrm{e}^{\mathrm{i}(\mathbf{k}\cdot\boldsymbol{\omega})t}$ for all $\mathbf{k}$ in the resulting expressions, and to determine polynomials $z_j^{\mathbf{k}}$ such that there is a small defect $d_j^{\mathbf{k}}$ in the modulation equations

$$(\omega_j^2 - (\mathbf{k}\cdot\boldsymbol{\omega})^2) z_j^{\mathbf{k}} + 2\mathrm{i}(\mathbf{k}\cdot\boldsymbol{\omega})\dot{z}_j^{\mathbf{k}} + \ddot{z}_j^{\mathbf{k}} = -\nabla_{-j}^{-\mathbf{k}}\mathcal{U}(\mathbf{z}) + d_j^{\mathbf{k}} \tag{8}$$

with the potential

$$\mathcal{U}(\mathbf{z}) = -\sum_{m\geq 3} \frac{g^{(m-1)}(0)}{m!} \sum_{j_1+\ldots+j_m=0}\ \sum_{\mathbf{k}^1+\ldots+\mathbf{k}^m=0} z_{j_1}^{\mathbf{k}^1} \ldots z_{j_m}^{\mathbf{k}^m}.$$

Here, the last sum is restricted to $\mathbf{k}^1 \in \mathcal{K}_{j_1}, \ldots, \mathbf{k}^m \in \mathcal{K}_{j_m}$. The symbol $\nabla_{-j}^{-\mathbf{k}}$ in (8) denotes the partial derivative with respect to $z_{-j}^{-\mathbf{k}}$, and $\mathbf{z} = (z_j^{\mathbf{k}})_{j,\mathbf{k}}$.

3.2. Non-resonance condition. In the construction of the modulated Fourier expansion (7) via (8) we encounter small denominators $\omega_j^2 - (\mathbf{k}\cdot\boldsymbol{\omega})^2$. We require the following non-resonance condition: there is a $\gamma > 0$ such that

$$\big|\omega_j - |\mathbf{k}\cdot\boldsymbol{\omega}|\big| \ge \gamma \quad \text{for} \quad (j,\mathbf{k}) \in \mathcal{K},\ \mathbf{k} \ne \pm\langle j\rangle, \tag{9}$$

where $\langle j\rangle = (0,\dots,0,1,0,\dots)$ is the $|j|$-th unit sequence (with the entry 1 at the $|j|$-th position) and the set $\mathcal{K}$ is defined in the following. We fix the integer K, and we denote

$$\mu(\mathbf{k}) = 2|k_0| + \sum_{l\ge 1} \min(l,K)|k_l|. \tag{10}$$

We set

$$\begin{aligned}\mathcal{K} = \{(j,\mathbf{k}) \,:\, &\max\big(|j|,\mu(\mathbf{k})\big) < 2K \ \text{ and } \ k_l = 0 \text{ for all } l \ge K\}\\ &\cup \{(j, \pm\langle j-r\rangle + \mathbf{k}) \,:\, |j| \ge K,\, |r| < K,\, \mu(\mathbf{k}) < K\}\end{aligned} \tag{11}$$

and let

$$\mathcal{K}_j = \{\mathbf{k} \,:\, (j,\mathbf{k}) \in \mathcal{K}\}. \tag{12}$$

The set of indices $\mathcal{K}$, which determines the linear combinations of frequencies that have to be controlled in the non-resonance condition (9), is smaller than the corresponding set for initial values with *all* mode energies of order ϵ as studied in [1,2,8]. This reflects the fact that, due to the special form of the initial condition, fewer interactions of Fourier modes are relevant on the considered time interval. The following theorem will be shown in Section 6.

Theorem 2. *Fix $K \ge 2$ and $\rho_0 > 0$. All except finitely many $0 < \rho < \rho_0$ satisfy the non-resonance condition* (9) *with a constant γ depending on ρ and K.*

This theorem is in contrast to the situation studied in [1,2,8], where the set of resonant ρ is (only) shown to be of zero Lebesgue measure [1, Theorem 6.5].

3.3. Weighted norms. For a sequence $\mathbf{u} = (u_j)_{j\in\mathbb{Z}}$ we consider the weighted ℓ^2 norm

$$\|\mathbf{u}\|^2 = \sum_{j\in\mathbb{Z}} \sigma_j |u_j|^2 \quad \text{with} \quad \sigma_j = \begin{cases} \epsilon^{-2(1-\theta)}, & j = 0,\\ \epsilon^{-|j|(1-\theta)}|j|^{2s}, & j \ne 0,\end{cases} \tag{13}$$

where $0 < \theta \le 1$. For $s > \frac{1}{2}$ this norm behaves well with convolutions $(\mathbf{u}*\mathbf{v})_j = \sum_{j_1+j_2=j} u_{j_1} v_{j_2}$. We then have

$$\|\mathbf{u}*\mathbf{v}\| \le c_s\, \|\mathbf{u}\|\cdot\|\mathbf{v}\| \tag{14}$$

with a constant c_s that is independent of $\mathbf{u}$, $\mathbf{v}$, ϵ (but depends on $s > 1/2$). This bound relies on the fact that the weights $\kappa_j = \max\big(1,|j|^{2s}\big)$ satisfy the inequality $\sum_{j_1+j_2=j} \kappa_{j_1}^{-1}\kappa_{j_2}^{-1} \le c_s^2 \kappa_j^{-1}$, so that the Cauchy–Schwarz inequality yields

$$\sum_{j\in\mathbb{Z}} \kappa_j \Big| \sum_{j_1+j_2=j} u_{j_1} v_{j_2}\Big|^2 \le c_s^2 \Big(\sum_{j_1\in\mathbb{Z}} \kappa_{j_1}|u_{j_1}|^2\Big)\Big(\sum_{j_2\in\mathbb{Z}} \kappa_{j_2}|v_{j_2}|^2\Big). \tag{15}$$

For the extended norm

$$\|(\mathbf{u}, \dot{\mathbf{u}})\|^2 = \|\boldsymbol{\Omega}\mathbf{u}\|^2 + \|\dot{\mathbf{u}}\|^2 \quad \text{with} \quad (\boldsymbol{\Omega}\mathbf{u})_j = \omega_j u_j, \tag{16}$$

we note

$$\frac{1}{2}\,\|(\mathbf{u}, \dot{\mathbf{u}})\|^2 = \sum_{j\in\mathbb{Z}} \sigma_j E_j.$$

3.4. The modulated Fourier expansion on a short time interval. The situation (4) is not met at any later time $t > 0$. Instead of (4) we now consider the situation where the initial mode energies satisfy

$$\begin{aligned} E_0(0) &\le C_0 \epsilon^2, \\ E_l(0) &\le C_0 \epsilon^l, \quad 0 < l < K, \\ \sum_{l\ge K} \sigma_l E_l(0) &\le C_0 \epsilon^{\theta K}. \end{aligned} \tag{17}$$

We then have the following result.

Theorem 3 (Modulated Fourier expansion). *Fix an integer $K \ge 2$, and let $\rho > 0$ be such that the non-resonance condition (9) is satisfied. Let $s > 1/2$ and $0 < \theta \le 1$. Let the initial mode energies satisfy (17). Then, the Fourier coefficients $u_j(t)$ ($j \in \mathbb{Z}$) of the solution of (1)-(2) admit an expansion*

$$u_j(t) = \sum_{\mathbf{k}\in\mathcal{K}_j} z_j^{\mathbf{k}}(t)\, \mathrm{e}^{\mathrm{i}(\mathbf{k}\cdot\boldsymbol{\omega})t} + r_j(t) \quad \text{for} \quad 0 \le t \le 1, \tag{18}$$

where $\mathcal{K}_j$ is a finite set of sequences $\mathbf{k} = (k_l)_{l\ge 0}$ with only finitely many non-zero integers k_l as given by (12), where the coefficient functions $z_j^{\mathbf{k}}$ are polynomials satisfying $z_{-j}^{-\mathbf{k}} = \overline{z_j^{\mathbf{k}}}$, and where the remainder $\mathbf{r}(t) = (r_j(t))_{j\in\mathbb{Z}}$ is bounded in the weighted norm (13), (16) by

$$\|(\mathbf{r}(t), \dot{\mathbf{r}}(t))\| \le C\epsilon^{\theta K} t \tag{19}$$

and the defect $\mathbf{d}(t) = (\sum_{\mathbf{k}\in\mathcal{K}_j} |d_j^{\mathbf{k}}(t)|)_{j\in\mathbb{Z}}$ of (8) is bounded by

$$\|\mathbf{d}(t)\| \le C\epsilon^{\theta K}. \tag{20}$$

The constant C is independent of ϵ and θ, but depends on K, on γ of (9), on s in (13) and on C_0 of (17).

