STRONG NONLINEAR INSTABILITY AND GROWTH OF SOBOLEV NORMS NEAR QUASIPERIODIC FINITE-GAP TORI FOR THE 2D CUBIC NLS EQUATION

M. GUARDIA, Z. HANI, E. HAUS, A. MASPERO, AND M. PROCESI

ABSTRACT. We consider the defocusing cubic nonlinear Schrödinger equation (NLS) on the two-dimensional torus. The equation admits a special family of elliptic invariant quasiperiodic tori called finite-gap solutions. These are inherited from the integrable 1D model (cubic NLS on the circle) by considering solutions that depend only on one variable. We study the long-time stability of such invariant tori for the 2D NLS model and show that, under certain assumptions and over sufficiently long timescales, they exhibit a strong form of *transverse instability* in Sobolev spaces $H^s(\mathbb{T}^2)$ ($0 < s < 1$). More precisely, we construct solutions of the 2D cubic NLS that start arbitrarily close to such invariant tori in the H^s topology and whose H^s norm can grow by any given factor. This work is partly motivated by the problem of infinite energy cascade for 2D NLS, and seems to be the first instance where (unstable) long-time nonlinear dynamics near (linearly stable) quasiperiodic tori is studied and constructed.

1. INTRODUCTION

A widely held principle in dynamical systems theory is that invariant quasiperiodic tori play an important role in understanding the complicated long-time behavior of Hamiltonian ODE and PDE. In addition to being important in their own right, the hope is that such quasiperiodic tori can play an important role in understanding other, possibly more generic, dynamics of the system by acting as *islands* in whose vicinity orbits might spend long periods of time before moving to other such islands. The construction of such invariant sets for Hamiltonian PDE has witnessed an explosion of activity over the past thirty years after the success of extending KAM techniques to infinite dimensions. However, the dynamics near such tori is still poorly understood, and often restricted to the linear theory. The purpose of this work is to take a step in the direction of understanding and constructing non-trivial *nonlinear* dynamics in the vicinity of certain quasiperiodic solutions for the cubic defocusing NLS equation. In line with the above philosophy emphasizing the role of invariant quasiperiodic tori for other types of behavior, another aim is to push forward a program aimed at proving infinite Sobolev norm growth for the 2D cubic NLS equation, an outstanding open problem.

1.1. The dynamical system and its quasiperiodic objects. We start by describing the dynamical system and its quasiperiodic invariant objects at the center of our analysis. Consider the periodic cubic defocusing nonlinear Schrödinger equation (NLS),

$$\text{(2D-NLS)} \qquad \qquad \mathrm{i}\partial_t u + \Delta u = |u|^2 u$$

where $(x, y) \in \mathbb{T}^2 = \mathbb{R}^2/(2\pi\mathbb{Z})^2$, $t \in \mathbb{R}$ and $u : \mathbb{R} \times \mathbb{T}^2 \to \mathbb{C}$. All the results in this paper extend trivially to higher dimensions $d \geq 3$ by considering solutions that only depend on two variables[1]. This is a Hamiltonian

Departament de Matemàtica Aplicada I, Departament de Matemàtiques, Universitat Politècnica de Catalunya & BGSMATH Barcelona Graduate School of Mathematics, Barcelona, Spain, `marcel.guardia@upc.edu`.

Department of Mathematics, University of Michigan, 530 Church Street, Ann Arbor, MI 48109-1043, USA, `zhani@umich.edu`.

Dipartimento di Matematica e Fisica, Università di Roma Tre, Largo S.L. Murialdo, 00146 Roma, Italy, `ehaus@mat.uniroma3.it`.

International School for Advanced Studies (SISSA), Via Bonomea 265, 34136, Trieste, Italy `alberto.maspero@sissa.it`.

Dipartimento di Matematica e Fisica, Università di Roma Tre, Largo S.L. Murialdo, 00146 Roma, Italy, `procesi@mat.uniroma3.it`.

[1]We expect that the results also extend to the focusing sign of the nonlinearity ($-|u|^2 u$ on the R. H. S. of (2D-NLS)). The reason why we restrict to the defocusing sign comes from the fact that the linear analysis around our quasiperiodic tori has only been established in full detail in [MP18] in this case.

PDE with conserved quantities: i) the Hamiltonian

$$(1.1) \qquad H_0(u) = \int_{\mathbb{T}^2} \left(|\nabla u(x,y)|^2 + \frac{1}{2}|u(x,y)|^4 \right) \mathrm{d}x\,\mathrm{d}y,$$

ii) the mass

$$(1.2) \qquad M(u) = \int_{\mathbb{T}^2} |u(x,y)|^2 \mathrm{d}x\,\mathrm{d}y,$$

which is just the square of the L^2 norm of the solution, and iii) the momentum

$$(1.3) \qquad P(u) = \mathrm{i} \int_{\mathbb{T}^2} \overline{u(x,y)}\nabla u(x,y)\,\mathrm{d}x\,\mathrm{d}y.$$

We also remark that the equation is locally well-posed for data in $H^s(\mathbb{T}^d)$ for all $s > 0$ [Bou93]. Thanks to the conservation of energy (and the subcritical nature of the local well-posedness result), one directly obtains global well-posedness in $H^s(\mathbb{T}^d)$ for $s \geq 1$. This global existence can be pushed further down to at least $s > 2/3$ using almost conservation inequalities (See for instance [DSPSN07]). All the solutions constructed in this manuscript are infinitely smooth, and as such there global-in-time existence is guaranteed.

Now, we describe the invariant objects around which we will study and construct our long-time nonlinear dynamics. Of course, such a task requires a very precise understanding of the linearized dynamics around such objects. For this reason, we take the simplest non-trivial family of invariant quasiperiodic tori admitted by (2D-NLS), namely those inherited from its completely integrable 1D counterpart

$$(1D\text{-NLS}) \qquad \mathrm{i}\partial_t q = -\partial_{xx} q + |q|^2 q, \quad x \in \mathbb{T}.$$

This is a subsystem of (2D-NLS) if we consider solutions that depend only on the first spatial variable. It is well known that equation (1D-NLS) is integrable and its phase space is foliated by tori of finite or infinite dimension with periodic, quasiperiodic, or almost periodic dynamics. The quasiperiodic orbits are usually called *finite-gap solutions*.

Such tori are Lyapunov stable (for all time!) as solutions of (1D-NLS) (as will be clear once we exhibit its integrable structure) and some of them are linearly stable as solutions of (2D-NLS), but we will be interested in their *long-time nonlinear stability* (or lack of it) as invariant objects for the 2D equation (2D-NLS). In fact, we shall show that they are *nonlinearly unstable* as solutions of (2D-NLS), and in a strong sense, in certain topologies and after very long times. Such instability is transversal in the sense that one drifts along the purely 2-dimensional directions: solutions which are initially very close to 1-dimensional become strongly 2-dimensional after some long time scales[2].

1.2. Energy Cascade, Sobolev norm growth, and Lyapunov instability. In addition to studying long-time dynamics close to invariant objects for NLS, another purpose of this work is to make progress on a fundamental problem in nonlinear wave theory, which is the transfer of energy between characteristically different scales for a nonlinear dispersive PDE. This is called the *energy cascade* phenomenon. It is a purely nonlinear phenomenon (energy is static in frequency space for the linear system), and will be the underlying mechanism behind the long-time instability of the finite gap tori mentioned above.

We shall exhibit solutions whose energy moves from very high frequencies towards low frequencies (*backward or inverse cascade*), as well as ones that exhibit cascade in the opposite direction (*forward or direct cascade*). Such cascade phenomena have attracted a lot of attention in the past few years as they are central aspects of various theories of turbulence for nonlinear systems. For dispersive PDE, this goes by the name of *wave turbulence theory* which predicts the existence of solutions (and statistical states) of (2D-NLS) that exhibit a cascade of energy between very different length-scales. In the mathematical community, Bourgain drew attention

---

[2] The tranversal instability phenomenon was already studied for solitary waves of the water waves equation [RT11] and the KP-I equation [RT12] by Rousset and Tzvetkov. However, their instability is a *linear effect*, in the sense that the linearized dynamics is unstable. In contrast, our result is a fundamentally nonlinear effect, as the linearized dynamics around some of the finite gap tori is stable.

to such questions of energy cascade by first noting that it can be captured in a quantitative way by studying the behavior of the Sobolev norms of the solution

$$\|u\|_{H^s} = \left(\sum_{n \in \mathbb{Z}^2} (1 + |n|)^{2s} |\widehat{u}_n|^2 \right)^{\frac{1}{2}}.$$

In his list of Problems on Hamiltonian PDE [Bou00], Bourgain asked whether there exist solutions that exhibit a quantitative version of the forward energy cascade, namely solutions whose Sobolev norms H^s, with $s > 1$, are unbounded in time

$$(1.4) \qquad \qquad \sup_{t \geq 0} \|u(t)\|_{H^s} = +\infty, \qquad s > 1.$$

We should point out here that such growth cannot happen for $s = 0$ or $s = 1$ due to the conservation laws of the equations. For other Sobolev indices, there exists polynomial upper bounds for the growth of Sobolev norms (cf. [Bou96, Sta97, CDKS01, Bou04, Zho08, CW10, Soh11a, Soh12, Soh11b, CKO12, PTV17]). Nevertheless, results proving actual growth of Sobolev norms are much more scarce. After seminal works by Bourgain himself [Bou96] and Kuksin [Kuk96, Kuk97a, Kuk97b], the landmark result in [CKS⁺10] played a fundamental importance in the recent progress, including this work: It showed that for any $s > 1$, $\delta \ll 1$, $K \gg 1$, there exist solutions u of (2D-NLS) such that

$$(1.5) \qquad \qquad \|u(0)\|_{H^s} \leq \delta \quad \text{and} \quad \|u(T)\|_{H^s} \geq K$$

for some $T > 0$. Even if not mentioned in that paper, the same techniques also lead to the same result for $s \in (0, 1)$. This paper induced a lot of activity in the area [GK15, Han14, Gua14, HPTV15, HP15, GHP16] (see also [GG10, Del10, Poc11, GG12, Poc13, GG15, Mas18a] on results about growth of Sobolev norms with different techniques). Despite all that, Bourgain's question about solutions exhibiting (1.4) remains open on $\mathbb{T}^d$ (however a positive answer holds for the cylindrical domain $\mathbb{R} \times \mathbb{T}^d$, [HPTV15]).

The above-cited works revealed an intimate connection between Lyapunov instability and Sobolev norm growth. Indeed, the solution $u = 0$ of (2D-NLS) is an elliptic critical point and is linearly stable in all H^s. From this point of view, the result in [CKS⁺10] given in (1.5) can be interpreted as a strong form of Lyapunov instability (see item (6) in Section 1.4) in H^s, $s \neq 1$, of the elliptic critical point $u = 0$ (the first integrals (1.1) and (1.2) imply Lyapunov stability in the H^1 and L^2 topology). It turns out that this connection runs further, particularly in relation to the question of finding solutions exhibiting (1.4). As was observed in [Han14], one way to prove the existence of such solutions is to prove that, for sufficiently many $\phi \in H^s$, an instability similar to that in (1.5) holds, but with $\|u(0) - \phi\|_{H^s} \leq \delta$. In other words, proving long-time instability as in (1.5) but with solutions starting δ−close to ϕ, and for sufficiently many $\phi \in H^s$ implies the existence (and possible genericness) of unbounded orbits satisfying (1.4). Such a program (based on a Baire-Category argument) was applied successfully for the Szegö equation on $\mathbb{T}$ in [GG15].

Motivated by this, one is naturally led to studying this strong form of Lyapunov instability of more general invariant objects of (2D-NLS) (or other Hamiltonian PDEs), or equivalently to investigate whether one can achieve Sobolev norm explosion starting arbitrarily close to a given invariant object. The first work in this direction is by one of the authors [Han14]. He considers the plane waves $u(t, x) = Ae^{i(mx - \omega t)}$ with $\omega = m^2 + A^2$, periodic orbits of (2D-NLS), and proves that there are orbits which start δ-close to them and undergo H^s Sobolev norm explosion, $0 < s < 1$. This implies that the plane waves are strongly Lyapunov unstable in these topologies. Stability results for plane waves in H^s, $s > 1$, on shorter time scales are provided in [FGL14].

1.3. Statement of results. Roughly speaking, we will construct solutions to (2D-NLS) that start very close to the finite-gap tori in appropriate topologies, and exhibit either backward cascade of energy from high to low frequencies, or forward cascade of energy from low to high frequencies. In the former case, the solutions that exhibit backward cascade start in an arbitrarily small vicinity of a finite-gap torus in Sobolev spaces $H^s(\mathbb{T}^2)$ with $0 < s < 1$, but grow to become larger than any pre-assigned factor $K \gg 1$ in the same H^s (higher Sobolev norms H^s with $s > 1$ decrease, but they are large for all times). In the latter case, the solutions that exhibit forward cascade start in an arbitrarily small vicinity of a finite-gap torus in $L^2(\mathbb{T}^2)$, but their H^s Sobolev norm

(for $s > 1$) exhibits a growth by a large multiplicative factor $K \gg 1$ after a large time. We shall comment further on those results after we state the theorems precisely.