3.5. Almost-invariants. For real τ and real sequences $\boldsymbol{\lambda} = (\lambda_l)_{l\ge 0}$ we note that

$$\mathcal{U}(\mathcal{S}_{\boldsymbol{\lambda}}(\tau)\mathbf{z}) = \mathcal{U}(\mathbf{z}) \quad \text{for} \quad \mathcal{S}_{\boldsymbol{\lambda}}(\tau)\mathbf{z} = (\mathrm{e}^{\mathrm{i}(\mathbf{k}\cdot\boldsymbol{\lambda})\tau} z_j^{\mathbf{k}})_{j,\mathbf{k}}.$$

We therefore obtain

$$0 \;=\; \frac{d}{d\tau}\Big|_{\tau=0} \mathcal{U}(\mathcal{S}_{\boldsymbol{\lambda}}(\tau)\mathbf{z}) = \sum_{j\in\mathbb{Z}} \sum_{\mathbf{k}\in\mathcal{K}_j} \mathrm{i}(\mathbf{k}\cdot\boldsymbol{\lambda}) z_{-j}^{-\mathbf{k}} \nabla_{-j}^{-\mathbf{k}} \mathcal{U}(\mathbf{z})$$

and by (8),

$$0 = \sum_{j\in\mathbb{Z}} \sum_{\mathbf{k}\in\mathcal{K}_j} \mathrm{i}(\mathbf{k}\cdot\boldsymbol{\lambda}) z_{-j}^{-\mathbf{k}} \Big((\omega_j^2 - (\mathbf{k}\cdot\boldsymbol{\omega})^2) z_j^{\mathbf{k}} + 2\mathrm{i}(\mathbf{k}\cdot\boldsymbol{\omega})\dot z_j^{\mathbf{k}} + \ddot z_j^{\mathbf{k}} - d_j^{\mathbf{k}} \Big). \qquad (21)$$

We note that

$$\begin{aligned} \sum_{j,\mathbf{k}} (\mathbf{k}\cdot\boldsymbol{\lambda}) z_{-j}^{-\mathbf{k}} (\omega_j^2 - (\mathbf{k}\cdot\boldsymbol{\omega})^2) z_j^{\mathbf{k}} &= 0, \\ 2\sum_{j,\mathbf{k}} (\mathbf{k}\cdot\boldsymbol{\lambda})(\mathbf{k}\cdot\boldsymbol{\omega}) z_{-j}^{-\mathbf{k}} \dot z_j^{\mathbf{k}} &= \sum_{j,\mathbf{k}} (\mathbf{k}\cdot\boldsymbol{\lambda})(\mathbf{k}\cdot\boldsymbol{\omega}) \frac{d}{dt}\big(z_{-j}^{-\mathbf{k}} z_j^{\mathbf{k}} \big), \\ \sum_{j,\mathbf{k}} (\mathbf{k}\cdot\boldsymbol{\lambda}) z_{-j}^{-\mathbf{k}} \ddot z_j^{\mathbf{k}} &= \sum_{j,\mathbf{k}} (\mathbf{k}\cdot\boldsymbol{\lambda}) \frac{d}{dt}\big(z_{-j}^{-\mathbf{k}} \dot z_j^{\mathbf{k}} \big), \end{aligned}$$

and hence the relation (21) can be rewritten as

$$\frac{d}{dt}\mathcal{I}_{\boldsymbol{\lambda}}(\mathbf{z},\dot{\mathbf{z}}) = -\sum_{j\in\mathbb{Z}} \sum_{\mathbf{k}\in\mathcal{K}_j} \mathrm{i}(\mathbf{k}\cdot\boldsymbol{\lambda}) z_{-j}^{-\mathbf{k}} d_j^{\mathbf{k}} \qquad (22)$$

with the almost-invariant action

$$\mathcal{I}_{\boldsymbol{\lambda}}(\mathbf{z},\dot{\mathbf{z}}) = \sum_{j\in\mathbb{Z}} \sum_{\mathbf{k}\in\mathcal{K}_j} (\mathbf{k}\cdot\boldsymbol{\lambda}) \Big((\mathbf{k}\cdot\boldsymbol{\omega}) z_{-j}^{-\mathbf{k}} z_j^{\mathbf{k}} - \mathrm{i} z_{-j}^{-\mathbf{k}} \dot z_j^{\mathbf{k}} \Big).$$

For $\boldsymbol{\lambda} = \langle l\rangle$ with $l \ge 0$, the l-th unit sequence $(0,\dots,0,1,0,\dots)$, we define the l-th almost-invariant energy as

$$\mathcal{E}_l(\mathbf{z},\dot{\mathbf{z}}) = \frac{\omega_l}{2}\, \mathcal{I}_{\langle l\rangle}(\mathbf{z},\dot{\mathbf{z}}) = \frac{1}{2} \sum_{j\in\mathbb{Z}} \sum_{\mathbf{k}\in\mathcal{K}_j} k_l\omega_l \Big((\mathbf{k}\cdot\boldsymbol{\omega}) |z_j^{\mathbf{k}}|^2 - \mathrm{i} z_{-j}^{-\mathbf{k}} \dot z_j^{\mathbf{k}} \Big) \qquad (23)$$

and we write briefly

$$\mathcal{E}_l(t) = \mathcal{E}_l(\mathbf{z}(t),\dot{\mathbf{z}}(t))$$

when it is clear to which function $\mathbf{z}$ we refer. We first show that the almost-invariant energies $\mathcal{E}_l$ can be bounded in terms of bounds on the mode energies E_j.

Theorem 4 (Almost-invariant energies controlled by mode energies). *Let the conditions of Theorem 3 be fulfilled. If the initial mode energies satisfy the bounds (17), then we have, for* $0 \le t \le 1$,

$$\begin{aligned} |\mathcal{E}_0(t)| &\le \mathcal{C}_0 \epsilon^2, \\ |\mathcal{E}_l(t)| &\le \mathcal{C}_0 \epsilon^l, \quad 0 < l < K, \\ \sum_{l \ge K} \sigma_l |\mathcal{E}_l(t)| &\le \mathcal{C}_0 \epsilon^{\theta K}, \end{aligned} \tag{24}$$

where $\mathcal{C}_0$ *is independent of* ϵ *and depends only through the constant* C_0 *on the estimates (17).*

The relation (22) together with bounds of the defects $d_j^{\mathbf{k}}$ leads to the following bounds, which show that the functions $\mathcal{E}_l(t)$ are nearly constant.

Theorem 5 (Variation of almost-invariant energies). *Under the conditions of Theorem 3 we have, for* $0 \le t \le 1$,

$$\begin{aligned} \left| \frac{d}{dt} \mathcal{E}_0(t) \right| &\le C \epsilon^2 \epsilon^{\theta K/2}, \\ \left| \frac{d}{dt} \mathcal{E}_l(t) \right| &\le C \epsilon^l \epsilon^{\theta K/2}, \quad 0 < l < K, \\ \sum_{l \ge K} \sigma_l \left| \frac{d}{dt} \mathcal{E}_l(t) \right| &\le C \epsilon^{3\theta K/2}, \end{aligned}$$

where C *is independent of* ϵ *and depends on the initial data only through the constant* C_0 *of (17).*

When we consider a modulated Fourier expansion constructed from the solution u at $t = 1$, we can relate the almost-invariants of that expansion with those of the modulated Fourier expansion starting from the initial data at $t = 0$. Together with Theorem 5 this will allow us to study longer times by patching together many intervals.

Theorem 6 (Transitions in the almost-invariant energies). *Let the conditions of Theorem 3 be fulfilled. Let* $\mathbf{z}(t) = (z_j^{\mathbf{k}}(t))_{j,\mathbf{k}}$ *for* $0 \le t \le 1$ *be the coefficient functions as in Theorem 3 for initial data* $(u(\cdot,0), u_t(\cdot,0))$*, and let* $\widetilde{\mathbf{z}}(t) = (\widetilde{z}_j^{\mathbf{k}}(t))_{j,\mathbf{k}}$ *be the coefficient functions of the modulated Fourier expansion for* $0 \le t \le 1$ *corresponding to the initial data* $(u(\cdot,1), u_t(\cdot,1))$*, constructed as in Theorem 3. Then,*

$$\begin{aligned} |\mathcal{E}_0(\mathbf{z}(1), \dot{\mathbf{z}}(1)) - \mathcal{E}_0(\widetilde{\mathbf{z}}(0), \dot{\widetilde{\mathbf{z}}}(0))| &\le C \epsilon^2 \epsilon^{\theta K/2}, \\ |\mathcal{E}_l(\mathbf{z}(1), \dot{\mathbf{z}}(1)) - \mathcal{E}_l(\widetilde{\mathbf{z}}(0), \dot{\widetilde{\mathbf{z}}}(0))| &\le C \epsilon^l \epsilon^{\theta K/2}, \quad 0 < l < K, \\ \sum_{l \ge K} \sigma_l |\mathcal{E}_l(\mathbf{z}(1), \dot{\mathbf{z}}(1)) - \mathcal{E}_l(\widetilde{\mathbf{z}}(0), \dot{\widetilde{\mathbf{z}}}(0))| &\le C \epsilon^{3\theta K/2}, \end{aligned}$$

where C *is independent of* ϵ *and depends on the initial data only through the constant* C_0 *of (17).*

As the following result shows, the mode energies E_j can be bounded in terms of bounds of the almost-invariant energies $\mathcal{E}_l$.

Theorem 7 (Mode energies controlled by almost-invariant energies). *Let the conditions of Theorem 3 be fulfilled. If the almost-invariant energies satisfy (24) for* $0 \le t \le 1$*, then the mode energies are bounded by*

$$\begin{aligned} E_0(t) &\le \mathcal{C}\epsilon^2, \\ E_l(t) &\le \mathcal{C}\epsilon^l, \quad 0 < l < K, \\ \sum_{l \ge K} \sigma_l E_l(t) &\le \mathcal{C}\epsilon^{\theta K}, \end{aligned} \tag{25}$$

and

$$|E_1(t) - \mathcal{E}_1(t)| \le \mathcal{C}\epsilon^2,$$

where $\mathcal{C}$ *depends on* $\mathcal{C}_0$ *in (24), but is independent of* ϵ *and* C_0 *of (17) if* ϵ^θ *is sufficiently small.*

3.6. From short to long intervals. Assume that the solution of (1) satisfies condition (25) with a possibly larger constant $\widehat{C}_0$ instead of $\mathcal{C}$ for $0 \le t \le n$ (this is true for $n = 0$ by assumption (17); else it is justified by induction). Theorems 3 to 7 can be applied on the next interval $n \le t \le n+1$. In particular, Theorem 3 gives us a modulated Fourier expansion whose coefficient functions we denote by $\mathbf{z}_n(t)$, for $0 \le t \le 1$. Theorem 5 yields, for $0 \le t \le 1$,

$$\begin{aligned} |\mathcal{E}_0(\mathbf{z}_n(0), \dot{\mathbf{z}}_0(0)) - \mathcal{E}_0(\mathbf{z}_n(t), \dot{\mathbf{z}}_n(t))| &\le \widehat{C}\epsilon^2\epsilon^{\theta K/2}\,, \\ |\mathcal{E}_l(\mathbf{z}_n(0), \dot{\mathbf{z}}_0(0)) - \mathcal{E}_l(\mathbf{z}_n(t), \dot{\mathbf{z}}_n(t))| &\le \widehat{C}\epsilon^l\epsilon^{\theta K/2}\,, \quad 0 < l < K, \\ \sum_{l \ge K} \sigma_l \,|\mathcal{E}_l(\mathbf{z}_n(0), \dot{\mathbf{z}}_0(0)) - \mathcal{E}_l(\mathbf{z}_n(t), \dot{\mathbf{z}}_n(t))| &\le \widehat{C}\epsilon^{3\theta K/2} \end{aligned}$$

with a constant $\widehat{C}$ that only depends on $\widehat{C}_0$. On the other hand, we can also apply Theorems 4 and 7 to guarantee that condition (25) is also satisfied on the interval $n \le t \le n+1$ with a possibly larger constant. Consequently, Theorem 3 allows us to consider the modulated Fourier expansion $\mathbf{z}_{n+1}(t)$ on the next interval. The transition is estimated by Theorem 6 as

$$\begin{aligned} |\mathcal{E}_0(\mathbf{z}_n(1), \dot{\mathbf{z}}_n(1)) - \mathcal{E}_0(\mathbf{z}_{n+1}(0), \dot{\mathbf{z}}_{n+1}(0))| &\le \widehat{C}\epsilon^2\epsilon^{\theta K/2}, \\ |\mathcal{E}_l(\mathbf{z}_n(1), \dot{\mathbf{z}}_n(1)) - \mathcal{E}_l(\mathbf{z}_{n+1}(0), \dot{\mathbf{z}}_{n+1}(0))| &\le \widehat{C}\epsilon^l\epsilon^{\theta K/2}, \quad 0 < l < K, \\ \sum_{l \ge K} \sigma_l \,|\mathcal{E}_l(\mathbf{z}_n(1), \dot{\mathbf{z}}_n(1)) - \mathcal{E}_l(\mathbf{z}_{n+1}(0), \dot{\mathbf{z}}_{n+1}(0))| &\le \widehat{C}\epsilon^{3\theta K/2}\,, \end{aligned}$$

where the constant $\widehat{C}$ can be assumed to be the same as above. Summing up these estimates over n and applying the triangle inequality yields, for $0 \le t \le 1$,

$$\begin{aligned} |\mathcal{E}_0(\mathbf{z}_0(0), \dot{\mathbf{z}}_0(0)) - \mathcal{E}_0(\mathbf{z}_n(t), \dot{\mathbf{z}}_n(t))| &\le \widehat{C}\epsilon^2\epsilon^{\theta K/2}\,(2n+1), \\ |\mathcal{E}_l(\mathbf{z}_0(0), \dot{\mathbf{z}}_0(0)) - \mathcal{E}_l(\mathbf{z}_n(t), \dot{\mathbf{z}}_n(t))| &\le \widehat{C}\epsilon^l\epsilon^{\theta K/2}\,(2n+1), \quad 0 < l < K, \\ \sum_{l \ge K} \sigma_l \,|\mathcal{E}_l(\mathbf{z}_0(0), \dot{\mathbf{z}}_0(0)) - \mathcal{E}_l(\mathbf{z}_n(t), \dot{\mathbf{z}}_n(t))| &\le \widehat{C}\epsilon^{3\theta K/2}\,(2n+1). \end{aligned}$$

Let $\mathcal{C}_0$ be the constant in condition (24) for the first subinterval $[0,1]$. This constant depends on C_0 of (17), but not on $\widehat{C}_0$. Set $c = \mathcal{C}_0/(2\widehat{C})$. For $n+1 \le c\,\epsilon^{-\theta K/2}$ we thus obtain from Theorem 4

$$\begin{aligned} |\mathcal{E}_0(\mathbf{z}_n(t), \dot{\mathbf{z}}_n(t))| &\le 2\mathcal{C}_0\epsilon^2, \\ |\mathcal{E}_l(\mathbf{z}_n(t), \dot{\mathbf{z}}_n(t))| &\le 2\mathcal{C}_0\epsilon^l, \quad 0 < l < K, \\ \sum_{l \ge K} \sigma_l |\mathcal{E}_l(\mathbf{z}_n(t), \dot{\mathbf{z}}_n(t))| &\le 2\mathcal{C}_0\epsilon^{\theta K}. \end{aligned}$$

Theorem 7 now yields for $t \le c\,\epsilon^{-\theta K/2}$,

$$\begin{aligned} E_0(t) &\le C\epsilon^2, \\ E_l(t) &\le C\epsilon^l, \quad 0 < l < K, \\ \sum_{l \ge K} \sigma_l E_l(t) &\le C\epsilon^{\theta K}, \\ |E_1(t) - E_1(0)| &\le C\epsilon^2, \end{aligned}$$

where C depends on C_0, but is independent of $\widehat{C}_0$. The last estimate is obtained from the bound $|E_1(t) - \mathcal{E}_1(t)| \le C\epsilon^2$ and from the near invariance of $\mathcal{E}_1(t)$. This gives us the bounds of Theorem 1. It remains to prove Theorems 3 to 7 .

4. Proof of Theorems 3 and 4

For ease of presentation we give the proof for the particular nonlinearity $g(u) = u^2$. The generalization to other nonlinearities satisfying (2) presents no mathematical difficulties. In the following we denote by $\lesssim$ an inequality $\le$ up to a factor that is independent of ϵ and θ.

4.1. Construction of the modulation functions. We work with the small parameter

$$\delta = \epsilon^{\theta/2}.$$

Condition (17) implies for the initial values

$$\begin{aligned} \omega_0 u_0(0) &= \delta^2 a_0, \\ \dot{u}_0(0) &= \delta^2 b_0, \\ \omega_j u_j(0) &= \delta^{\min(|j|,K)} a_j, \quad j \ne 0, \\ \dot{u}_j(0) &= \delta^{\min(|j|,K)} b_j, \quad j \ne 0, \end{aligned} \tag{26}$$

where $\overline{a_j} = a_{-j}, \overline{b_j} = b_{-j}$ and by (17)

$$\sum_{j \in \mathbb{Z}} \sigma_j (|a_j|^2 + |b_j|^2) \lesssim 1.$$

The modulation functions will be constructed from an ansatz

$$z_j^{\mathbf{k}}(t) = \sum_{m=1}^{2K-1} \delta^m z_{j,m}^{\mathbf{k}}(t) \tag{27}$$

with polynomials $z_{j,m}^{\mathbf{k}}(t)$. Inserting this ansatz into (8), using $g(u) = u^2$, and comparing the coefficients of δ^m yields the equation

$$\begin{aligned}(\omega_j^2 - (\mathbf{k}\cdot\boldsymbol{\omega})^2) z_{j,m}^{\mathbf{k}} + 2\mathrm{i}(\mathbf{k}\cdot\boldsymbol{\omega})\dot z_{j,m}^{\mathbf{k}} + \ddot z_{j,m}^{\mathbf{k}} \\ = \sum_{j_1+j_2=j} \sum_{\mathbf{k}^1+\mathbf{k}^2=\mathbf{k}} \sum_{m_1+m_2=m} z_{j_1,m_1}^{\mathbf{k}^1} z_{j_2,m_2}^{\mathbf{k}^2}.\end{aligned} \tag{28}$$

For the initial values we obtain from (26) by requiring that (7) is exactly fulfilled at $t = 0$,

$$\begin{aligned}\omega_j \sum_{\mathbf{k}\in\mathcal{K}_j} z_{j,m}^{\mathbf{k}}(0) &= \begin{cases} a_j & \text{for } m = \min(|j|, K) \text{ if } j \neq 0, \\ & \text{and for } m = 2 \text{ if } j = 0, \\ 0 & \text{else,} \end{cases} \\ \sum_{\mathbf{k}\in\mathcal{K}_j} \Big(\dot z_{j,m}^{\mathbf{k}}(0) + \mathrm{i}(\mathbf{k}\cdot\boldsymbol{\omega}) z_{j,m}^{\mathbf{k}}(0) \Big) &= \begin{cases} b_j & \text{for } m = \min(|j|, K) \text{ if } j \neq 0, \\ & \text{and for } m = 2 \text{ if } j = 0, \\ 0 & \text{else.} \end{cases}\end{aligned} \tag{29}$$

We determine an approximate solution of (28)-(29) for $m = 1, \dots, 2K-1$ consecutively. For $\mathbf{k} \neq \pm\langle j\rangle$, after division by $(\omega_j^2 - (\mathbf{k}\cdot\boldsymbol{\omega})^2)$, equation (28) is of the form

$$z - \alpha \dot z - \beta \ddot z = p.$$

For a polynomial p of degree d, the unique polynomial solution is given by

$$z = \sum_{l=0}^{d} \big(\alpha \frac{d}{dt} + \beta \frac{d^2}{dt^2}\big)^l p. \tag{30}$$