To do that, we need to introduce the Birkhoff coordinates for equation 1D-NLS. Grébert and Kappeler showed in [GK14a] that there exists a globally defined map, called the Birkhoff map, such that $\forall s \geq 0$

$$(1.6) \qquad \begin{aligned} \Phi : H^s(\mathbb{T}) &\longrightarrow h^s(\mathbb{Z}) \times h^s(\mathbb{Z}) \\ q &\longmapsto (z_m, \overline{z}_m)_{m \in \mathbb{Z}}, \end{aligned}$$

such that equation (1D-NLS) is transformed in the new coordinates $(z_m, \overline{z}_m)_{m \in \mathbb{Z}} = \Phi(q)$ to:

$$(1.7) \qquad \mathrm{i}\dot{z}_m = \alpha_m(I)z_m$$

where $I = (I_m)_{m \in \mathbb{Z}}$ and $I_m = |z_m|^2$ are the actions, which are conserved in time (since $\alpha_m(I) \in \mathbb{R}$). Therefore in these coordinates, called Birkhoff coordinates, equation (1D-NLS) becomes a chain of nonlinear harmonic oscillators. Of course the solutions of (1.7) live on finite and infinite dimensional tori with periodic, quasiperiodic or almost periodic dynamics, depending on how many of the actions I_m (which are constant!) are nonzero and on the properties of rational dependence of the frequencies. Hence the 1D-NLS equation admits a family of finite dimensional integrable subsystems, denoted by $\mathcal{G}_\mathcal{S}$, where $\mathcal{S}$ runs through the nonempty, finite subsets of $\mathbb{Z}$. $\mathcal{G}_\mathcal{S}$ is contained in $\cap_{n \geq 0} H^n(\mathbb{T}, \mathbb{C})$ and its elements are called $\mathcal{S}$-gap solutions. In particular it follows that $\mathcal{G}_\mathcal{S}$ is foliated by $\mathcal{T}_\mathcal{S}^I := \Phi^{-1}(\mathbb{T}_\mathcal{S}^I)$ where

$$\mathbb{T}_\mathcal{S}^I := \left\{ z \in \ell^2 : |z_m|^2 = I_m, \text{ for } m \in \mathcal{S}, \quad |z_m|^2 = 0 \text{ for } m \notin \mathcal{S} \right\}$$

is a torus of dimension $|\mathcal{S}|$ parametrized by the action variables $I = (I_j)_{j \in \mathcal{S}} \in \mathbb{R}_{>0}^\mathcal{S}$. This torus, as an invariant object of equation 1D-NLS, is stable for all times in the sense of Lyapunov.

We will abuse notation, and identify $H^s(\mathbb{T})$ with the closed subspace of $H^s(\mathbb{T}^2)$ of functions depending only on the x variable. Consequently, $\mathcal{T}_\mathcal{S}^I$ is a closed torus of $H^s(\mathbb{T}) \subset H^s(\mathbb{T}^2)$ which is invariant for the (2D-NLS) dynamics.

The main result of this paper will show the instability (in the sense of Lyapunov) of many of these invariant objects under the dynamics of (2D-NLS). Roughly speaking, we show that, under certain assumptions on the choices of modes $\mathcal{S}$ and actions I, these tori are unstable in the $H^s(\mathbb{T}^2)$ topology for $s \in (0, 1)$. Even more, there exist orbits which start arbitrarily close to these tori and undergo an arbitrarily large H^s-norm explosion. In order to state our result precisely, we introduce the definition of generic set $\mathcal{S}$:

Definition 1.1 (L−genericity). *Given* $\mathsf{L} \in \mathbb{N}$, *we say that* $\mathcal{S} = \{\mathsf{m}_1, \ldots, \mathsf{m}_\mathsf{d}\}$ *is* L-generic *if it satisfies the condition*

$$(1.8) \qquad \sum_{i=1}^\mathsf{d} \ell_i \mathsf{m}_i \neq 0 \qquad \forall \ell \in \mathbb{Z}^\mathsf{d} : \quad 0 < |\ell| := \sum_{i=1}^\mathsf{d} |\ell_i| \leq \mathsf{L},$$

where d *is the cardinality of* $\mathcal{S}$.

Our main result is the following:

Theorem 1.2. *Fix a positive integer* $\mathsf{d} \geq 2$ *and a sufficiently large* $\mathsf{L} \in \mathbb{N}$. *Assume that* $\mathcal{S}_0$ *has cardinality* d *and is* L-generic. *Then there exists* $\varepsilon_* > 0$ *such that for any* $\varepsilon \in (0, \varepsilon_*)$ *there exists a positive measure Cantor-like set* $\mathcal{I} \subset (\varepsilon/2, \varepsilon)^\mathsf{d}$ *such that the torus* $\mathcal{T}_{\mathcal{S}_0}^I$ *with* $I = (I_j)_{j \in \mathcal{S}_0} \in \mathcal{I}$ *has the following properties:*

(1) **Long-time instability of** $\mathcal{S}_0$**-gap solutions in** $H^s(\mathbb{T}^2, \mathbb{C})$ **for** $0 < s < 1$:
   For any $s \in (0, 1)$, $\delta > 0$ *small enough, and for any* $K > 1$ *sufficiently large, there exists a smooth solution* $u(t)$ *of* (2D-NLS), $u : \mathbb{R} \to \cap_{n \geq 0} H^n(\mathbb{T}^2, \mathbb{C})$, *and a time* $0 < T \leq e^{\left(\frac{K}{\delta}\right)^\beta}$ *so that*

$$\mathrm{dist}\left(u(0), \mathcal{T}_{\mathcal{S}_0}^I\right)_{H^s(\mathbb{T}^2)} \leq \delta \quad \text{and} \quad \|u(T)\|_{H^s(\mathbb{T}^2)} \geq K.$$

   Here the exponent $\beta > 1$ *can be chosen independently of* K, δ. *In particular, the* $\mathcal{S}_0$*-gap solutions in* $\mathcal{T}_{\mathcal{S}_0}^I$ *are Lyapunov (orbitally) unstable.*

(2) **Long-time instability of $\mathcal{S}_0$-gap solutions in $H^s(\mathbb{T}^2, \mathbb{C})$ for $s > 1$:**

For any $s > 1$ and for any $K > 1$ sufficiently large, there exists a smooth solution $u : \mathbb{R} \to \cap_{n \geq 0} H^n(\mathbb{T}^2, \mathbb{C})$ of (2D-NLS) *and a time $0 < T \leq e^{K^\sigma}$ such that*

$$\operatorname{dist}\big(u(0), \mathcal{T}_{\mathcal{S}_0}^I\big)_{L^2(\mathbb{T}^2)} \leq K^{-\sigma'} \quad \text{and} \quad \|u(T)\|_{H^s(\mathbb{T}^2)} \geq K \|u(0)\|_{H^s(\mathbb{T}^2)}.$$

Here $\sigma, \sigma' > 0$ depend on s, but not on K. Note that $\operatorname{dist}\big(u(0), \mathcal{T}_{\mathcal{S}_0}^I\big)_{H^s(\mathbb{T}^2)}$ *might not be small.*

1.4. Comments and remarks on Theorem 1.2:

(1) The relative measure of the set $\mathcal{I}$ of admissible actions can be taken as close to 1 as desired. Indeed, by taking smaller ε_*, one has that the relative measure satisfies

$$|1 - \operatorname{Meas}(\mathcal{I})| \leq C \varepsilon_*^\kappa$$

for some constant $C > 0$ and $0 < \kappa < 1$ independent of $\varepsilon_* > 0$. The genericity condition on the set $\mathcal{S}_0$ and the actions $(I_{\mathfrak{m}})_{\mathfrak{m} \in \mathcal{S}_0} \in \mathcal{I}$ ensure that the *linearized* dynamics around the resulting torus $\mathcal{T}_{\mathcal{S}_0}^I$ is stable for the perturbations we need to induce the nonlinear instability. In fact, a subset of those tori is even linearly stable for much more general perturbations as we remark below.

(2) *Why does the finite gap solution need to be small?* To prove Theorem 1.2 we need to analyze the linearization of equation (2D-NLS) at the finite gap solution (see Section 4). Roughly speaking, this leads to a Schrödinger equation with a quasi-periodic potential. Luckily, such operators can be *reduced* to constant coefficients via a KAM scheme. This is known as *reducibility theory* which allows one to construct a change of variables that casts the linearized operator into an essentially constant coefficient diagonal one. This KAM scheme was carried out in [MP18], and requires the quasi-periodic potential, given by the finite gap solution here, to be small for the KAM iteration to converge. That being said, we suspect a similar result to be true for non-small finite gap solutions.

(3) To put the complexity of this result in perspective, it is instructive to compare it with the stability result in [MP18]. In that paper, it is shown that a proper *subset* $\mathcal{I}' \subset \mathcal{I}$ of the tori considered in Theorem 1.2 are Lyapunov stable in H^s, $s > 1$, but for shorter time scales than those considered in this theorem. More precisely, all orbits that are initially δ-close to $\mathcal{T}_{\mathcal{S}_0}^I$ in H^s stay $C\delta$-close for some fixed $C > 0$ for time scales $t \sim \delta^{-2}$. The same stability result (with a completely identical proof) holds if we replace H^s by $\mathcal{F}\ell_1$ norm (functions whose Fourier series is in ℓ^1). In fact, by trivially modifying the proof, one could also prove stability on the δ^{-2} timescale in $\mathcal{F}\ell_1 \cap H^s$ for $0 < s < 1$. What this means is that the solutions in the first part of Theorem 1.2 remains within $C\delta$ of $\mathcal{T}_{\mathcal{S}_0}^I$ up to times $\sim \delta^{-2}$ but can diverge vigorously afterwards at much longer time scales.

It is also worth mentioning that the complementary subset $\mathcal{I} \setminus \mathcal{I}'$ has a positive measure subset where tori are linearly unstable since they possess a finite set of modes that exhibit hyperbolic behavior. In principle, hyperbolic directions are good for instability, but they are not useful for our purposes since they live at very low frequencies, and hence cannot be used (at least not by themselves alone) to produce a substantial growth of Sobolev norms. We avoid dealing with these linearly unstable directions by restricting our solution to an invariant subspace on which these modes are at rest[3].

(4) The growth in part (1) of the theorem is the result of the so-called *inverse cascade of mass* from high frequencies towards smaller ones, whereas the growth for $s > 1$ in part (2) is the result of *forward cascade on kinetic energy* from low to high frequencies. Both phenomena are predicted by the physical theory of *wave turbulence* but their rigorous justification is highly nontrivial from a mathematical viewpoint as we discussed earlier. For part (1), initially the mass of the perturbation is concentrated on the "high-frequency set" Λ_{03} in Theorem 7.3), but becomes concentrated on the "low-frequency set" $\Lambda_{0\mathfrak{g}-1}$ at time T. This leads to the inflation of the H^s norm for $0 < s < 1$ (cf. (7.5)) whereas the Sobolev norms for $s > 1$ actually contract. In contrast, in part (2), the initial kinetic energy of the perturbation is concentrated on the set $\Lambda_{0\mathfrak{g}-1}$ and ends up being concentrated on the "high frequency set" Λ_{03}, which

---

[3]We expect that such hyperbolic directions should imply a transverse instability result similar to the one obtained by Rousset and Tzvetkov [RT11, RT12] for solitary waves.

yields the growth of Sobolev norms for $s > 1$. It is here that the dependence of the solution on s starts to make a difference in the proof (cf. Sections 7 and 8).

(5) It is expected that a similar statement to the first part of Theorem 1.2 is also true for $s > 1$. This would be a stronger instability compared to that in the second part (for which the initial perturbation is small in L^2 but not in H^s). Nevertheless, this case cannot be tackled with the techniques considered in this paper. Indeed, one of the key points in the proof is to perform a (partial) Birkhoff normal form up to order 4 around the finite gap solution. The terms which lead to the instabilities in Theorem 1.2 are quasi-resonant instead of being completely resonant. Working in the H^s topology with $s \in (0, 1)$, such terms can be considered completely resonant with little error on the timescales where instability happens. However, this cannot be done for $s > 1$, for which one might be able to eliminate those terms by a higher order normal form ($s > 1$ gives a stronger topology and can thus handle worse small divisors). This would mean that one needs other resonant terms to achieve growth of Sobolev norms. The same difficulties were encountered in [Han14] to prove the instability of the plane waves of (2D-NLS).