For $\mathbf{k} = \pm\langle j\rangle$ we have a differential equation of the form $\dot z - \beta \ddot z = p$ which admits a polynomial solution with one free parameter. This parameter is fixed by the initial value that is obtained from (29):

$$\begin{aligned}2i\omega_j z_{j,m}^{\pm\langle j\rangle}(0) = -\mathrm{i} \sum_{\mathbf{k}\neq\langle j\rangle, \mathbf{k}\neq-\langle j\rangle} (\omega_j \pm \mathbf{k}\cdot\boldsymbol{\omega}) z_{j,m}^{\mathbf{k}}(0) \mp \sum_{\mathbf{k}} \dot z_{j,m}^{\mathbf{k}}(0) \\ + \begin{cases} \mathrm{i}a_j \pm b_j & \text{for } m = \min(|j|, K) \text{ if } j \neq 0, \\ & \text{and for } m = 2 \text{ if } j = 0, \\ 0 & \text{else.} \end{cases}\end{aligned} \tag{31}$$

We now discuss the iterative process used to determine the modulation functions $z_{j,m}^{\mathbf{k}}$ in more detail. In particular, we will show that these functions vanish for sequences $\mathbf{k}$ not belonging to the set $\mathcal{K}_j$ from (12). Moreover, the leading powers of δ in the functions $z_j^{\mathbf{k}}$ will be examined in more detail, and it will be found that

$$
\begin{aligned}
z_{j,m}^{\mathbf{k}} &= 0 \quad \text{for } m < \min(|j|,K) \text{ if } j \neq 0, \text{ and for } m < 2 \text{ if } j = 0, && \text{(32a)}\\
\dot z_{j,m}^{\mathbf{k}} &= 0 \quad \text{for } m = \min(|j|,K) \text{ if } j \neq 0, \text{ and for } m = 2 \text{ if } j = 0, && \text{(32b)}\\
z_{j,m}^{\mathbf{k}} &= 0 \quad \text{for } m < \mu(\mathbf{k}), && \text{(32c)}\\
z_{j,K}^{\mathbf{k}} &= 0 \quad \text{for } \mathbf{k} \neq \pm\langle j\rangle \text{ and } |j| > K, && \text{(32d)}\\
z_{j,m}^{\pm\langle l\rangle} &= 0 \quad \text{for } |j| \neq l \text{ and } m = \min(|j|,K) \text{ if } j \neq 0, \text{ and } m = 2 \text{ if } j = 0. && \text{(32e)}
\end{aligned}
$$

Case $m = 1$: Here $p = 0$, since the sum in (28) is empty. Hence,

$$
\begin{aligned}
&z_{j,1}^{\mathbf{k}} \equiv 0 \text{ for } \mathbf{k} \neq \pm\langle j\rangle,\\
&z_{j,1}^{\pm\langle j\rangle} \text{ is a constant.}
\end{aligned}
$$

For the initial values we obtain from (31)

$$
\begin{aligned}
z_{1,1}^{\pm\langle 1\rangle}(0) &= \frac{1}{2\omega_1}(a_1 \mp \mathrm{i}b_1),\\
z_{-1,1}^{\pm\langle 1\rangle}(0) &= \frac{1}{2\omega_1}(a_{-1} \mp \mathrm{i}b_{-1}) = \frac{1}{2\omega_1}(\bar a_1 \mp \mathrm{i}\bar b_1)
\end{aligned}
$$

for the only non-vanishing coefficients with $m = 1$. We note that $z_{-j,1}^{-\mathbf{k}} = \overline{z_{j,1}^{\mathbf{k}}}$.

Cases $1 < m < K$: We recall the definition (10) of $\mu(\mathbf{k})$. By induction, it follows that $z_{j,m}^{\mathbf{k}}$ is a polynomial of degree at most $m - \max(|j|,\mu(\mathbf{k}))$ and in the case $j\mu(\mathbf{k}) = 0$ of degree not exceeding $m - \max(|j|,\mu(\mathbf{k}),2)$, where a negative degree corresponds to the zero polynomial. In particular, we note that the sums in (28) are finite sums, and $z_{j,m}^{\mathbf{k}} \neq 0$ only if $\max(|j|,\mu(\mathbf{k})) \leq m$. Using $a_{-j} = \overline{a_j}$ and $b_{-j} = \overline{b_j}$ we obtain $z_{-j,m}^{-\mathbf{k}} = \overline{z_{j,m}^{\mathbf{k}}}$.

Cases $K \leq m \leq 2K-1$: The case $m = K$ is different for the diagonal elements $\dot z_{j,K}^{\pm\langle j\rangle}$ for $|j| > K$, since by our assumption on the initial values we can no longer conclude $z_{j,K}^{\pm\langle j\rangle} = 0$ from $\dot z_{j,K}^{\pm\langle j\rangle} = 0$. For $m = K + n$ with $n > 0$ we decompose the right-hand side of (28) as

$$
\sum_{j_1+j_2=j}\ \sum_{\mathbf{k}^1+\mathbf{k}^2=\mathbf{k}} \Big(2\sum_{l=1}^{n} z_{j_1,m-l}^{\mathbf{k}^1} z_{j_2,l}^{\mathbf{k}^2} + \sum_{\substack{m_1+m_2=m,\\ m_i<K}} z_{j_1,m_1}^{\mathbf{k}^1} z_{j_2,m_2}^{\mathbf{k}^2}\Big).
$$

The number of terms in the sum is thus finite and can be estimated independently of j and $\mathbf{k}$. The first expression in the sum makes evident how the index set for non-vanishing modulation functions,

$$
\big\{\mathbf{k} = \pm\langle j+r\rangle + \bar{\mathbf{k}} \,:\, |j| > m = K+n, |r| \leq n, \mu(\bar{\mathbf{k}}) \leq n\big\},
$$

is built up: from $\mathbf{k}^1 = \pm\langle j_1 + r_1\rangle + \bar{\mathbf{k}}^1$ with $|r_1| \leq n - l, \mu(\mathbf{k}^1) \leq n - l$ and $\max(|j_2|,\mu(\mathbf{k}^2)) \leq l$ it follows that $\mathbf{k} = \mathbf{k}^1 + \mathbf{k}^2$ equals $\mathbf{k} = \pm\langle j+r\rangle + \bar{\mathbf{k}}$ with

$r = r_1 - j_2$ and $\bar{\mathbf{k}} = \bar{\mathbf{k}}^1 + \mathbf{k}^2$, which are bounded by $|r| \le n, \mu(\bar{\mathbf{k}}) \le n$. Induction over n further shows that for $(j,\mathbf{k})$ with $\max(|j|,\mu(\mathbf{k})) \le K+n$, the polynomial $z_{j,m}^{\mathbf{k}}$ has degree at most $m - \max(|j|,\mu(\mathbf{k}))$ and in the case $j\mu(\mathbf{k}) = 0$ even $\le m - \max(|j|,\mu(\mathbf{k}),2)$, and $z_{j,m}^{\mathbf{k}}$ with $\max(|j|,\mu(\mathbf{k})) > K+n$ is a polynomial of degree at most n. In all cases we obtain $z_{-j,m}^{-\mathbf{k}} = \overline{z_{j,m}^{\mathbf{k}}}$, and $z_{j,m}^{\mathbf{k}} \ne 0$ only if $\mathbf{k} \in \mathcal{K}_j$ as defined in (11)–(12).

4.2. Bounds of the modulation functions. In view of the expansion (27) of the modulation functions in terms of $\delta = \epsilon^{\theta/2}$ we expect that the coefficient functions $z_{j,m}^{\mathbf{k}}$ still carry some power of ϵ. The structure of the equation (28) suggests that a control of these coefficient functions multiplied with

$$\tilde{\gamma}_j^{\mathbf{k}} = \max\bigl(\epsilon^{-|j|}, \epsilon^{-2|k_0| - \sum_{l\ge1} l|k_l|}\bigr)^{(1-\theta)/2}$$

for $j \ne 0$ and

$$\tilde{\gamma}_j^{\mathbf{k}} = \max\bigl(\epsilon^{-2}, \epsilon^{-2|k_0| - \sum_{l\ge1} l|k_l|}\bigr)^{(1-\theta)/2}$$

for $j = 0$ should be possible (note that $\tilde{\gamma}_j^{\mathbf{k}} \le \tilde{\gamma}_{j_1}^{\mathbf{k}^1}\tilde{\gamma}_{j_2}^{\mathbf{k}^2}$ for $j_1 + j_2 = j$ and $\mathbf{k}^1 + \mathbf{k}^2 = \mathbf{k}$). We use the norm, for $0 \le t \le 1$,

$$|\!|\!| z |\!|\!|_t = \sum_{l\ge0} \frac{1}{l!}\Bigl|\frac{d^l}{dt^l} z(t)\Bigr|$$

and note the properties $|\!|\!| z\cdot w |\!|\!|_t \le |\!|\!| z |\!|\!|_t \cdot |\!|\!| w |\!|\!|_t$, $|\!|\!| z |\!|\!|_t \le |z(0)| + 2\sup_{0\le\tau\le t} |\!|\!| \dot{z} |\!|\!|_\tau$, and $|\!|\!| \dot{p} |\!|\!|_t \le d |\!|\!| p |\!|\!|_t$, if p is a polynomial of degree d. In particular, for z of (30) we have

$$|\!|\!| z |\!|\!|_t \le C |\!|\!| p |\!|\!|_t,$$

where C depends only on an upper bound of the coefficients $|\alpha|$ and $|\beta|$ and of the degree of the polynomial p.

Lemma 1. *Let $s > 1/2$ and $\kappa_j = \max(1,|j|^{2s})$ for $j \in \mathbb{Z}$. For $1 \le m \le 2K-1$ and $0 \le t \le 1$ we have*

$$\sum_{j\in\mathbb{Z}} \kappa_j \Bigl(\sum_{\mathbf{k}\in\mathcal{K}_j} \gamma_j^{\mathbf{k}} |\!|\!| z_{j,m}^{\mathbf{k}} |\!|\!|_t \Bigr)^2 \lesssim 1$$

with

$$\gamma_j^{\mathbf{k}} = \max(1,\omega_j,|\mathbf{k}\cdot\boldsymbol{\omega}|)\cdot\tilde{\gamma}_j^{\mathbf{k}}.$$

Proof. We use induction over m. The statement is evident for $m=1$. Let $\mathbf{k} \ne \pm\langle j\rangle$ and

$$\alpha_j^{\mathbf{k}} = -\frac{2\mathrm{i}(\mathbf{k}\cdot\boldsymbol{\omega})}{\omega_j^2 - (\mathbf{k}\cdot\boldsymbol{\omega})^2}, \qquad \beta_j^{\mathbf{k}} = -\frac{1}{\omega_j^2 - (\mathbf{k}\cdot\boldsymbol{\omega})^2}.$$

By the non-resonance condition (9), $|\alpha_j^{\mathbf{k}}| + |\beta_j^{\mathbf{k}}| \lesssim 1$ for $\mathbf{k} \in \mathcal{K}_j$. For $\mathbf{k} \neq \pm\langle j\rangle$ the polynomial solution of (28) as given by (30) therefore satisfies the bound

$$|\!|\!| z_{j,m}^{\mathbf{k}} |\!|\!|_t \lesssim \sum_{j_1+j_2=j} \sum_{\mathbf{k}^1+\mathbf{k}^2=\mathbf{k}} \sum_{m_1+m_2=m} |\!|\!| z_{j_1,m_1}^{\mathbf{k}^1} |\!|\!|_t |\!|\!| z_{j_2,m_2}^{\mathbf{k}^2} |\!|\!|_t .$$

By the definition of $\gamma_j^{\mathbf{k}}$ we have $\gamma_j^{\mathbf{k}} \lesssim \gamma_{j_1}^{\mathbf{k}^1} \gamma_{j_2}^{\mathbf{k}^2}$ for $j_1 + j_2 = j$ and $\mathbf{k}^1 + \mathbf{k}^2 = \mathbf{k}$. Together with (15) this yields

$$\begin{aligned} &\sum_{j\in\mathbb{Z}} \kappa_j \Big(\sum_{\mathbf{k}\in\mathcal{K}_j, \mathbf{k}\neq\pm\langle j\rangle} \gamma_j^{\mathbf{k}} |\!|\!| z_{j,m}^{\mathbf{k}} |\!|\!|_t \Big)^2 \\ &\quad \lesssim \sum_{m_1+m_2=m} \bigg(\sum_{j\in\mathbb{Z}} \kappa_j \Big(\sum_{\mathbf{k}\in\mathcal{K}_j} \gamma_j^{\mathbf{k}} |\!|\!| z_{j,m_1}^{\mathbf{k}} |\!|\!|_t \Big)^2 \bigg) \bigg(\sum_{j\in\mathbb{Z}} \kappa_j \Big(\sum_{\mathbf{k}\in\mathcal{K}_j} \gamma_j^{\mathbf{k}} |\!|\!| z_{j,m_2}^{\mathbf{k}} |\!|\!|_t \Big)^2 \bigg) \lesssim 1. \end{aligned}$$

For $\mathbf{k} = \pm\langle j\rangle$ we obtain analogously

$$\sum_{j\in\mathbb{Z}} \kappa_j \Big(\gamma_j^{\pm\langle j\rangle} |\!|\!| \dot z_{j,m}^{\pm\langle j\rangle} |\!|\!|_t \Big)^2 \lesssim 1.$$

With $|\!|\!| z |\!|\!|_t \le |z(0)| + 2\sup_{0\le\tau\le t} |\!|\!| \dot z |\!|\!|_\tau$ we conclude

$$\sum_{j\in\mathbb{Z}} \kappa_j \Big(\gamma_j^{\pm\langle j\rangle} |\!|\!| z_{j,m}^{\pm\langle j\rangle} |\!|\!|_t \Big)^2 \lesssim 1$$

since the initial values $z_{j,m}^{\pm\langle j\rangle}(0)$ can be estimated accordingly by (31) and the bounds already given above. □

4.3. Proof of Theorem 4. We shall prove that, for $0 \le t \le 1$,

$$\begin{aligned} \sigma_0 \, |\mathcal{E}_0(t)| &\lesssim \delta^4 \\ \sigma_l \, |\mathcal{E}_l(t)| &\lesssim \delta^{2l}, \quad 0 < l < K, \\ \sum_{l\ge K} \sigma_l \, |\mathcal{E}_l(t)| &\lesssim \delta^{2K}. \end{aligned}$$

Because of $\delta = \epsilon^{\theta/2}$ the statement of Theorem 4 then follows from the definition of σ_l in (13). We start from

$$|\mathcal{E}_l(\mathbf{z}, \dot{\mathbf{z}})| \lesssim \sum_{j\in\mathbb{Z}} \Big(\sum_{\mathbf{k}\in\mathcal{K}_j} |k_l| \omega_l |\mathbf{k}\cdot\boldsymbol{\omega}| |z_j^{\mathbf{k}}|^2 + \sum_{\mathbf{k}\in\mathcal{K}_j} |k_l| \omega_l |z_j^{\mathbf{k}}| |\dot z_j^{\mathbf{k}}| \Big). \tag{33}$$

From $k_l \neq 0$ we infer $\mu(\mathbf{k}) \ge m_0(l)$, where $m_0(l) = \min(l, K)$ if $l \ge 1$, and $m_0(0) = 2$. Therefore, (32c) yields

$$z_j^{\mathbf{k}}(t) = \sum_{m=m_0(l)}^{2K-1} \delta^m z_{j,m}^{\mathbf{k}}(t).$$

Lemma 1 and the inequality $\sum_{l\ge0} \sigma_l |k_l| \omega_l \max(1, |\mathbf{k}\cdot\boldsymbol{\omega}|) \lesssim \kappa_j \big(\gamma_j^{\mathbf{k}}\big)^2$ for $\mathbf{k} \in \mathcal{K}_j$ thus imply that the sums in (33) are estimated to yield the above bounds.

4.4. Bounds of the defect. The following bound implies (20).

Lemma 2. *The defect in (8) satisfies, for $0 \le t \le 1$,*

$$\sum_{j\in\mathbb{Z}} \kappa_j \Big(\sum_{\mathbf{k}\in\mathcal{K}_j} \gamma_j^{\mathbf{k}} |\!|\!| d_j^{\mathbf{k}} |\!|\!|_t \Big)^2 \lesssim \delta^{2K}$$

with $\gamma_j^{\mathbf{k}}$ as in Lemma 1.

Proof. By construction of the z_j^k,

$$d_j^{\mathbf{k}} = - \sum_{m=2K}^{2(2K-1)} \delta^m \sum_{j_1+j_2=j} \sum_{\mathbf{k}^1+\mathbf{k}^2=\mathbf{k}} \sum_{m_1+m_2=m} z_{j_1,m_1}^{\mathbf{k}^1} z_{j_2,m_2}^{\mathbf{k}^2}.$$

The result then follows by proceeding as in the proof of Lemma 1 and using the bounds of that lemma. □

4.5. Bounds of the remainder. We consider the differential equation (6) for the sequence of Fourier coefficients $\mathbf{u} = (u_j)_{j\in\mathbb{Z}}$, which reads

$$\begin{aligned} \boldsymbol{\Omega}\dot{\mathbf{u}} &= \boldsymbol{\Omega}\mathbf{v}, \\ \dot{\mathbf{v}} &= -\boldsymbol{\Omega}(\boldsymbol{\Omega}\mathbf{u}) + \mathbf{u} * \mathbf{u} \end{aligned}$$

for $g(u) = u^2$. Using the variation-of-constants formula, the assumptions on the initial values and a bootstrap argument involving (14), we obtain in the norm (13)

$$\|\mathbf{u}(t)\| \lesssim \delta \quad \text{for} \quad 0 \le t \le 1.$$

For the solution $\widetilde{\mathbf{u}}$ of the system $\ddot{\widetilde{\mathbf{u}}} = -\boldsymbol{\Omega}^2\widetilde{\mathbf{u}} + \widetilde{\mathbf{u}} * \widetilde{\mathbf{u}} + \boldsymbol{d}$ with perturbation $\boldsymbol{d}(t) = (d_j(t))_{j\in\mathbb{Z}}$ given as

$$d_j(t) = \sum_{\mathbf{k}\in\mathcal{K}_j} d_j^{\mathbf{k}}(t) \mathrm{e}^{\mathrm{i}(\mathbf{k}\cdot\boldsymbol{\omega})t},$$

we likewise obtain

$$\|\widetilde{\mathbf{u}}(t)\| \lesssim \delta \quad \text{for} \quad 0 \le t \le 1.$$

Using (14) we obtain the bound

$$\|\mathbf{u} * \mathbf{u} - \widetilde{\mathbf{u}} * \widetilde{\mathbf{u}}\| \lesssim \delta \|\mathbf{u} - \widetilde{\mathbf{u}}\| \lesssim \delta \|\boldsymbol{\Omega}(\mathbf{u} - \widetilde{\mathbf{u}})\|$$

Using this bound together with the variation-of-constants formula and the Gronwall lemma, we conclude for the error $\mathbf{r} = \mathbf{u} - \widetilde{\mathbf{u}}$,

$$\|(\mathbf{r}(t), \dot{\mathbf{r}}(t))\| \le \int_0^t \|\boldsymbol{d}(\tau)\| \, d\tau \lesssim \delta^{2K} t \quad \text{for} \quad 0 \le t \le 1, \tag{34}$$

using (20) in the last inequality. This gives (19) and completes the proof of Theorem 3.