(6) The first part of the result of Theorem 1.2 can be interpreted as a *strong* form of Lyapunov instability in H^s norm ($0 < s < 1$) of the tori $\mathcal{T}_{\mathcal{S}_0}^I$ where the $\mathcal{S}_0$-gap solutions are supported. Indeed, for an invariant subset X of the phase space, being Lyapunov stable means that for all $\epsilon > 0$ there exists $\delta > 0$ such that all solutions that are δ-close to X at time $t = 0$ stay ϵ-close to X for all times. Thus, Lyapunov instability of X means that there exists $K > 0$ such that for all $\delta > 0$ there exist a solution $u(t)$ and a time T such that $\mathrm{dist}(u(0), X) < \delta$ and $\mathrm{dist}(u(T), X) > K$. In the first part of Theorem 1.2, we prove that for $X = \mathcal{T}_{\mathcal{S}_0}^I$ such an instability property holds true for all $\delta > 0$ and **for all** $K > 0$ in H^s norm (with $s \in (0, 1)$). Thus, a stronger form of instability holds: one can start as close to $\mathcal{T}_{\mathcal{S}_0}^I$ as desired but still end up *as far as desired* from $\mathcal{T}_{\mathcal{S}_0}^I$ after some time $T = T(K, \delta) > 0$.

(7) For finite dimensional Hamiltonian dynamical systems, proving Lyapunov instability for quasi-periodic Diophantine elliptic (or maximal dimensional Lagrangian) tori is an extremely difficult task. Actually all the obtained results [CZ13, GK14b] deal with C^r or C^∞ Hamiltonians, and not a single example of such instability is known for analytic Hamiltonian systems. In fact, there are no results of instabilities in the vicinity of non-resonant elliptic critical points or periodic orbits for analytic Hamiltonian systems (see [LCD83, Dou88, KMV04] for results on the C^∞ topology). The present paper proves the existence of unstable Diophantine elliptic tori in an analytic infinite dimensional Hamiltonian system. Obtaining such instabilities in infinite dimensions is, in some sense, easier: having infinite dimensions gives "more room" for instabilities.

(8) It is well known that many Hamiltonian PDEs possess quasiperiodic invariant tori [Way90, Pös96, KP96, Bou98, BB13, EK10, GXY11, BB11, Wan16, PX13, BCP15, PP12, PP15, BBHM18]. Most of these tori are normally elliptic and thus linearly stable. It is widely expected that the behavior given by Theorem 1.2 also arises in the neighborhoods of (many of) those tori. Nevertheless, it is not clear how to apply the techniques of the present paper to these settings.

1.5. Scheme of the proof. Let us explain the main steps to prove Theorem 1.2:

(1) Analysis of the 1-dimensional cubic Schrödinger equation. We express the 1-dimensional cubic NLS in terms of the Birkhoff coordinates. We need a quite precise knowledge of the Birkhoff map (see Theorem 3.1). In particular, we need that it "behaves well" in ℓ^1. This is done in the paper [Mas18b] and summarized in Section 3. In Birkhoff coordinates, the finite gap solutions are supported in a finite set of variables. We use such coordinates to express the Hamiltonian (1.1) in a more convenient way.

(2) Reducibility of the 2-dimensional cubic NLS around a finite gap solution. We reduce the linearization of the vector field around the finite gap solutions to a constant coefficients diagonal vector field. This is done in [MP18] and explained in Section 4. In Theorem 4.3 we give the conditions to achieve full reducibility. In effect, this transforms the linearized operator around the finite gap into a constant coefficient diagonal (in Fourier space) operator, with eigenvalues $\{\Omega_{\vec{j}}\}_{\vec{j} \in \mathbb{Z}^2 \setminus \mathcal{S}_0}$. We give the asymptotics of these eigenvalues in Theorem 4.4, which roughly speaking look like

$$(1.9) \qquad\qquad \Omega_{\vec{j}} = |\vec{j}|^2 + O(J^{-2})$$

for frequencies $\vec{\jmath} = (m, n)$ satisfying $|m|, |n| \sim J$. This seemingly harmless $O(J^{-2})$ correction to the unperturbed Laplacian eigenvalues is sharp and will be responsible for the restriction to $s \in (0, 1)$ in the first part of Theorem 1.2 as we shall explain below.

(3) Degree three Birkhoff normal form around the finite gap solution. This is done in [MP18], but we shall need more precise information from this normal form that will be crucial for Steps 5 and 6 below. This is done in 5 (see Theorem 5.2).

(4) Partial normal form of degree four. We remove all degree four monomials which are not (too close to) resonant. This is done in Section 6, and leaves us with a Hamiltonian with (close to) resonant degree-four terms plus a higher-degree part which will be treated as a remainder in our construction.

(5) We follow the paradigm set forth in [CKS$^+$10, GK15] to construct solutions to the truncated Hamiltonian consisting of the (close to) resonant degree-four terms isolated above, and then afterwards to the full Hamiltonian by an approximation argument. This construction will be done at frequencies $\vec{\jmath} = (m, n)$ such that $|m|, |n| \sim J$ with J very large, and for which the dynamics is effectively given by the following system of ODE

$$
\begin{cases}
\mathrm{i}\dot{a}_{\vec{\jmath}} &= -|a_{\vec{\jmath}}|^2 a_{\vec{\jmath}} + \sum_{\mathcal{R}(\vec{\jmath})} a_{\vec{\jmath}_1} \overline{a_{\vec{\jmath}_2}} a_{\vec{\jmath}_3} e^{\mathrm{i}\Gamma t} \\
\mathcal{R}(\vec{\jmath}) &:= \{(\vec{\jmath}_1, \vec{\jmath}_2, \vec{\jmath}_3) \in \mathbb{Z}^2 \setminus \mathcal{S}_0 : \vec{\jmath}_1, \vec{\jmath}_3 \neq \vec{\jmath}, \quad \vec{\jmath}_1 - \vec{\jmath}_2 + \vec{\jmath}_3 = \vec{\jmath}, \quad |\vec{\jmath}_1|^2 - |\vec{\jmath}_2|^2 + |\vec{\jmath}_3|^2 = |\vec{\jmath}|^2\} \\
\Gamma &:= \Omega_{\vec{\jmath}_1} - \Omega_{\vec{\jmath}_2} + \Omega_{\vec{\jmath}_3} - \Omega_{\vec{\jmath}}.
\end{cases}
$$

We remark that the conditions of the set $\mathcal{R}(\vec{\jmath})$ are essentially equivalent to saying that $(\vec{\jmath}_1, \vec{\jmath}_2, \vec{\jmath}_3, \vec{\jmath})$ form a rectangle in $\mathbb{Z}^2$. Also note that by the asymptotics of $\Omega_{\vec{\jmath}}$ mentioned above in (1.9), one obtains that $\Gamma = O(J^{-2})$ if all the frequencies involved are in $\mathcal{R}(\vec{\jmath})$ and satisfy $|m|, |n| \sim J$. The idea now is to reduce this system into a finite dimensional system called the "Toy Model" which is tractable enough for us to construct a solution that cascades energy. An obstruction to this plan is presented by the presence of the oscillating factor $e^{\mathrm{i}\Gamma t}$ for which Γ is not zero (in contrast to [CKS$^+$10]) but rather $O(J^{-2})$. The only way to proceed with this reduction is to approximate $e^{\mathrm{i}\Gamma t} \sim 1$ which is only possible provided $J^{-2}T \ll 1$. The solution coming from the Toy Model is supported on a finite number of modes $\vec{\jmath} \in \mathbb{Z}^2 \setminus \mathcal{S}_0$ satisfying $|j| \sim J$, and the time it takes for the energy to diffuse across its modes is $T \sim O(\nu^{-2})$ where ν is the characteristic size of the modes in ℓ^1 norm. Requiring the solution to be initially close in H^s to the finite gap would necessitate that $\nu J^s \lesssim \delta$ which gives that $T \gtrsim_\delta J^{-2s}$, and hence the condition $J^{-2}T \ll 1$ translates into the condition $s < 1$. This explains the restriction to $s < 1$ in the first part of Theorem 1.2. If we only require our solutions to be close to the finite gap in L^2, then no such restriction on ν is needed, and hence there is no restriction on s beyond being $s > 0$ and $s \neq 1$, which is the second part of the theorem.

This analysis is done in Section 7 and 8. In the former, we perform the reduction to the effective degree 4 Hamiltonian taking into account all the changes of variables performed in the previous sections; while in Section 8 we perform the above approximation argument allowing to shadow the Toy Model solution mentioned above with a solution of (2D-NLS) exhibiting the needed norm growth, thus completing the proof of Theorem 1.2.

In Appendix B we give a list of notations and parameters used throughout the paper.

2. Notation and functional setting

2.1. **Notation.** For a complex number z, it is often convenient to use the notation

$$z^\sigma = \begin{cases} z & \text{if } \sigma = +1, \\ \bar{z} & \text{if } \sigma = -1. \end{cases}$$

For any subset $\Gamma \subset \mathbb{Z}^2$, we denote by $h^s(\Gamma)$ the set of sequences $(a_{\vec{\jmath}})_{\vec{\jmath} \in \Gamma}$ with norm

$$\|a\|_{h^s(\Gamma)} = \left(\sum_{\vec{\jmath} \in \Gamma} \langle \vec{\jmath} \rangle^{2s} |a_{\vec{\jmath}}|^2 \right)^{1/2} < \infty.$$

Our phase space will be obtained by an appropriate linearization around the finite gap solution with $\mathtt{d}$ frequencies/actions. For a finite set $\mathcal{S}_0 \subset \mathbb{Z} \times \{0\}$ of $\mathtt{d}$ elements, we consider the phase space $\mathcal{X} = (\mathbb{C}^{\mathtt{d}} \times \mathbb{T}^{\mathtt{d}}) \times \ell^1(\mathbb{Z}^2 \setminus \mathcal{S}_0) \times \ell^1(\mathbb{Z}^2 \setminus \mathcal{S}_0)$. The first part $(\mathbb{C}^{\mathtt{d}} \times \mathbb{T}^{\mathtt{d}})$ corresponds to the finite-gap sites in action angle coordinates, whereas $\ell^1(\mathbb{Z}^2 \setminus \mathcal{S}_0) \times \ell^1(\mathbb{Z}^2 \setminus \mathcal{S}_0)$ corresponds to the remaining orthogonal sites in frequency space. We shall often denote the ℓ^1 norm by $\| \cdot \|_1$. We shall denote variables on $\mathcal{X}$ by

$$\mathcal{X} \ni (\mathcal{Y}, \theta, \mathbf{a}) : \qquad \mathcal{Y} \in \mathbb{C}^{\mathtt{d}}, \quad \theta \in \mathbb{T}^{\mathtt{d}}, \quad \mathbf{a} = (a, \bar{a}) \in \ell^1(\mathbb{Z}^2 \setminus \mathcal{S}_0) \times \ell^1(\mathbb{Z}^2 \setminus \mathcal{S}_0).$$

We shall use multi-index notation to write monomials like $\mathcal{Y}^l$ and $\mathfrak{m}_{\alpha,\beta} = a^\alpha \bar{a}^\beta$ where $l \in \mathbb{N}^{\mathtt{d}}$ and $\alpha, \beta \in (\mathbb{N})^{\mathbb{Z}^2 \setminus \mathcal{S}_0}$. Often times, we will abuse notation, and simply use the notation $\mathbf{a} \in \ell^1$ to mean $\mathbf{a} = (a, \bar{a}) \in \ell^1(\mathbb{Z}^2 \setminus \mathcal{S}_0) \times \ell^1(\mathbb{Z}^2 \setminus \mathcal{S}_0)$, and $\|\mathbf{a}\|_1 = \|a\|_{\ell^1(\mathbb{Z}^2 \setminus \mathcal{S}_0)}$.

Definition 2.1. *For a monomial of the form $e^{\mathrm{i}\ell \cdot \theta} \mathcal{Y}^l \mathfrak{m}_{\alpha,\beta}$, we define its degree to be $2|l| + |\alpha| + |\beta| - 2$, where the modulus of a multi-index is given by its ℓ^1 norm.*

2.2. **Regular Hamiltonians.** Given a Hamiltonian function $F(\mathcal{Y}, \theta, \mathbf{a})$ on the phase space $\mathcal{X}$, we associate to it the Hamiltonian vector field

$$X_F := \{ -\partial_\theta F, \partial_{\mathcal{Y}} F, \quad -\mathrm{i}\partial_{\bar{a}} F, \mathrm{i}\partial_a F \},$$

where we have used the standard complex notation to denote the Fréchet derivatives of F with respect to the variable $\mathbf{a} \in \ell^1$.

We will often need to complexify the variable $\theta \in \mathbb{T}^{\mathtt{d}}$ into the domain

$$\mathbb{T}^{\mathtt{d}}_\rho := \{ \theta \in \mathbb{C}^{\mathtt{d}} : \mathrm{Re}(\theta) \in \mathbb{T}^{\mathtt{d}}, \quad |\mathrm{Im}(\theta)| \le \rho \}$$

and consider vector fields which are functions from

$$\mathbb{C}^{\mathtt{d}} \times \mathbb{T}^{\mathtt{d}}_\rho \times \ell^1 \to \mathbb{C}^{\mathtt{d}} \times \mathbb{C}^{\mathtt{d}} \times \ell^1 : (\mathcal{Y}, \theta, \mathbf{a}) \to (X^{(\mathcal{Y})}, X^{(\theta)}, X^{(a)}, X^{(\bar{a})})$$

which are analytic in $\mathcal{Y}, \theta, \mathbf{a}$. Our vector fields will be defined on the domain

$$(2.1) \qquad D(\rho, r) := \mathbb{T}^{\mathtt{d}}_\rho \times D(r) \quad \text{where} \quad D(r) := \{ |\mathcal{Y}| \le r^2, \quad \|\mathbf{a}\|_1 \le r \}.$$

On the vector field, we use as norm

$$\|X\|_r := |X^{(\theta)}| + \frac{|X^{(\mathcal{Y})}|}{r^2} + \frac{\|X^{(a)}\|_1}{r} + \frac{\|X^{(\bar{a})}\|_1}{r}.$$

All Hamiltonians F considered in this article are analytic, real valued and can be expanded in Taylor Fourier series which are well defined and pointwise absolutely convergent

$$(2.2) \qquad F(\mathcal{Y}, \theta, \mathbf{a}) = \sum_{\alpha, \beta \in \mathbb{N}^{\mathbb{Z}^2 \setminus \mathcal{S}_0}, \ell \in \mathbb{Z}^{\mathtt{d}}, l \in \mathbb{N}^{\mathtt{d}}} F_{\alpha, \beta, l, \ell} \, e^{\mathrm{i}\ell \cdot \theta} \mathcal{Y}^l \mathfrak{m}_{\alpha, \beta}.$$

Correspondingly we expand vector fields in Taylor Fourier series (again well defined and pointwise absolutely convergent)

$$X^{(v)}(\mathcal{Y}, \theta, \mathbf{a}) = \sum_{\alpha, \beta \in \mathbb{N}^{\mathbb{Z}^2 \setminus \mathcal{S}_0}, \ell \in \mathbb{Z}^{\mathtt{d}}, l \in \mathbb{N}^{\mathtt{d}}} X^{(v)}_{\alpha, \beta, l, \ell} \, e^{\mathrm{i}\ell \cdot \theta} \mathcal{Y}^l \mathfrak{m}_{\alpha, \beta},$$

where v denotes the components $\theta_i, \mathcal{Y}_i$ for $1 \le i \le \mathtt{d}$ or $a_{\vec{\jmath}}, \bar{a}_{\vec{\jmath}}$ for $\vec{\jmath} \in \mathbb{Z}^2 \setminus \mathcal{S}_0$.

To a vector field we associate its majorant

$$\underline{X}_\rho^{(v)}[\mathcal{Y}, \mathbf{a}] := \sum_{\ell \in \mathbb{Z}^{\mathtt{d}}, l \in \mathbb{N}^{\mathtt{d}}, \alpha, \beta \in \mathbb{N}^{\mathbb{Z}^2}} |X_{\alpha,\beta,\ell}^{(v)}| \, e^{\rho |\ell|} \, \mathcal{Y}^l \, \mathfrak{m}_{\alpha,\beta}$$

and require that this is an analytic map on $D(r)$. Such a vector field is called *majorant analytic*. Since Hamiltonian functions are defined modulo constants, we give the following definition of the *norm* of F:

$$|F|_{\rho,r} := \sup_{(\mathcal{Y},\mathbf{a}) \in D(r)} \left\| (X_F)_\rho \right\|_r .$$

Note that the norm $|\cdot|_{\rho,r}$ controls the $|\cdot|_{\rho',r'}$ whenever $\rho' < \rho$, $r' < r$.

Finally, we will also consider Hamiltonians $F(\lambda; \theta, a, \bar{a}) \equiv F(\lambda)$ depending on an external parameter $\lambda \in \mathcal{O} \subset \mathbb{R}^{\mathtt{d}}$. For those, we define the *inhomogeneous Lipschitz* norm:

$$|F|_{\rho,r}^{\mathcal{O}} := \sup_{\lambda \in \mathcal{O}} |F(\lambda)|_{\rho,r} + \sup_{\lambda_1 \neq \lambda_2 \in \mathcal{O}} \frac{|F(\lambda_1) - F(\lambda_2)|_{\rho,r}}{|\lambda_1 - \lambda_2|} .$$

2.3. Commutation rules.

Given two Hamiltonians F and G, we define their Poisson bracket as $\{F, G\} := \mathrm{d}F(X_G)$; in coordinates

$$\{F, G\} = -\partial_{\mathcal{Y}} F \cdot \partial_\theta G + \partial_\theta F \cdot \partial_{\mathcal{Y}} G + \mathrm{i} \left(\sum_{\vec{j} \in \mathbb{Z}^2 \setminus \mathcal{S}_0} \partial_{\bar{a}_{\vec{j}}} F \partial_{a_{\vec{j}}} G - \partial_{a_{\vec{j}}} F \partial_{\bar{a}_{\vec{j}}} G \right) .$$

Given $\alpha, \beta \in \mathbb{N}^{\mathbb{Z}^2 \setminus \mathcal{S}_0}$ we denote $\mathfrak{m}_{\alpha,\beta} := a^\alpha \bar{a}^\beta$. To the monomial $e^{\mathrm{i}\ell \cdot \theta} \mathcal{Y}^l \mathfrak{m}_{\alpha,\beta}$ with $\ell \in \mathbb{Z}^{\mathtt{d}}, l \in \mathbb{N}^{\mathtt{d}}$ we associate various numbers. We denote by

$$(2.3) \qquad \eta(\alpha, \beta) := \sum_{\vec{j} \in \mathbb{Z}^2 \setminus \mathcal{S}_0} (\alpha_{\vec{j}} - \beta_{\vec{j}}) , \qquad \eta(\ell) := \sum_{i=1}^{\mathtt{d}} \ell_i .$$

We also associate to $e^{\mathrm{i}\ell \cdot \theta} \mathcal{Y}^l \mathfrak{m}_{\alpha,\beta}$ the quantities $\pi(\alpha, \beta) = (\pi_x, \pi_y)$ and $\pi(\ell)$ defined by

$$(2.4) \qquad \pi(\alpha, \beta) = \begin{bmatrix} \pi_x(\alpha, \beta) \\ \pi_y(\alpha, \beta) \end{bmatrix} = \sum_{\vec{j} = (m,n) \in \mathbb{Z}^2 \setminus \mathcal{S}_0} \begin{bmatrix} m \\ n \end{bmatrix} (\alpha_{\vec{j}} - \beta_{\vec{j}}) , \qquad \pi(\ell) = \sum_{i=1}^{\mathtt{d}} \mathtt{m}_i \ell_i .$$

The above quantities are associated with the following mass $\mathcal{M}$ and momentum $\mathcal{P} = (\mathcal{P}_x, \mathcal{P}_y)$ functionals given by

$$(2.5) \qquad
\begin{aligned}
\mathcal{M} &:= \sum_{i=1}^{\mathtt{d}} \mathcal{Y}_i + \sum_{\vec{j} \in \mathbb{Z}^2 \setminus \mathcal{S}_0} |a_{\vec{j}}|^2 \\
\mathcal{P}_x &:= \sum_{i=1}^{\mathtt{d}} \mathtt{m}_i \mathcal{Y}_i + \sum_{(m,n) \in \mathbb{Z}^2 \setminus \mathcal{S}_0} m \, |a_{(m,n)}|^2 \\
\mathcal{P}_y &:= \sum_{(m,n) \in \mathbb{Z}^2 \setminus \mathcal{S}_0} n |a_{(m,n)}|^2
\end{aligned}$$

via the following commutation rules: given a monomial $e^{\mathrm{i}\ell \cdot \theta} \mathcal{Y}^l \mathfrak{m}_{\alpha,\beta}$

$$\{\mathcal{M}, e^{\mathrm{i}\ell \cdot \theta} \mathcal{Y}^l \mathfrak{m}_{\alpha,\beta}\} = \mathrm{i}(\eta(\alpha, \beta) + \eta(\ell)) e^{\mathrm{i}\ell \cdot \theta} \mathcal{Y}^l \mathfrak{m}_{\alpha,\beta}$$

$$\{\mathcal{P}_x, e^{\mathrm{i}\ell \cdot \theta} \mathcal{Y}^l \mathfrak{m}_{\alpha,\beta}\} = \mathrm{i}(\pi_x(\alpha, \beta) + \pi(\ell)) e^{\mathrm{i}\ell \cdot \theta} \mathcal{Y}^l \mathfrak{m}_{\alpha,\beta}$$

$$\{\mathcal{P}_y, e^{\mathrm{i}\ell \cdot \theta} \mathcal{Y}^l \mathfrak{m}_{\alpha,\beta}\} = \mathrm{i} \, \pi_y(\alpha, \beta) \, e^{\mathrm{i}\ell \cdot \theta} \mathcal{Y}^l \mathfrak{m}_{\alpha,\beta}$$

Remark 2.2. *An analytic hamiltonian function $\mathcal{F}$ (expanded as in (2.2)) commutes with the mass $\mathcal{M}$ and the momentum $\mathcal{P}$ if and only if the following selection rules on its coefficients hold:*

$$\begin{aligned}
\{\mathcal{F}, \mathcal{M}\} = 0 &\quad\Leftrightarrow\quad \mathcal{F}_{\alpha,\beta,l,\ell}\left(\eta(\alpha,\beta) + \eta(\ell)\right) = 0 \\
\{\mathcal{F}, \mathcal{P}_x\} = 0 &\quad\Leftrightarrow\quad \mathcal{F}_{\alpha,\beta,l,\ell}\left(\pi_x(\alpha,\beta) + \pi(\ell)\right) = 0 \\
\{\mathcal{F}, \mathcal{P}_y\} = 0 &\quad\Leftrightarrow\quad \mathcal{F}_{\alpha,\beta,l,\ell}\left(\pi_y(\alpha,\beta)\right) = 0
\end{aligned}$$

where $\eta(\alpha,\beta), \eta(\ell)$ are defined in (2.3) and $\pi(\alpha,\beta), \pi(\ell)$ are defined in (2.4).

Definition 2.3. *We will denote by $\mathcal{A}_{\rho,r}$ the set of all real-valued Hamiltonians of the form (2.2) with finite $|\cdot|_{\rho,r}$ norm and which Poisson commute with $\mathcal{M}, \mathcal{P}$. Given a compact set $\mathcal{O} \subset \mathbb{R}^{\mathrm{d}}$, we denote by $\mathcal{A}_{\rho,r}^{\mathcal{O}}$ the Banach space of Lipschitz maps $\mathcal{O} \to \mathcal{A}_{\rho,r}$ with the norm $|\cdot|_{\rho,r}^{\mathcal{O}}$.*

From now on, all our Hamiltonians will belong to some set $\mathcal{A}_{\rho,r}$ for some $\rho, r > 0$.

3. Adapted variables and Hamiltonian formulation

3.1. Fourier expansion and phase shift.
Let us start by expanding u in Fourier coefficients

$$u(x,y,t) = \sum_{\vec{j}=(m,n)\in\mathbb{Z}^2} u_{\vec{j}}(t)\, e^{\mathrm{i}(mx+ny)}.$$

Then, the Hamiltonian H_0 introduced in (1.1) can be written as

$$H_0(u) = \sum_{\vec{j}\in\mathbb{Z}^2} |\vec{j}|^2 |u_{\vec{j}}|^2 + \frac{1}{2} \sum_{\substack{\vec{j}_i \in \mathbb{Z}^2 \\ \vec{j}_1 - \vec{j}_2 + \vec{j}_3 - \vec{j}_4 = 0}} u_{\vec{j}_1} \bar{u}_{\vec{j}_2} u_{\vec{j}_3} \bar{u}_{\vec{j}_4}$$

$$= \sum_{\vec{j}\in\mathbb{Z}^2} |\vec{j}|^2 |u_{\vec{j}}|^2 - \frac{1}{2} \sum_{\vec{j}\in\mathbb{Z}^2} |u_{\vec{j}}|^4 + 2 \overbrace{\left(\sum_{\vec{j}\in\mathbb{Z}^2} |u_{\vec{j}}|^2\right)^2}^{M(u)^2} + \frac{1}{2} \sum_{\substack{\vec{j}_i \in \mathbb{Z}^2 \\ \vec{j}_1 - \vec{j}_2 + \vec{j}_3 - \vec{j}_4 = 0}}^{\star} u_{\vec{j}_1} \bar{u}_{\vec{j}_2} u_{\vec{j}_3} \bar{u}_{\vec{j}_4}$$

where the $\sum^{\star}$ means the sum over the quadruples $\vec{j}_i$ such that $\{\vec{j}_1, \vec{j}_3\} \neq \{\vec{j}_2, \vec{j}_4\}$.