5. Proof of Theorems 5 to 7

5.1. Almost-invariant energies: proof of Theorem 5. From (22) and (23) we have

$$\frac{d}{dt}\mathcal{E}_l(\mathbf{z},\dot{\mathbf{z}}) = -\tfrac{1}{2}\sum_{j\in\mathbb{Z}}\sum_{\mathbf{k}\in\mathcal{K}_j} \mathrm{i}k_l\omega_l z_{-j}^{-\mathbf{k}} d_j^{\mathbf{k}}.$$

For $0 \le l \le K$ we therefore obtain, using the Cauchy-Schwarz inequality and $\sigma_l\omega_l \lesssim \kappa_j\big(\gamma_j^{\mathbf{k}}\big)^2$ if $k_l \neq 0$ and $\mathbf{k}\in\mathcal{K}_j$, with $\gamma_j^{\mathbf{k}}$ as in Lemma 1,

$$\begin{aligned}\sigma_l\left|\frac{d}{dt}\mathcal{E}_l(\mathbf{z},\dot{\mathbf{z}})\right| &\lesssim \sum_{j\in\mathbb{Z}}\kappa_j \sum_{\substack{\mathbf{k}\in\mathcal{K}_j\\ k_l\neq 0}} \gamma_j^{\mathbf{k}}|z_j^{\mathbf{k}}|\,\gamma_j^{\mathbf{k}}|d_j^{\mathbf{k}}| \\ &\le \Big(\sum_{j\in\mathbb{Z}}\kappa_j\Big(\sum_{\substack{\mathbf{k}\in\mathcal{K}_j\\ k_l\neq 0}}\gamma_j^{\mathbf{k}}|z_j^{\mathbf{k}}|\Big)^2\Big)^{1/2}\Big(\sum_{j\in\mathbb{Z}}\kappa_j\Big(\sum_{\substack{\mathbf{k}\in\mathcal{K}_j\\ k_l\neq 0}}\gamma_j^{\mathbf{k}}|d_j^{\mathbf{k}}|\Big)^2\Big)^{1/2}.\end{aligned}$$

Noting that for $l \ge 1$,

$$z_j^{\mathbf{k}} = \sum_{m=l}^{2K-1}\delta^m z_{j,m}^{\mathbf{k}} \quad \text{if} \quad k_l \neq 0,$$

and for $l = 0$, $z_j^{\mathbf{k}} = \sum_{m=2}^{2K-1}\delta^m z_{j,m}^{\mathbf{k}}$ if $k_0 \neq 0$, the stated bound follows from Lemma 1 and Lemma 2. For

$$\sum_{l\ge K}\sigma_l\left|\frac{d}{dt}\mathcal{E}_l(\mathbf{z},\dot{\mathbf{z}})\right| \le \sum_{j\in\mathbb{Z}}\sum_{\substack{\mathbf{k}\in\mathcal{K}_j\\ \mu(\mathbf{k})\ge K}}\sum_{l\ge K}\sigma_l|k_l|\omega_l|z_j^{\mathbf{k}}||d_j^{\mathbf{k}}|$$

the bound is obtained similarly on using $\sum_{l\ge K}\sigma_l|k_l|\omega_l \lesssim \kappa_j\big(\gamma_j^{\mathbf{k}}\big)^2$ for $\mathbf{k}\in\mathcal{K}_j$.

5.2. Transitions in the almost-invariant energies: proof of Theorem 6. First we note that $\mathbf{z}(t+1)$ contains the modulation functions that are uniquely constructed (up to $O(\delta^{2K})$) by starting from $(\widehat{\mathbf{u}}(1),\dot{\widehat{\mathbf{u}}}(1))$, where $\widehat{\mathbf{u}} = (\widehat{u}_j)$ is the truncated modulated Fourier expansion (18) without the remainder term $\mathbf{r} = (r_j)$. On the other hand, $\widetilde{\mathbf{z}}(t)$ contains the modulation functions constructed by starting from the Fourier coefficients $\mathbf{u} = (u_j)$ of the solution. By Theorem 3 we have $\mathbf{u} = \widehat{\mathbf{u}} + \mathbf{r}$ with the remainder estimate (19). We thus need to estimate $|||\mathbf{z}(\cdot+1)-\widetilde{\mathbf{z}}|||_t$ at $t=0$ in terms of $\|\widehat{\mathbf{u}}(1)-\mathbf{u}(1)\| = \|\mathbf{r}(1)\|$. We proceed similarly to the proof of Lemma 1, taking differences in the recursions instead of direct bounds. Omitting the details, we obtain

$$\sum_{j\in\mathbb{Z}}\kappa_j\Big(\sum_{\mathbf{k}\in\mathcal{K}_j}\gamma_j^{\mathbf{k}}|||z_j^{\mathbf{k}}(\cdot+1)-\widetilde{z}_j^{\mathbf{k}}|||_t\Big)^2 \lesssim \delta^{2K}$$

at $t=0$. Together with the definition of $\mathcal{E}_l$ and the bounds of Lemma 1, this yields the stated bound.

5.3. Controlling mode energies by almost-invariant energies: proof of Theorem 7. For the proof of Theorem 7 we first show that we can control the mode energies if we can control the dominant terms in the modulation functions.

Lemma 3. *For $0 \le t \le 1$, assume that for all $\mathbf{k}$ with $\mu(\mathbf{k}) \le K$,*

$$\begin{aligned} \sigma_0 |z_{0,2}^{\mathbf{k}}(t)|^2 &\le \mathcal{C}_0, \\ \sigma_j |z_{j,|j|}^{\mathbf{k}}(t)|^2 &\le \mathcal{C}_0 \quad \text{for} \quad 0 < |j| \le K, \\ \sum_{l \ge K} \sigma_l \omega_l^2 |z_{l,K}^{\pm\langle l\rangle}(t)|^2 &\le \mathcal{C}_0. \end{aligned}$$

Then,

$$\begin{aligned} \sigma_0 E_0(t) &\le \mathcal{C}\delta^4 \\ \sigma_l E_l(t) &\le \mathcal{C}\delta^{2l} \quad \text{for} \quad l = 1, \dots, K-1, \\ \sum_{l \ge K} \sigma_l E_l(t) &\le \mathcal{C}\delta^{2K} \end{aligned}$$

and

$$\sigma_1 \Big| E_1(t) - \omega_1^2 \big(|z_1^{\langle 1\rangle}|^2 + |z_1^{-\langle 1\rangle}|^2 \big) \Big| \le \mathcal{C}\delta^4,$$

where $\mathcal{C}$ depends on $\mathcal{C}_0$ but is independent of C_0 of (17), if δ is sufficiently small.

Proof. We have

$$\begin{aligned} E_l(t) = \frac{1}{2}|\omega_l u_l(t)|^2 + \frac{1}{2}|\dot u_l(t)|^2 &\le \Big(\sum_{\mathbf{k}} \omega_l |z_l^{\mathbf{k}}(t)| \Big)^2 \\ &+ 2\Big(\sum_{\mathbf{k}} |\dot z_l^{\mathbf{k}}(t)| \Big)^2 + 2\Big(\sum_{\mathbf{k}} |\mathbf{k}\cdot\boldsymbol{\omega}| |z_l^{\mathbf{k}}(t)| \Big)^2 + \omega_l^2 |r_l(t)|^2 + |\dot r_l(t)|^2. \end{aligned}$$

Let $m_0(l) = \min(l, K)$ if $l \ge 1$, and $m_0(l) = 2$ in case $l = 0$, so that $z_{l,m}^{\mathbf{k}}(t) = 0$ for $m < m_0(l)$ by (32a) and $\dot z_{l,m_0(l)}^{\mathbf{k}} = 0$ for all $\mathbf{k}$ by (32b). Therefore,

$$\begin{aligned} \sigma_l E_l(t) &\le 6\delta^{2m_0(l)} \sigma_l \Big(\sum_{\mathbf{k}} \max(\omega_l, |\mathbf{k}\cdot\boldsymbol{\omega}|) |z_{l,m_0(l)}^{\mathbf{k}}(t)| \Big)^2 \\ &+ 8\delta^{2m_0(l)+2} \kappa_l \Big(\sum_{\mathbf{k}} \gamma_l^{\mathbf{k}} \sum_{m > m_0(l)} \| z_{l,m}^{\mathbf{k}} \|_t \Big)^2 + \sigma_l \omega_l^2 |r_l(t)|^2 + \sigma_l |\dot r_l(t)|^2. \end{aligned}$$

For $l > K$ only diagonal elements are among the dominant terms in the modulated Fourier expansion by (32d), $z_{l,m_0(l)}^{\mathbf{k}} = 0$ for $\mathbf{k} \ne \pm\langle l\rangle$. With the assumption, with Lemma 1 and the estimate (34) of the remainder we obtain

$$\sum_{l > K} \sigma_l E_l(t) \le 12\delta^{2K} + C\delta^{2K+2} + C\delta^{4K},$$

where C depends on C_0, but not on δ.