Since the mass $M(u)$ in (1.2) is a constant of motion, we make a trivial phase shift and consider an equivalent Hamiltonian $H(u) = H_0(u) - M(u)^2$,

$$(3.1) \qquad H(u) = \int_{\mathbb{T}^2} |\nabla u(x,y)|^2 \, \mathrm{d}x\,\mathrm{d}y + \frac{1}{2} \int_{\mathbb{T}^2} |u(x,y)|^4 \, \mathrm{d}x\,\mathrm{d}y - M(u)^2$$

corresponding to the Hamilton equation

$$(3.2) \qquad \mathrm{i}\partial_t u = -\Delta u + |u|^2 u - 2M(u)u \,, \qquad (x,y) \in \mathbb{T}^2 \,.$$

Clearly the solutions of (3.2) differ from the solutions of (2D-NLS) only by a phase shift[4]. Then,

$$(3.3) \qquad H(u) = \sum_{\vec{j}\in\mathbb{Z}^2} |\vec{j}|^2 |u_{\vec{j}}|^2 - \frac{1}{2} \sum_{\vec{j}\in\mathbb{Z}^2} |u_{\vec{j}}|^4 + \frac{1}{2} \sum_{\substack{\vec{j}_i \in \mathbb{Z}^2 \\ \vec{j}_1 - \vec{j}_2 + \vec{j}_3 - \vec{j}_4 = 0}}^{\star} u_{\vec{j}_1} \bar{u}_{\vec{j}_2} u_{\vec{j}_3} \bar{u}_{\vec{j}_4}.$$

---

[4]In order to show the equivalence we consider any solution $u(x,t)$ of (3.2) and consider the invertible map

$$u \mapsto v = u\, e^{-2\mathrm{i}M(u)t} \quad \text{with inverse} \quad v \mapsto u = v\, e^{2\mathrm{i}M(v)t}.$$

Then a direct computation shows that v solves 2D-NLS.

3.2. **The Birkhoff map for the 1D cubic NLS.** We devote this section to gathering some properties of the Birkhoff map for the integrable 1D NLS equation. These will be used to write the Hamiltonian (3.3) in a more convenient way. The main reference for this section is [Mas18b]. We shall denote by $B^s(r)$ the ball of radius r and center 0 in the topology of $h^s \equiv h^s(\mathbb{Z})$.

Theorem 3.1. *There exist $r_* > 0$ and a symplectic, real analytic map Φ with $d\Phi(0) = \mathbb{I}$ such that $\forall s \geq 0$ one has the following*

(i) $\Phi : B^s(r_*) \to h^s$. *More precisely, there exists a constant $C > 0$ such that for all $0 \leq r \leq r_*$*

$$\sup_{\|q\|_{h^s} \leq r} \|(\Phi - \mathbb{I})(q)\|_{h^s} \leq C \, r^3 \ .$$

The same estimate holds for $\Phi^{-1} - \mathbb{I}$ or by replacing the space h^s with the space ℓ^1.

(ii) *Moreover, if $q \in h^s$ for $s \geq 1$, Φ introduces local Birkhoff coordinates for (NLS-1d) in h^s as follows: the integrals of motion of (NLS-1d) are real analytic functions of the actions $I_j = |z_j|^2$ where $(z_j)_{j \in \mathbb{Z}} = \Phi(q)$. In particular, the Hamiltonian $H_{\mathrm{NLS1d}}(q) \equiv \int_{\mathbb{T}} |\partial_x q(x)|^2 \, dx - M(q)^2 + \frac{1}{2} \int_{\mathbb{T}} |q(x)|^4 \, dx$, the mass $M(q) := \int_{\mathbb{T}} |q(x)|^2 \, dx$ and the momentum $P(q) := -\int_{\mathbb{T}} \bar{q}(x) i \partial_x q(x) dx$ have the form*

$$\tag{3.4} \left(H_{\mathrm{NLS1d}} \circ \Phi^{-1}\right)(z) \equiv h_{\mathrm{nls1d}}\left((|z_m|^2)_{m \in \mathbb{Z}}\right) = \sum_{m \in \mathbb{Z}} m^2 |z_m|^2 - \frac{1}{2} \sum_{m \in \mathbb{Z}} |z_m|^4 + O(|z|^6) \ ,$$

$$\left(M \circ \Phi^{-1}\right)(z) = \sum_{m \in \mathbb{Z}} |z_m|^2 \ ,$$

$$\left(P \circ \Phi^{-1}\right)(z) = \sum_{m \in \mathbb{Z}} m |z_m|^2 \ .$$

(iii) *Define the (NLS-1d) action-to-frequency map $I \mapsto \alpha^{\mathrm{nls1d}}(I)$, where $\alpha_m^{\mathrm{nls1d}}(I) := \frac{\partial h_{\mathrm{nls1d}}}{\partial I_m}$, $\forall m \in \mathbb{Z}$. Then one has the asymptotic expansion*

$$\tag{3.5} \alpha_m^{\mathrm{nls1d}}(I) = m^2 - I_m + \frac{\varpi_m(I)}{\langle m \rangle}$$

where $\varpi_m(I)$ is at least quadratic in I.

Proof. Item (i) is the main content of [Mas18b], where it is proved that the Birkhoff map is majorant analytic between some Fourier-Lebesgue spaces. Item (ii) is proved in [GK14a]. Item (iii) is Theorem 1.3 of [KST17]. □

Remark 3.2. *Theorem 3.1 implies that all solutions of 1D NLS have Sobolev norms uniformly bounded in time (as it happens for other integrable systems, like KdV and Toda lattice, see e.g. [BM16, KMMT16]). On the contrary, the Szegő equation is an integrable system which exhibits growth of Sobolev norms [GG15].*

3.3. **Adapted variables.** The aim of this section is to write the Hamiltonian (3.1), the mass M (1.2) and the momentum P (1.3) in the local variables around the finite gap solution corresponding to

$$\begin{cases} |z_m|^2 &= I_m, \qquad m \in \mathcal{S}_0 \\ z_m &= 0, \qquad m \in \mathbb{Z} \setminus \mathcal{S}_0. \end{cases}$$

To begin with, we start from the Hamiltonian in Fourier coordinates (3.3), and set

$$q_m := u_{(m,0)} \quad \text{if } m \in \mathbb{Z}, \qquad a_{\vec{j}} = u_{\vec{j}} \quad \text{if } \vec{j} = (m,n) \in \mathbb{Z}^2, \, n \neq 0 \ .$$

We rewrite the Hamiltonian accordingly in increasing degree in a, obtaining

$$H(q,a) = \sum_{m \in \mathbb{Z}} m^2 |q_m|^2 - \frac{1}{2} \sum_{m \in \mathbb{Z}} |q_m|^4 + \frac{1}{2} \sum_{\substack{m_i \in \mathbb{Z} \\ m_1 - m_2 + m_3 - m_4 = 0}}^{\star} q_{m_1} \bar{q}_{m_2} q_{m_3} \bar{q}_{m_4} +$$

$$+ \sum_{\vec{j} \in \mathbb{Z}^2 \setminus \mathbb{Z}} |\vec{j}|^2 |a_{\vec{j}}|^2 + 2 \sum_{\substack{\vec{j}_i = (m_i,n_i), i=3,4, n_i \neq 0 \\ m_1 - m_2 + m_3 - m_4 = 0 \\ n_3 - n_4 = 0}}^{\star} q_{m_1} \bar{q}_{m_2} a_{\vec{j}_3} \bar{a}_{\vec{j}_4} + \operatorname{Re} \sum_{\substack{\vec{j}_i = (m_i,n_i), i=2,4, n_i \neq 0 \\ m_1 - m_2 + m_3 - m_4 = 0 \\ n_2 + n_4 = 0}} \bar{q}_{m_1} a_{\vec{j}_2} \bar{q}_{m_3} a_{\vec{j}_4}$$

$$+ 2 \operatorname{Re} \sum_{\substack{\vec{j}_i = (m_i,n_i), i=2,3,4, n_i \neq 0 \\ m_1 - m_2 + m_3 - m_4 = 0 \\ -n_2 + n_3 - n_4 = 0}} q_{m_1} \bar{a}_{\vec{j}_2} a_{\vec{j}_3} \bar{a}_{\vec{j}_4}$$

$$+ \frac{1}{2} \sum_{\substack{\vec{j}_i = (m_i,n_i), i=1,2,3,4, n_i \neq 0 \\ \vec{j}_1 - \vec{j}_2 + \vec{j}_3 - \vec{j}_4 = 0}}^{\star} a_{\vec{j}_1} \bar{a}_{\vec{j}_2} a_{\vec{j}_3} \bar{a}_{\vec{j}_4} - \frac{1}{2} \sum_{\vec{j} \in \mathbb{Z}^2 \setminus \mathbb{Z}} |a_{\vec{j}}|^4$$

$$=: H_{\mathrm{nls1d}}(q) + H^{\mathrm{II}}(q,a) + H^{\mathrm{III}}(q,a) + H^{\mathrm{IV}}(a).$$

Step 1: First we do the following change of coordinates, which amounts to introducing Birkhoff coordinates on the line $\mathbb{Z} \times \{0\}$. We set

$$\left((z_m)_{m \in \mathbb{Z}}, (a_{\vec{j}})_{\vec{j} \in \mathbb{Z}^2 \setminus \mathbb{Z}} \right) \mapsto \left((q_m)_{m \in \mathbb{Z}}, (a_{\vec{j}})_{\vec{j} \in \mathbb{Z}^2 \setminus \mathbb{Z}} \right)$$

$$(q_m)_{m \in \mathbb{Z}} = \Phi^{-1} \left((z_m)_{m \in \mathbb{Z}} \right), \quad a_{\vec{j}} = u_{\vec{j}}, \quad \vec{j} \in \mathbb{Z}^2 \setminus \mathbb{Z}.$$

In those new coordinates, the Hamiltonian becomes

$$H(z,a) = H_{\mathrm{nls1d}}(\Phi^{-1}(z)) + H^{\mathrm{II}}(\Phi^{-1}(z),a) + H^{\mathrm{III}}(\Phi^{-1}(z),a) + H^{\mathrm{IV}}(a),$$

where

$$H_{\mathrm{nls1d}}(\Phi^{-1}(z)) = h_{\mathrm{nls1d}}((|z_m|^2)_{m \in \mathbb{Z}}).$$

Step 2: Next, we go to action-angle coordinates only on the set $\mathcal{S}_0 = \{\mathtt{m}_1, \ldots, \mathtt{m}_d\} \subset \mathbb{Z} \times \{0\}$ and rename z_m for $m \notin \mathcal{S}_0$ as $a_{(m,0)}$, as follows

$$(\mathcal{Y}_i, \theta_i, a_{\vec{j}}) \underset{\substack{1 \leq i \leq \mathtt{d} \\ \vec{j} \in \mathbb{Z}^2 \setminus \mathcal{S}_0}}{} \mapsto (z_m, a_{\vec{j}})_{m \in \mathbb{Z}, \vec{j} \in \mathbb{Z}^2 \setminus \mathbb{Z}}$$

$$z_{\mathtt{m}_i} = \sqrt{I_{\mathtt{m}_i} + \mathcal{Y}_i}\, e^{\mathrm{i}\theta_i}, \qquad \mathtt{m}_i \in \mathcal{S}_0,$$

$$z_m = a_{(m,0)}, \qquad m \in \mathbb{Z} \setminus \mathcal{S}_0,$$

$$a_{\vec{j}} = a_{\vec{j}}, \qquad \vec{j} \in \mathbb{Z}^2 \setminus \mathbb{Z}.$$

In those coordinates, the Hamiltonian becomes (using (3.4))

$$(3.6) \qquad \mathcal{H}(\mathcal{Y}, \theta, a) = h_{\mathrm{nls1d}}(I_{\mathtt{m}_1} + \mathcal{Y}_1, \ldots, I_{\mathtt{m}_d} + \mathcal{Y}_d, (|a_{(m,0)}|^2)_{m \notin \mathcal{S}_0})$$

$$(3.7) \qquad + H^{\mathrm{II}} \left(\Phi^{-1} \left(\sqrt{I_{\mathtt{m}_1} + \mathcal{Y}_1} e^{\mathrm{i}\theta_1}, \ldots, \sqrt{I_{\mathtt{m}_d} + \mathcal{Y}_d} e^{\mathrm{i}\theta_d}, (a_{(m,0)})_{m \notin \mathcal{S}_0} \right), (a_{(m,n)})_{n \neq 0} \right)$$

$$(3.8) \qquad + H^{\mathrm{III}} \left(\Phi^{-1} \left(\sqrt{I_{\mathtt{m}_1} + \mathcal{Y}_1} e^{\mathrm{i}\theta_1}, \ldots, \sqrt{I_{\mathtt{m}_d} + \mathcal{Y}_d} e^{\mathrm{i}\theta_d}, (a_{(m,0)})_{m \notin \mathcal{S}_0} \right), (a_{(m,n)})_{n \neq 0} \right)$$

$$(3.9) \qquad + H^{\mathrm{IV}} \left((a_{(m,n)})_{n \neq 0} \right).$$

We first remark that $\mathcal{T}^I_{\mathcal{S}_0}$ is described in the $(\mathcal{Y}, \theta, a)$ coordinates by $\mathcal{Y} = 0$, $a = 0$. Furthermore, it is proved in Proposition 4.2 of [MP18] that a neighbourhood of $(0, \theta, 0)$ corresponds in the original variables to a neighbourhood of the torus $\mathcal{T}^I_{\mathcal{S}_0}$; in particular

$$(3.10) \qquad |\mathcal{Y}| \leq r^2, \quad \|a\|_{h^s(\mathbb{Z}^2 \setminus \mathcal{S}_0)} \leq r \quad \Longrightarrow \quad \operatorname{dist} \left(u(\mathcal{Y}, \theta, a), \mathcal{T}^I_{\mathcal{S}_0} \right)_{H^s(\mathbb{T}^2)} \leq cr$$

for some $c > 0$ and any sufficiently small $r \geq 0$.

Step 3: Now, we expand each line by itself. By Taylor expanding around the finite-gap torus corresponding to $(\mathcal{Y}, \theta, a) = (0, \theta, 0)$ we obtain, up to an additive constant,

$$h_{\mathrm{nls1d}}\left(I_{\mathtt{m}_1}+\mathcal{Y}_1,\ldots,I_{\mathtt{m}_{\mathtt{d}}}+\mathcal{Y}_{\mathtt{d}},(|a_{(m,0)}|^2)_{m\notin\mathcal{S}_0}\right)=\sum_{i=1}^{\mathtt{d}}\partial_{\mathtt{m}_i}h_{\mathrm{nls1d}}(I_{\mathtt{m}_1},\ldots,I_{\mathtt{m}_{\mathtt{d}}},0)\mathcal{Y}_i+\sum_{m\in\mathbb{Z}\setminus\mathcal{S}_0}\partial_m h_{\mathrm{nls1d}}(I_{\mathtt{m}_1},\ldots,I_{\mathtt{m}_{\mathtt{d}}},0)|a_{(m,0)}|^2$$

$$-\frac{1}{2}(|\mathcal{Y}|^2+\sum_{m\in\mathbb{Z}\setminus\mathcal{S}_0}|a_{(m,0)}|^4)+O\left(|I|\left\{\sum_{j=1}^{\mathtt{d}}\mathcal{Y}_j+\sum_{m\notin\mathcal{S}_0}|a_{(m,0)}|^2\right\}^2\right)$$

$$+O\left(\left\{\sum_{j=1}^{\mathtt{d}}\mathcal{Y}_j+\sum_{m\notin\mathcal{S}_0}|a_{(m,0)}|^2\right\}^3\right),$$

where we have used formula (3.4) in order to deduce that $\frac{\partial^2 h_{nls1d}}{\partial I_m \partial I_n}(0)=-\delta_n^m$ where δ_n^m is the Kronecker delta.

The following lemma follows easily from Theorem 3.1 (particularly formulae (3.4) and (3.5)):

Lemma 3.3 (Frequencies around the finite gap torus). *Denote*

$$\partial_{I_{\mathtt{m}_j}}h_{\mathrm{nls1d}}(I_{\mathtt{m}_1},\ldots,I_{\mathtt{m}_{\mathtt{d}}},0)\equiv\alpha_{\mathtt{m}_j}^{\mathrm{nls1d}}(I_{\mathtt{m}_1},\ldots,I_{\mathtt{m}_{\mathtt{d}}},0)=\mathtt{m}_j^2-\widetilde{\lambda}_j(I_{\mathtt{m}_1},\ldots,I_{\mathtt{m}_{\mathtt{d}}}).$$

Then,

(1) *The map $(I_{\mathtt{m}_1},\ldots,I_{\mathtt{m}_{\mathtt{d}}})\mapsto\widetilde{\lambda}(I_{\mathtt{m}_1},\ldots,I_{\mathtt{m}_{\mathtt{d}}})=(\widetilde{\lambda}_i(I_{\mathtt{m}_1},\ldots,I_{\mathtt{m}_{\mathtt{d}}}))_{1\leq i\leq\mathtt{d}}$ is a diffeomorphism from a small neighborhood of 0 of $\mathbb{R}^{\mathtt{d}}$ to a small neighborhood of 0 in $\mathbb{R}^{\mathtt{d}}$. Indeed, $\widetilde{\lambda}=$Identity $+$(quadratic in I). More precisely, there exists $\varepsilon_{1d}>0$ such that if $0<\varepsilon<\varepsilon_{1d}$ and*

$$\widetilde{\lambda}(I_{\mathtt{m}_1},\ldots,I_{\mathtt{m}_{\mathtt{d}}})=\varepsilon\lambda,\quad\lambda\in\left(\frac{1}{2},1\right)^{\mathtt{d}}$$

then $(I_{\mathtt{m}_1},\ldots,I_{\mathtt{m}_{\mathtt{d}}})=\varepsilon\lambda+O(\varepsilon^2)$. From now on, and to simplify notation, we will use the vector λ as a parameter as opposed to $(I_{\mathtt{m}_1},\ldots,I_{\mathtt{m}_{\mathtt{d}}})$, and we shall set the vector

$$\omega_i(\lambda)=\mathtt{m}_i^2-\varepsilon\lambda_i,\qquad 1\leq i\leq\mathtt{d}$$

to denote the frequencies at the tangential sites in $\mathcal{S}_0$.

(2) *For $m\in\mathbb{Z}\setminus\mathcal{S}_0$, denoting $\Omega_m(\lambda):=\partial_{I_m}h_{\mathrm{nls1d}}(I_{\mathtt{m}_1}(\lambda),\ldots,I_{\mathtt{m}_{\mathtt{d}}}(\lambda),0)$, we have*

$$\Omega_m(\lambda):=m^2+\frac{\varpi_m(I(\lambda))}{\langle m\rangle},\quad\text{with}\quad\sup_{\lambda\in(\frac{1}{2},1)^{\mathtt{d}}}\sup_{m\in\mathbb{Z}}|\varpi_m(I(\lambda))|\leq C\varepsilon^2\ .$$

With this in mind, line (3.6) becomes

$$h_{\mathrm{nls1d}}\left(I_{\mathtt{m}_1}+\mathcal{Y}_1,\ldots,I_{\mathtt{m}_{\mathtt{d}}}+\mathcal{Y}_{\mathtt{d}},(|a_{(m,0)}|^2)_{m\notin\mathcal{S}_0}\right)=\omega(\lambda)\cdot\mathcal{Y}+\sum_{m\in\mathbb{Z}\setminus\mathcal{S}_0}\Omega_m(\lambda)\left|a_{(m,0)}\right|^2$$

$$-\frac{1}{2}\left(|\mathcal{Y}|^2+\sum_{m\in\mathbb{Z}\setminus\mathcal{S}_0}\left|a_{(m,0)}\right|^4\right)$$

$$+O\left(|I|\left\{\sum_{j=1}^{\mathtt{d}}\mathcal{Y}_j+\sum_{m\notin\mathcal{S}_0}\left|a_{(m,0)}\right|^2\right\}^2\right)$$

$$+O\left(\left\{\sum_{j=1}^{\mathtt{d}}\mathcal{Y}_j+\sum_{m\notin\mathcal{S}_0}\left|a_{(m,0)}\right|^2\right\}^3\right).$$

We now analyze (3.7). This is given by

$$(3.7) = \sum_{\vec{\jmath} \in \mathbb{Z}^2 \setminus \mathbb{Z}} |\vec{\jmath}|^2 |a_{\vec{\jmath}}|^2 + 2 \sum_{\substack{\vec{\jmath}_i = (m_i, n_i), i=3,4, n_i \neq 0 \\ m_1 - m_2 + m_3 - m_4 = 0 \\ n_3 - n_4 = 0}}^{\star} q_{m_1} \bar{q}_{m_2} a_{\vec{\jmath}_3} \bar{a}_{\vec{\jmath}_4} + \mathrm{Re} \sum_{\substack{\vec{\jmath}_i = (m_i, n_i), i=2,4, n_i \neq 0 \\ m_1 - m_2 + m_3 - m_4 = 0 \\ n_2 + n_4 = 0}} \bar{q}_{m_1} a_{\vec{\jmath}_2} \bar{q}_{m_3} a_{\vec{\jmath}_4}$$

where we now think q_m as a function of $\mathcal{Y}, \theta, a$. By Taylor expanding it at $\mathcal{Y} = 0$ and $a = 0$,

$$q_m = q_m(\lambda; \mathcal{Y}, \theta, (a_{(m_1,0)})_{m_1 \in \mathbb{Z} \setminus \mathcal{S}_0}) = \overbrace{q_m(\lambda; 0, \theta, 0)}^{=: q_m^{\mathrm{fg}}(\lambda; \theta)} + \sum_{i=1}^{\mathtt{d}} \frac{\partial q_m}{\partial \mathcal{Y}_i}(\lambda; 0, \theta, 0) \mathcal{Y}_i$$

$$(3.11) \qquad + \sum_{m_1 \in \mathbb{Z} \setminus \mathcal{S}_0} \frac{\partial q_m}{\partial a_{(m_1,0)}}(\lambda; 0, \theta, 0) a_{(m_1,0)} + \frac{\partial q_m}{\partial \bar{a}_{(m_1,0)}}(\lambda; 0, \theta, 0) \bar{a}_{(m_1,0)}$$

$$+ \sum_{\substack{m_1, m_2 \in \mathbb{Z} \setminus \mathcal{S}_0 \\ \sigma_1, \sigma_2 = \pm 1}} Q_{m, m_1 m_2}^{\sigma_1 \sigma_2}(\lambda; \theta) a_{(m_1,0)}^{\sigma_1} a_{(m_2,0)}^{\sigma_2} + \mathcal{O}(\mathcal{Y}^2, \mathcal{Y}a, a^3),$$

where we have denoted $(q_m^{\mathrm{fg}}(\lambda; \theta))_{m \in \mathbb{Z}}$ the finite gap torus (which corresponds to $\mathcal{Y} = 0$, $\mathbf{a} = 0$), and

$$Q_{m, m_1 m_2}^{\sigma_1 \sigma_2}(\lambda; \theta) = \frac{1}{2} \frac{\partial^2 q_m}{\partial a_{m_1}^{\sigma_1} \partial a_{m_2}^{\sigma_2}}(\lambda; 0, \theta, 0).$$

Therefore, we obtain

$$(3.7) = \sum_{\vec{\jmath} \in \mathbb{Z}^2 \setminus \mathbb{Z}} |\vec{\jmath}|^2 |a_{\vec{\jmath}}|^2 + 2 \sum_{\substack{\vec{\jmath}_i = (m_i, n_i), i=3,4, n_i \neq 0 \\ m_1 - m_2 + m_3 - m_4 = 0 \\ n_3 - n_4 = 0}}^{\star} q_{m_1}^{\mathrm{fg}}(\lambda; \theta) \bar{q}_{m_2}^{\mathrm{fg}}(\lambda; \theta) a_{\vec{\jmath}_3} \bar{a}_{\vec{\jmath}_4} + \mathrm{Re} \sum_{\substack{\vec{\jmath}_i = (m_i, n_i), i=2,4, n_i \neq 0 \\ m_1 - m_2 + m_3 - m_4 = 0 \\ n_2 + n_4 = 0}} \bar{q}_{m_1}^{\mathrm{fg}}(\lambda; \theta) a_{\vec{\jmath}_2} \bar{q}_{m_3}^{\mathrm{fg}}(\lambda; \theta) a_{\vec{\jmath}_4}$$

$$+ \left\{ 2 \sum_{\substack{\vec{\jmath}_i = (m_i, n_i), i=3,4, n_i \neq 0 \\ m_1 - m_2 + m_3 - m_4 = 0 \\ n_3 - n_4 = 0}}^{\star} \sum_{m_2' \in \mathbb{Z} \setminus \mathcal{S}_0} \frac{\partial \bar{q}_{m_2}}{\partial \bar{a}_{(m_2',0)}}(\lambda; 0, \theta, 0) q_{m_1}^{\mathrm{fg}}(\lambda; \theta) \bar{a}_{(m_2',0)} a_{\vec{\jmath}_3} \bar{a}_{\vec{\jmath}_4} + \text{similar cubic terms in } (a, \bar{a}) \right\}$$

$$+ (3.7)^{(2)} + (3.7)^{(\geq 3)}$$

where $(3.7)^{(2)}$ are degree 2 terms (cf. Definition 2.1), $(3.7)^{(\geq 3)}$ are of degree ≥ 3. More precisely,

$$(3.7)^{(2)} = 2 \sum_{\substack{\vec{\jmath}_i = (m_i, n_i), i=3,4, n_i \neq 0 \\ m_1 - m_2 + m_3 - m_4 = 0 \\ n_3 - n_4 = 0 \\ 1 \leq i \leq \mathtt{d}}}^{\star} q_{m_1}^{\mathrm{fg}}(\lambda; \theta) \frac{\partial \bar{q}_{m_2}}{\partial \mathcal{Y}_i}(\lambda; 0, \theta, 0) \mathcal{Y}_i a_{\vec{\jmath}_3} \bar{a}_{\vec{\jmath}_4} + \text{similar terms}$$

$$(3.12) \qquad + \sum_{\substack{\vec{\jmath}_i = (m_i, n_i), i=3,4, n_i \neq 0 \\ m_1 - m_2 + m_3 - m_4 = 0 \\ n_3 - n_4 = 0 \\ \sigma_1, \sigma_2 = \pm 1, m_1', m_2' \in \mathbb{Z} \setminus \mathcal{S}_0}}^{\star} L_{m_1, m_2, m_1', m_2'}^{\sigma_1, \sigma_2}(\lambda; \theta) a_{(m_1',0)}^{\sigma_1} a_{(m_2',0)}^{\sigma_2} a_{\vec{\jmath}_3} \bar{a}_{\vec{\jmath}_4} + \text{similar terms},$$

for some uniformly bounded coefficients $L_{m_1, m_2, m_1', m_2'}^{\sigma_1, \sigma_2}$.