We now study the case $l \le K$, where off-diagonal elements $z_l^{\mathbf{k}}$ with $\mathbf{k} \ne \langle l\rangle$ can be among the dominant terms in the modulated Fourier expansion. Such modulation functions (whose number depends only on K) are estimated by the assumption. Together with Lemma 1 and the remainder estimate we obtain the bound of $E_l(t)$ for $l < K$.

Finally we study E_1 in more detail. We have (omitting the argument t in the notation)

$$\begin{aligned} 2E_1 &= |\omega_1 u_1|^2 + |\dot u_1|^2 \\ &= \omega_1^2 \big| z_1^{\langle 1\rangle} \mathrm{e}^{\mathrm{i}\omega_1 t} + z_1^{-\langle 1\rangle} e^{-\mathrm{i}\omega_1 t} \big|^2 + 2\omega_1^2 \mathrm{Re}\Big(\overline{\eta} \big(z_1^{\langle 1\rangle} \mathrm{e}^{\mathrm{i}\omega_1 t} + z_1^{-\langle 1\rangle} e^{-\mathrm{i}\omega_1 t} \big) \Big) + \omega_1^2 |\eta|^2 \\ &\quad + \omega_1^2 \big| \mathrm{i} z_1^{\langle 1\rangle} \mathrm{e}^{\mathrm{i}\omega_1 t} - \mathrm{i} z_1^{-\langle 1\rangle} e^{-\mathrm{i}\omega_1 t} \big|^2 + 2\omega_1 \mathrm{Re}\Big(\overline{\vartheta} \big(\mathrm{i} z_1^{\langle 1\rangle} \mathrm{e}^{\mathrm{i}\omega_1 t} - \mathrm{i} z_1^{-\langle 1\rangle} e^{-\mathrm{i}\omega_1 t} \big) \Big) + |\vartheta|^2 \end{aligned}$$

with

$$\eta = \sum_{\mathbf{k} \neq \pm\langle 1\rangle} z_1^{\mathbf{k}} \mathrm{e}^{\mathrm{i}(\mathbf{k}\cdot\boldsymbol{\omega})t} + r_1 \quad \text{and} \quad \vartheta = \sum_{\mathbf{k} \neq \pm\langle 1\rangle} \mathrm{i}(\mathbf{k}\cdot\boldsymbol{\omega}) z_1^{\mathbf{k}} \mathrm{e}^{\mathrm{i}(\mathbf{k}\cdot\boldsymbol{\omega})t} + \sum_{\mathbf{k}} \dot z_1^{\mathbf{k}} \mathrm{e}^{\mathrm{i}(\mathbf{k}\cdot\boldsymbol{\omega})t} + \dot r_1.$$

The estimate of $|E_1 - \omega_1^2(|z_1^{\langle 1\rangle}|^2 + |z_1^{-\langle 1\rangle}|^2)|$ thus follows from Lemma 1, the estimates of the remainder (34), and the fact that $z_{1,m}^{\mathbf{k}} = 0$ for $m = 2$ and all $\mathbf{k}$. □

We now control the off-diagonal modulation functions $z_j^{\mathbf{k}}$ for small $|j| \le K$ in terms of the diagonal entries $z_l^{\langle l\rangle}$.

Lemma 4. *For $0 \le t \le 1$ assume that*

$$\begin{aligned} &\sigma_0 |z_{0,2}^{\pm\langle 0\rangle}(t)| \le \mathcal{C}_0, \\ &\sigma_j |z_{j,|j|}^{\pm\langle j\rangle}(t)| \le \mathcal{C}_0 \quad \textit{for } 0 < |j| \le K. \end{aligned}$$

Then,

$$\begin{aligned} &\sigma_0 |z_{0,2}^{\mathbf{k}}(t)| \le \mathcal{C} \quad \textit{for all } \mathbf{k}, \\ &\sigma_j |z_{j,|j|}^{\mathbf{k}}(t)| \le \mathcal{C} \quad \textit{for } 0 < |j| \le K \textit{ and all } \mathbf{k}, \end{aligned}$$

where $\mathcal{C}$ depends on $\mathcal{C}_0$, but is independent of C_0 of (17).

Proof. Let $m_0(j)$ be the smallest index m for which $z_{j,m}^{\pm\langle j\rangle}$ can be different from zero, i.e., $m_0(j) = \min(|j|, K)$ for $j \neq 0$ and $m_0(0) = 2$. We prove the result by induction over $m_0(j)$. For $m_0(j) = 1$, which requires $j = \pm 1$, the estimate follows from the explicit formulas (case $m = 1$) of Section 4.1. Let now $j \in \mathbb{Z}$ with $m_0(j) > 1$. Since $\dot z_{j,m_0(j)}^{\mathbf{k}} = 0$, the recurrence relation of the modulation functions yields

$$(\omega_j^2 - (\mathbf{k}\cdot\boldsymbol{\omega})^2)\, z_{j,m_0(j)}^{\mathbf{k}} = \sum_{j_1+j_2=j} \sum_{\mathbf{k}^1+\mathbf{k}^2=\mathbf{k}} \sum_{m_1+m_2=m_0(j)} z_{j_1,m_1}^{\mathbf{k}^1} z_{j_2,m_2}^{\mathbf{k}^2}.$$

Since $z_{j_1,m_1}^{\mathbf{k}^1} = 0$ for $m_1 < m_0(j_1)$ and $m_0(j) \le m_0(j_1) + m_0(j_2)$, this simplifies to

$$(\omega_j^2 - (\mathbf{k}\cdot\boldsymbol{\omega})^2)\, z_{j,m_0(j)}^{\mathbf{k}} = \sum_{\substack{j_1+j_2=j \\ m_0(j_1)+m_0(j_2)=m_0(j)}} \sum_{\mathbf{k}^1+\mathbf{k}^2=\mathbf{k}} z_{j_1,m_0(j_1)}^{\mathbf{k}^1} z_{j_2,m_0(j_2)}^{\mathbf{k}^2}.$$

The number of terms in the sum depends only on K. Since $m_0(j_1) < m_0(j)$ and $m_0(j_2) < m_0(j)$, the result follows by induction using (9) and $\sigma_j \lesssim \sigma_{j_1}\sigma_{j_2}$. □

Finally we show that the almost-invariant energies control the diagonal modulation functions.

Lemma 5. *For* $0 \le t \le 1$ *we have*

$$\sigma_0 \left| \mathcal{E}_0(t) - \omega_0^2 \big(|z_0^{\langle 0\rangle}(t)|^2 + |z_0^{-\langle 0\rangle}(t)|^2 \big) \right| \lesssim \delta^5$$

$$\sigma_1 \left| \mathcal{E}_1(t) - \omega_1^2 \big(|z_1^{\langle 1\rangle}(t)|^2 + |z_1^{-\langle 1\rangle}(t)|^2 \big) \right| \lesssim \delta^4$$

$$\sigma_l \left| \mathcal{E}_l(t) - \omega_l^2 \big(|z_l^{\langle l\rangle}(t)|^2 + |z_l^{-\langle l\rangle}(t)|^2 \big) \right| \lesssim \delta^{2l+1}, \quad 1 < l < K,$$

$$\sum_{l \ge K} \sigma_l \left| \mathcal{E}_l(t) - \omega_l^2 \big(|z_l^{\langle l\rangle}(t)|^2 + |z_l^{-\langle l\rangle}(t)|^2 \big) \right| \lesssim \delta^{2K+1}.$$

Proof. The proof is very similar to that of Theorem 4 in Section 4.3. The only difference is that we subtract the dominant term $\omega_l^2 \big(|z_l^{\langle l\rangle}|^2 + |z_l^{-\langle l\rangle}|^2 \big)$ from $\mathcal{E}_l(\mathbf{z}, \dot{\mathbf{z}})$, so that the first sum over $\mathbf{k} \in \mathcal{K}_j$ in (33) is only over multi-indices $\mathbf{k}$ satisfying $\mathbf{k} \ne \pm\langle j\rangle$. For $k_l \ne 0$ and $\mathbf{k} \ne \pm\langle l\rangle$ we have $\mu(\mathbf{k}) \ge m_0(l) + 1$, so that by (32c)

$$z_j^{\mathbf{k}}(t) = \sum_{m=m_0(l)+1}^{2K-1} \delta^m z_{j,m}^{\mathbf{k}}(t), \qquad k_l \ne 0 \;\text{ and }\; \mathbf{k} \ne \pm\langle l\rangle.$$

By (32b) we have $\dot{z}_{j,m_0(l)}^{\pm\langle l\rangle} = 0$ for $|j| = l$ and by (32e) we have $z_{j,m_0(l)}^{\pm\langle l\rangle} = 0$ for $|j| \ne l$. Compared to the estimates in Section 4.3 we thus gain one power of δ. This yields the stated bounds for $l \ne 1$. For $l = 1$ we use that $z_{1,m}^{\mathbf{k}} = 0$ for $m = 2$ and all $\mathbf{k}$. $\square$

Lemmas 3 to 5 yield Theorem 7. For the estimate of $\mathcal{E}_1 - E_1$ we use these lemmas with $\theta = 1$.

6. On the non-resonance condition — Proof of Theorem 2

Fix $K \ge 2$ and $\rho_0 > 0$. Recall that for $\rho \ge 0$ the frequencies are defined as $\omega_j = \sqrt{j^2 + \rho}$. Throughout the proof of Theorem 2 we consider these frequencies as functions of ρ. The proof is based on the corresponding proof of Bambusi and Grébert [2] for the situation of initial values with all mode energies of order ϵ.