Next, we move on to (3.8), for which we have using equation (3.11)

$$
\begin{aligned}
(3.8) =\, & 2\operatorname{Re}\sum_{\substack{\vec{j}_i=(m_i,n_i)\,,\,i=2,3,4\,,\,n_i\neq 0\\ m_1-m_2+m_3-m_4=0\\ -n_2+n_3-n_4=0}} q^{\mathrm{fg}}_{m_1}(\lambda;\theta)\bar{a}_{\vec{j}_2}a_{\vec{j}_3}\bar{a}_{\vec{j}_4}
\end{aligned}
$$

$$
(3.13)\qquad \underbrace{+\,2\operatorname{Re}\sum_{\substack{\vec{j}_i=(m_i,n_i)\,,\,i=2,3,4\,,\,n_i\neq 0\\ m_1-m_2+m_3-m_4=0\\ -n_2+n_3-n_4=0}} \frac{\partial q_{m_1}}{\partial a_{(m'_1,0)}}(\lambda;0,\theta,0)a_{(m'_1,0)}\bar{a}_{\vec{j}_2}a_{\vec{j}_3}\bar{a}_{\vec{j}_4}+\text{similar terms}}_{(3.8)^{(2)}}+(3.8)^{(\geq 3)},
$$

where $(3.8)^{(2)}$ are terms of degree 2 and $(3.8)^{(\geq 3)}$ are terms of degree ≥ 3.

In conclusion, we obtain

$$
(3.14)\qquad \mathcal{H}(\lambda;\mathcal{Y},\theta,\mathbf{a})=\mathcal{N}+\mathcal{H}^{(0)}(\lambda;\theta,\mathbf{a})+\mathcal{H}^{(1)}(\lambda;\theta,\mathbf{a})+\mathcal{H}^{(2)}(\lambda;\mathcal{Y},\theta,\mathbf{a})+\mathcal{H}^{(\geq 3)}(\lambda;\mathcal{Y},\theta,\mathbf{a}),
$$

where

$$
(3.15)\qquad \mathcal{N}=\sum_{i=1}^{\mathtt{d}}\omega_{\mathtt{m}_i}(\lambda)\mathcal{Y}_i+\sum_{m\notin\mathcal{S}_0}\Omega_m(\lambda)|a_{(m,0)}|^2+\sum_{\substack{\vec{j}=(m,n)\in\mathbb{Z}^2\\ n\neq 0}}|\vec{j}|^2|a_{\vec{j}}|^2
$$

$$
\begin{aligned}
(3.16)\qquad \mathcal{H}^{(0)}(\lambda;\theta,\mathbf{a})=\,&2\sum^{\star}_{\substack{\vec{j}_i=(m_i,n_i)\,,\,i=3,4\,,\,n_i\neq 0\\ m_1-m_2+m_3-m_4=0\\ n_3-n_4=0}} q^{\mathrm{fg}}_{m_1}(\lambda;\theta)\bar{q}^{\mathrm{fg}}_{m_2}(\lambda;\theta)a_{\vec{j}_3}\bar{a}_{\vec{j}_4}\\
&+\operatorname{Re}\sum_{\substack{\vec{j}_i=(m_i,n_i)\,,\,i=2,4\,,\,n_i\neq 0\\ m_1-m_2+m_3-m_4=0\\ n_2+n_4=0}} \bar{q}^{\mathrm{fg}}_{m_1}(\lambda;\theta)a_{\vec{j}_2}\bar{q}^{\mathrm{fg}}_{m_3}(\lambda;\theta)a_{\vec{j}_4}
\end{aligned}
$$

$$
\begin{aligned}
(3.17)\qquad \mathcal{H}^{(1)}(\lambda;\theta,\mathbf{a})=\,&2\operatorname{Re}\sum_{\substack{\vec{j}_i=(m_i,n_i)\,,\,i=2,3,4\,,\,n_i\neq 0\\ m_1-m_2+m_3-m_4=0\\ -n_2+n_3-n_4=0}} q^{\mathrm{fg}}_{m_1}(\lambda;\theta)\bar{a}_{\vec{j}_2}a_{\vec{j}_3}\bar{a}_{\vec{j}_4}\\
&+2\sum^{\star}_{\substack{\vec{j}_i=(m_i,n_i)\,,\,i=3,4\,,\,n_i\neq 0\\ m_1-m_2+m_3-m_4=0\\ n_3-n_4=0}}\sum_{m'_2\in\mathbb{Z}\backslash\mathcal{S}_0} \frac{\partial\bar{q}_{m_2}}{\partial\bar{a}_{(m'_2,0)}}(\lambda;0,\theta,0)q^{\mathrm{fg}}_{m_1}(\lambda;\theta)\bar{a}_{(m'_2,0)}a_{\vec{j}_3}\bar{a}_{\vec{j}_4}\\
&+\text{similar cubic terms in }(a,\bar{a})
\end{aligned}
$$

$$
\begin{aligned}
(3.18)\qquad \mathcal{H}^{(2)}(\lambda;\theta,\mathbf{a})=\,&H^{\mathrm{IV}}\big((a_{(m,n)})_{n\neq 0}\big)-\frac{1}{2}\left(|\mathcal{Y}|^2+\sum_{m\in\mathbb{Z}\backslash\mathcal{S}_0}|a_{(m,0)}|^4\right)\\
&+O\left(\varepsilon\left\{\sum_{j=1}^{\mathtt{d}}\mathcal{Y}_j+\sum_{m\notin\mathcal{S}_0}|a_{(m,0)}|^2\right\}^2\right)+(3.7)^{(2)}+(3.8)^{(2)},
\end{aligned}
$$

where $(3.7)^{(2)}$ and $(3.8)^{(2)}$ were defined in (3.12) and (3.13) respectively. Finally, $\mathcal{H}^{(\geq 3)}$ collects all remainder terms of degree ≥ 3.

For short we write $\mathcal{N}$ as $\mathcal{N}=\omega(\lambda)\cdot\mathcal{Y}+\mathcal{D}$ where $\mathcal{D}$ is the diagonal operator

$$
\mathcal{D}:=\sum_{\vec{j}=(m,n)\in\mathbb{Z}^2\backslash\mathcal{S}_0}\Omega^{(0)}_{\vec{j}}|a_{\vec{j}}|^2
$$

and the normal frequencies $\Omega_{\vec{j}}^{(0)}$ are defined by

$$(3.19) \qquad \Omega_{\vec{j}}^{(0)} := \begin{cases} |\vec{j}|^2 & \text{if } \vec{j} = (m,n) \text{ with } n \neq 0 \\ \Omega_m(\lambda) & \text{if } \vec{j} = (m,0), \ m \notin \mathcal{S}_0 \end{cases} .$$

Proceeding as in [MP18], one can prove the following result:

Lemma 3.4. *Fix $\rho > 0$. There exists $\varepsilon_* > 0$ and for any $0 \leq \varepsilon \leq \varepsilon_*$, there exist $r_* \leq \sqrt{\varepsilon}/4$ and $C > 0$ such that $\mathcal{H}^{(0)}, \mathcal{H}^{(1)}, \mathcal{H}^{(2)}$ and $\mathcal{H}^{(\geq 3)}$ belong to $\mathcal{A}_{\rho,r_*}^{\mathcal{O}}$ and $\forall 0 < r \leq r_*$*

$$(3.20) \qquad |\mathcal{H}^{(0)}|_{\rho,r}^{\mathcal{O}} \leq C\varepsilon, \qquad |\mathcal{H}^{(1)}|_{\rho,r}^{\mathcal{O}} \leq C\sqrt{\varepsilon}\, r, \qquad |\mathcal{H}^{(2)}|_{\rho,r}^{\mathcal{O}} \leq Cr^2, \qquad |\mathcal{H}^{(\geq 3)}|_{\rho,r}^{\mathcal{O}} \leq C\frac{r^3}{\sqrt{\varepsilon}}.$$

4. REDUCIBILITY THEORY OF THE QUADRATIC PART

In this section, we review the reducibility of the quadratic part $\mathcal{N} + \mathcal{H}^{(0)}$ (see (3.15) and (3.16)) of the Hamiltonian, which is the main part of the work [MP18]. This will be a symplectic linear change of coordinates that transforms the quadratic part into an effectively diagonal, time independent expression.

4.1. **Restriction to an invariant sublattice $\mathbb{Z}_N^2$.** For $N \in \mathbb{N}$, we define the sublattice $\mathbb{Z}_N^2 := \mathbb{Z} \times N\mathbb{Z}$ and remark that it is invariant for the flow in the sense that the subspace

$$E_N := \{a_{\vec{j}} = \bar{a}_{\vec{j}} = 0, \quad \text{for} \quad \vec{j} \notin \mathbb{Z}_N^2\}$$

is invariant for the original NLS dynamics and that of the Hamiltonian (3.14). From now on, we restrict our system to this invariant sublattice, with

$$(4.1) \qquad N > \max_{1 \leq i \leq \mathtt{d}} |\mathtt{m}_i|.$$

The reason for this restriction is that it simplifies (actually eliminates the need for) some genericity requirements that are needed for the work [MP18] as well as some of the normal forms that we will perform later.

It will also be important to introduce the following two subsets of $\mathbb{Z}_N^2$:

$$(4.2) \qquad \mathscr{S} := \{(\mathtt{m}, n) : \mathtt{m} \in \mathcal{S}_0, \ n \in N\mathbb{Z}, \ n \neq 0\}, \qquad \mathbf{Z} = \mathbb{Z}_N^2 \setminus (\mathscr{S} \cup \mathcal{S}_0).$$

4.2. **Admissible monomials and reducibility.** The reducibility of the quadratic part of the Hamiltonian will introduce a change of variables that modifies the expression of the mass $\mathcal{M}$ and momentum $\mathcal{P}$ as follows. Let us set

$$(4.3) \qquad \begin{aligned} \widetilde{\mathcal{M}} &:= \sum_{i=1}^{\mathtt{d}} \mathcal{Y}_i + \sum_{(m,n) \in \mathbf{Z}} |a_j|^2, \\ \widetilde{\mathcal{P}}_x &:= \sum_{i=1}^{\mathtt{d}} \mathtt{m}_i \mathcal{Y}_i + \sum_{(m,n) \in \mathbf{Z}} m\, |a_{(m,n)}|^2, \\ \widetilde{\mathcal{P}}_y &:= \sum_{(m,n) \in \mathbb{Z}_N^2} n|a_{(m,n)}|^2. \end{aligned}$$

These will be the expressions for the mass and momentum after the change of variables introduced in the following two theorems. Notice the absence of the terms $\sum_{\substack{1 \leq i \leq \mathtt{d} \\ n \in N\mathbb{Z}}} |a_{(\mathtt{m}_i, n)}|^2$ and $\sum_{\substack{1 \leq i \leq \mathtt{d} \\ n \in N\mathbb{Z}}} \mathtt{m}_i |a_{(\mathtt{m}_i, n)}|^2$ from the expressions of $\widetilde{\mathcal{M}}$ and $\widetilde{\mathcal{P}}_x$ above. These terms are absorbed in the new definition of the $\mathcal{Y}$ and $\mathbf{a}$ variables.