In a first step we show that it suffices to control a finite number of linear combinations of frequencies in the non-resonance condition (9) (despite the fact that the set $\mathcal{K}$ from (11) is infinite). We denote by c_0 a constant depending only on ρ_0 such that

$$\omega_j - |j| \le \frac{c_0}{|j|+1} \quad \text{for all} \quad 0 \le \rho \le \rho_0.$$

Moreover, we introduce for $L \ge 2$

$$\gamma_0(L, \rho) = \min\big\{ |\mathbf{k} \cdot \boldsymbol{\omega}(\rho) + r| : r \in \mathbb{Z},\, |r| < K,\, |r| + \mu(\mathbf{k}) \ne 0, \\ \mu(\mathbf{k}) < 3K,\, k_l = 0 \text{ for all } l \ge L \big\}.$$

In this first step we assume ρ such that $\gamma_0(K, \rho) \ne 0$. Besides the value $L = K$ we are particularly interested in $L = \widetilde{K} := (1 + c_0)K + \frac{4c_0}{\gamma_0(K,\rho)}$.

Now let $(j,\mathbf{k})$ be such that $|\omega_j - |\mathbf{k}\cdot\boldsymbol{\omega}||$ should satisfy the non-resonance condition (9), i.e., let $(j,\mathbf{k}) \in \mathcal{K}$ with $\mathbf{k} \neq \pm\langle j\rangle$. We distinguish two cases: Either

$$\big|\omega_j - |\mathbf{k}\cdot\boldsymbol{\omega}|\big| = \big|\widetilde{\mathbf{k}}\cdot\boldsymbol{\omega}\big|$$

with $\widetilde{\mathbf{k}} \neq 0$, $\mu(\widetilde{\mathbf{k}}) < 3K$ and $\widetilde{k}_l = 0$ for $l \geq \widetilde{K}$ or

$$\big|\omega_j - |\mathbf{k}\cdot\boldsymbol{\omega}|\big| = \big|\omega_j \pm \omega_{j-r} + \widetilde{\mathbf{k}}\cdot\boldsymbol{\omega}\big|$$

with $\mu(\widetilde{\mathbf{k}}) < K$, $|r| < K$, $|r| + \mu(\widetilde{\mathbf{k}}) \neq 0$ and $\max(|j|, |j-r|) \geq \widetilde{K}$. The first case comprises the indices in the first part of the set $\mathcal{K}$ and the indices in the second part of this set that are not too large. The second case deals with indices in the second part of the set $\mathcal{K}$, where j and $j-r$ are large indices. By definition of γ_0 we have in the first case

$$\big|\omega_j - |\mathbf{k}\cdot\boldsymbol{\omega}|\big| \geq \gamma_0(\widetilde{K},\rho).$$

In the second case we either have (in case "+")

$$\big|\omega_j - |\mathbf{k}\cdot\boldsymbol{\omega}|\big| \geq \omega_j + \omega_{j-r} - \big|\widetilde{\mathbf{k}}\cdot\boldsymbol{\omega}\big| \geq \widetilde{K} - (1+c_0)\mu(\widetilde{\mathbf{k}}) \geq 1,$$

or (in case "−")

$$\big|\omega_j - |\mathbf{k}\cdot\boldsymbol{\omega}|\big| \geq \big|\widetilde{\mathbf{k}}\cdot\boldsymbol{\omega} \pm r\big| - \big|\omega_j - |j|\big| - \big|\omega_{j-r} - |j-r|\big| \geq \frac{1}{2}\gamma_0(K,\rho).$$

What remains to be shown in a second step is that $\gamma_0(\widetilde{K},\rho) = 0$ (and hence also $\gamma_0(K,\rho) = 0$) only for finitely many $0 < \rho < \rho_0$. We first consider the function

$$f_{\mathbf{k},r}(\rho) = \mathbf{k}\cdot\boldsymbol{\omega}(\rho) + r$$

for fixed $r \in \mathbb{Z}$ and a fixed sequence of integers $\mathbf{k}$ with $|r| + \mu(\mathbf{k}) \neq 0$, $\mu(\mathbf{k}) < 3K$ and $k_l = 0$ for all $l \geq \widetilde{K}$. The square matrix

$$\Big(\frac{d^m}{d\rho^m}\omega_l\Big)_{m=0,\ldots,M-1;\, l:k_l\neq 0},$$

with M equal to the number of indices l with $k_l \neq 0$, is invertible for $\rho > 0$, as is shown by computing its determinant of Vandermonde form, see [2, Lemma 5.1]. Hence there exists for all $\rho > 0$ a $0 \leq m \leq M-1$ such that $f^{(m)}_{\mathbf{k},r}(\rho) \neq 0$. The continuity of $f^{(m)}_{\mathbf{k},r}$ implies for any $\rho > 0$ the existence of an open neighbourhood where $f^{(m)}_{\mathbf{k},r}$ is never zero, and consequently where $f_{\mathbf{k},r}$ has only finitely many zeros. For $\rho = 0$ the same argument shows the existence of an half-open interval $[0,a)$ where $f_{\mathbf{k},r}$ has only finitely many zeros provided that $k_0 = 0$. If $k_0 \neq 0$ we use the fact that $\omega_0 = \sqrt{\rho}$ grows in a neighbourhood of zero faster than all other frequencies to show the existence of an half-open interval $[0,a)$ with the same property. We have thus constructed an open covering of $[0,\rho_0]$ where each set of the covering contains only finitely many zeros of $f_{\mathbf{k},r}$. This implies that $f_{\mathbf{k},r}$ has only finitely many zeros in $[0,\rho_0]$. Taking the union of these zeros over the finite number of r and $\mathbf{k}$ allowed in the definition of γ_0 we get a finite number of $0 \leq \rho \leq \rho_0$ for which $\gamma_0(\widetilde{K},\rho) = 0$. This completes the proof of Theorem 2.

Acknowledgements. This work has been supported by Fonds National Suisse, Project No. 200020-126638, and by DFG, Project LU 532/4-1 and SFB/TR7.

References

1. D. Bambusi, *Birkhoff normal form for some nonlinear PDEs*, Comm. Math. Phys. 234 (2003), 253–285.
2. D. Bambusi and B. Grébert, *Birkhoff normal form for PDEs with tame modulus*, Duke Math. J. 135 (2006), 507–567.
3. D. Bambusi and A. Ponno, *On metastability in FPU*, Comm. Math. Phys. 264 (2006), 539–561.
4. D. Bambusi, J.-M. Delort, B. Grébert, J. Szeftel, *Almost global existence for Hamiltonian semi-linear Klein-Gordon equations with small Cauchy data on Zoll manifolds*, Comm. Pure Appl. Math. 60 (2007), 1665–1690.
5. J. Bourgain, *Construction of approximative and almost periodic solutions of perturbed linear Schrödinger and wave equations*, Geom. Funct. Anal. 6 (1996), 201–230.
6. J. Bourgain, *On diffusion in high-dimensional Hamiltonian systems and PDE*, J. Anal. Math. 80 (2000) , 1–35.
7. R. Carles and E. Faou, *Energy cascades for NLS on the torus*, Discrete Contin. Dyn. Syst. 32 (2012), 2063–2077.
8. D. Cohen, E. Hairer, and C. Lubich, *Long-time analysis of nonlinearly perturbed wave equations via modulated Fourier expansions*, Arch. Ration. Mech. Anal. 187 (2008), 341–368.
9. J.-M. Delort and J. Szeftel, *Long-time existence for small data nonlinear Klein-Gordon equations on tori and spheres*, Int. Math. Res. Not. 37 (2004), 1897–1966.
10. E. Faou and B. Grébert, *A Nekhoroshev type theorem for the nonlinear Schrödinger equation on the d-dimensional torus*, Preprint, 2010, arXiv:1003.4845v1.
11. E. Fermi, J. Pasta, and S. Ulam, *Studies of non linear problems*, Tech. Report LA-1940, Los Alamos, 1955, Later published in E. Fermi: Collected Papers, Chicago, 1965 and Lect. Appl. Math. 15, 143, 1974.
12. G. Gallavotti (ed.), *The Fermi-Pasta-Ulam problem. A status report*, Lecture Notes in Physics, vol. 728, Springer, Berlin, 2008.
13. L. Gauckler and C. Lubich, *Nonlinear Schrödinger equations and their spectral semi-discretizations over long times*, Found. Comput. Math. 10 (2010), 141–169.
14. L. Gauckler, *Long-time analysis of Hamiltonian partial differential equations and their discretizations*, Doctoral thesis, Univ. Tübingen, 2010. `http://nbn-resolving.de/urn:nbn:de:bsz:21-opus-47540`
15. B. Grébert, *Birkhoff normal form and Hamiltonian PDEs*, Partial differential equations and applications, Sémin. Congr., 15, Soc. Math. France, Paris, 2007, 1–46
16. E. Hairer and C. Lubich, *Long-time energy conservation of numerical methods for oscillatory differential equations*, SIAM J. Numer. Anal. 38 (2001), 414–441.
17. E. Hairer, C. Lubich, and G. Wanner, *Geometric numerical integration. Structure-preserving algorithms for ordinary differential equations*, 2nd ed., Springer Series in Computational Mathematics 31, Springer-Verlag, Berlin, 2006.
18. E. Hairer and C. Lubich, *On the energy distribution in Fermi–Pasta–Ulam lattices*, Preprint, 2010, `http://na.uni-tuebingen.de/preprints.shtml`.