Definition 4.1 (Admissible monomials). *Given $\mathbf{j} = (\vec{j}_1, \ldots, \vec{j}_p) \in (\mathbb{Z}_N^2 \setminus \mathcal{S}_0)^p$, $\ell \in \mathbb{Z}^{\mathtt{d}}$, $l \in \mathbb{N}^{\mathtt{d}}$, and $\sigma = (\sigma_1, \ldots, \sigma_p) \in \{-1, 1\}^p$, we say that $(\mathbf{j}, \ell, \sigma)$ is admissible, and denote $(\mathbf{j}, \ell, \sigma) \in \mathfrak{A}_p$, if the monomial $\mathfrak{m} = e^{\mathrm{i}\theta \cdot \ell} \mathcal{Y}^l a_{\vec{j}_1}^{\sigma_1} \ldots a_{\vec{j}_p}^{\sigma_p}$ Poisson commutes with $\widetilde{\mathcal{M}}, \widetilde{\mathcal{P}}_x, \widetilde{\mathcal{P}}_y$. We call a monomial $e^{\mathrm{i}\theta \cdot \ell} \mathcal{Y}^l a_{\vec{j}_1}^{\sigma_1} \ldots a_{\vec{j}_p}^{\sigma_p}$ admissible if $(\mathbf{j}, \ell, \sigma)$ is admissible.*

Definition 4.2. *We define the resonant set at degree 0,*

$$(4.4) \qquad \mathfrak{R}_2 := \{(\vec{j}_1, \vec{j}_2, \ell, \sigma_1, \sigma_2)\} \in \mathfrak{A}_2 : \ell = 0, \ \sigma_1 = -\sigma_2, \ \vec{j}_1 = \vec{j}_2\}.$$

Theorem 4.3. *Fix $\varepsilon_0 > 0$ sufficiently small. There exist positive $\rho_0, \gamma_0, \tau_0, r_0, \mathsf{L}_0$ (with L_0 depending only on d) such that the following holds true uniformly for all $0 < \varepsilon \leq \varepsilon_0$: For an L_0-generic choice of the set $\mathcal{S}_0$ (in the sense of Definition 1.1), there exist a compact domain $\mathcal{O}_0 \subseteq (1/2, 1)^\mathsf{d}$, satisfying $|(1/2, 1)^\mathsf{d} \setminus \mathcal{O}_0| \leq \varepsilon_0$, and Lipschitz (in λ) functions $\{\Omega_{\vec{j}}\}_{\vec{j} \in \mathbb{Z}_N^2 \setminus \mathcal{S}_0}$ defined on $\mathcal{O}_0$ (described more precisely in Theorem 4.4 below) such that:*

(1) *The set*

(4.5)
$$\mathcal{C}^{(0)} := \left\{ \lambda \in \mathcal{O}_0 : \; |\omega \cdot \ell + \sigma_1 \Omega_{\vec{j}_1}(\lambda, \varepsilon) + \sigma_2 \Omega_{\vec{j}_2}(\lambda, \varepsilon)| \geq \gamma_0 \frac{\varepsilon}{\langle \ell \rangle^{\tau_0}} \;, \; \forall (\vec{j}, \ell, \sigma) \in \mathfrak{A}_2 \setminus \mathfrak{K}_2 \right\}$$

has positive measure. In fact $|\mathcal{O}_0 \setminus \mathcal{C}^{(0)}| \lesssim \varepsilon_0^{\kappa_0}$ for some $\kappa_0 > 0$ independent of ε_0.

(2) *For each $\lambda \in \mathcal{C}^{(0)}$ and all $r \in [0, r_0]$, $\rho \in [\frac{\rho_0}{64}, \rho_0]$, there exists an invertible symplectic change of variables $\mathcal{L}^{(0)}$, that is well defined and majorant analytic from $D(\rho/8, \zeta_0 r) \to D(\rho, r)$ (here $\zeta_0 > 0$ is a constant depending only on $\rho_0, \max(|\mathfrak{m}_k|^2)$) and such that if $\mathbf{a} \in h^1(\mathbb{Z}_N^2 \setminus \mathcal{S}_0)$, then*

$$(\mathcal{N} + \mathcal{H}^{(0)}) \circ \mathcal{L}^{(0)}(\mathcal{Y}, \theta, \mathbf{a}) = \omega \cdot \mathcal{Y} + \sum_{\vec{j} \in \mathbb{Z}_N^2 \setminus \mathcal{S}_0} \Omega_{\vec{j}} |a_{\vec{j}}|^2.$$

(3) *The mass $\mathcal{M}$ and the momentum $\mathcal{P}$ (defined in (2.5)) in the new coordinates are given by*

(4.6)
$$\mathcal{M} \circ \mathcal{L}^{(0)} = \widetilde{\mathcal{M}}, \quad \mathcal{P} \circ \mathcal{L}^{(0)} = \widetilde{\mathcal{P}},$$

where $\widetilde{\mathcal{M}}$ and $\widetilde{\mathcal{P}}$ are defined in (4.3).

(4) *The map $\mathcal{L}^{(0)}$ maps h^1 to itself and has the following form*

$$\mathcal{L}^{(0)}: \quad \mathbf{a} \mapsto L(\lambda; \theta, \varepsilon)\mathbf{a}, \qquad \mathcal{Y} \mapsto \mathcal{Y} + (\mathbf{a}, Q(\lambda; \theta, \varepsilon)\mathbf{a}), \qquad \theta \mapsto \theta.$$

The same holds for the inverse map $(\mathcal{L}^{(0)})^{-1}$.

(5) *The linear maps $L(\lambda; \theta, \varepsilon)$ and $Q(\lambda; \theta, \varepsilon)$ are block diagonal in the y Fourier modes, in the sense that $L = \mathrm{diag}_{n \in N\mathbb{N}}(L_n)$ with each L_n acting on the sequence $\{a_{(m,n)}, a_{(m,-n)}\}_{m \in \mathbb{Z}}$ (and similarly for Q). Moreover, $L_0 = \mathrm{Id}$ and L_n is of the form $\mathrm{Id} + S_n$ where S_n is a smoothing operator in the following sense: with the smoothing norm $\lceil \cdot \rceil_{\rho, -1}$ defined in (4.7) below*

$$\sup_{n \neq 0} \lceil S_n \circ P_{\{|m| \geq (\mathfrak{m}_\mathsf{d} + 1)\}} \rceil_{\rho, -1} \lesssim \varepsilon,$$

where $P_{\{|m| \geq K\}}$ is the orthogonal projection of a sequence $(c_m)_{m \in \mathbb{Z}}$ onto the modes $|m| \geq K$.

The above smoothing norm is defined as follows: Let $S(\lambda; \theta, \varepsilon)$ be an operator acting on sequences $(c_k)_{k \in \mathbb{Z}}$ through its matrix elements $S(\lambda; \theta, \varepsilon)_{m,k}$. Let us denote by $S(\lambda; \ell, \varepsilon)_{m,k}$ the θ-Fourier coefficients of $S(\lambda; \theta, \varepsilon)_{m,k}$. For $\rho, \nu > 0$ we define $\lceil S(\lambda; \theta, \varepsilon) \rceil_{\rho, \nu}$ as:

(4.7)
$$\lceil S(\lambda; \theta, \varepsilon) \rceil_{\rho, \nu} := \sup_{\|c\|_{\ell^1} \leq 1} \left\| \left(\sum_{\substack{k \in \mathbb{Z} \\ \ell \in \mathbb{Z}^\mathsf{d}}} e^{\rho|\ell|} |S_{m,k}(\lambda; \ell, \varepsilon)| \langle k \rangle^{-\nu} c_k \right) \right\|_{\ell^1}.$$

This definition is equivalent to the more general norm used in Definition 3.9 of [MP18]. Roughly speaking, the boundedness of this norm means that, in terms of its action on sequences, S maps $\langle k \rangle^\nu \ell^1 \to \ell^1$. As observed in Remark 3.10 of [MP18], thanks to the conservation of momentum this also means that S maps $\ell^1 \to \langle k \rangle^{-\nu} \ell^1$.

Proof of Theorem 4.3. The result follows from [MP18], by applying first the change of variables in Theorem 5.1 and then the one in Theorem 7.1 to the quadratic part of the NLS Hamiltonian (hence ignoring the terms $\widetilde{\mathcal{H}}^{(1)}, \widetilde{\mathcal{H}}^{(\geq 2)}$ in formula (5.2) and the terms $\mathcal{K}^{(1)}, \mathcal{K}^{(\geq 2)}$ in formula (7.3) of [MP18]). Note that in [MP18] the results are proved in h^s norm with $s > 1$, for instance in (4.7) the ℓ^1 norm is substituted with the h^s one. However the proofs only rely on momentum conservation and on the fact that h^s is an algebra w.r.t. convolution, which holds true also for ℓ^1. Hence the proof of our case is identical and we do not repeat it.

We are able to describe quite precisely the asymptotics of the frequencies $\Omega_{\vec{j}}$ of Theorem 4.3.

Theorem 4.4. *For any $0 < \varepsilon \le \varepsilon_0$ and $\lambda \in \mathcal{C}^{(0)}$, the frequencies $\Omega_{\vec{\jmath}} \equiv \Omega_{\vec{\jmath}}(\lambda, \varepsilon)$, $\vec{\jmath} = (m, n) \in \mathbb{Z}_N^2 \setminus \mathcal{S}_0$, introduced in Theorem [4.3] have the following asymptotics:*

$$
(4.8) \qquad \Omega_{\vec{\jmath}}(\lambda, \varepsilon) = \begin{cases} \widetilde{\Omega}_{\vec{\jmath}}(\lambda, \varepsilon) + \dfrac{\varpi_m(\lambda, \varepsilon)}{\langle m \rangle}, & n = 0 \\[2mm] \widetilde{\Omega}_{\vec{\jmath}}(\lambda, \varepsilon) + \dfrac{\Theta_m(\lambda, \varepsilon)}{\langle m \rangle^2} + \dfrac{\Theta_{m,n}(\lambda, \varepsilon)}{\langle m \rangle^2 + \langle n \rangle^2}, & n \ne 0 \end{cases} \quad ,
$$

where

$$
\widetilde{\Omega}_{\vec{\jmath}}(\lambda, \varepsilon) := \begin{cases} m^2, & \vec{\jmath} = (m, 0), m \notin \mathcal{S}_0 \\ m^2 + n^2, & \vec{\jmath} = (m, n) \in \mathbf{Z}, n \ne 0 \\ \varepsilon \mu_i(\lambda) + n^2, & \vec{\jmath} = (\mathtt{m}_i, n) \in \mathscr{S}, n \ne 0 \end{cases}
$$

where $\mathbf{Z}$ and $\mathscr{S}$ are the sets defined in [(4.2)].
    Here the $\{\mu_i(\lambda)\}_{1 \le i \le \mathtt{d}}$ are the roots of the polynomial

$$
P(t, \lambda) := \prod_{i=1}^{\mathtt{d}} (t + \lambda_i) - 2 \sum_{i=1}^{\mathtt{d}} \lambda_i \prod_{k \ne i} (t + \lambda_k),
$$

which is irreducible over $\mathbb{Q}(\lambda)[t]$.
    Finally $\mu_i(\lambda)$, $\{\varpi_m(\lambda, \varepsilon)\}_{m \in \mathbb{Z} \setminus \mathcal{S}_0}$, $\{\Theta_m(\lambda, \varepsilon)\}_{m \in \mathbb{Z}}$ and $\{\Theta_{m,n}(\lambda, \varepsilon)\}_{(m,n) \in \mathbb{Z}_N^2 \setminus \mathcal{S}_0}$ fulfill

$$
(4.9) \qquad \sum_{1 \le i \le \mathtt{d}} |\mu_i(\cdot)|^{\mathcal{O}_0} + \sup_{\varepsilon \le \varepsilon_0} \frac{1}{\varepsilon^2} \Big(\sup_{m \in \mathbb{Z} \setminus \mathcal{S}_0} |\varpi_m(\cdot, \varepsilon)|^{\mathcal{O}_0} + \sup_{m \in \mathbb{Z}} |\Theta_m(\cdot, \varepsilon)|^{\mathcal{O}_0} + \sup_{\substack{(m,n) \in \mathbb{Z}_N^2 \\ n \ne 0}} |\Theta_{m,n}(\cdot, \varepsilon)|^{\mathcal{O}_0} \Big) \le \mathtt{M}_0
$$

for some $\mathtt{M}_0$ independent of ε.

    Theorem [4.4] follows from Theorem 2.10 (see also Corollary 7.5) of [MP18], together with the observation that the set $\mathscr{C}$ defined in Definition 2.3 of [MP18] satisfies $\mathscr{C} \cap \mathbb{Z}_N^2 = \emptyset$ if $N > \max_i |\mathtt{m}_i|$.
    We conclude this section with a series of remarks.

Remark 4.5. *Notice that the $\{\mu_i(\lambda)\}_{1 \le i \le \mathtt{d}}$ depend on the number $\mathtt{d}$ of tangential sites but not on the $\{\mathtt{m}_i\}_{1 \le i \le \mathtt{d}}$.*

Remark 4.6. *The asymptotic expansion [(4.8)] of the normal frequencies does not contain any constant term. The reason is that we canceled such a term when we subtracted the quantity $M(u)^2$ from the Hamiltonian at the very beginning (see the footnote in Section [3.1]). Of course if we had not removed $M(u)^2$, we would have had a constant correction to the frequencies, equal to $\|q(\omega t, \cdot)\|_{L^2}^2$. Since $q(\omega t, x)$ is a solution of ([2D-NLS]), it enjoys mass conservation, and thus $\|q(\omega t, \cdot)\|_{L^2}^2 = \|q(0, \cdot)\|_{L^2}^2$ is independent of time.*

Remark 4.7. *In the new variables, the* selection rules *of Remark [2.2] become (with $\mathcal{H}$ expanded as in [(2.2)]):*

$$
\begin{aligned}
\{\mathcal{H}, \widetilde{\mathcal{M}}\} = 0 \quad &\Leftrightarrow \quad \mathcal{H}_{\alpha, \beta, \ell} \left(\widetilde{\eta}(\alpha, \beta) + \eta(\ell) \right) = 0 \\
\{\mathcal{H}, \widetilde{\mathcal{P}}_x\} = 0 \quad &\Leftrightarrow \quad \mathcal{H}_{\alpha, \beta, \ell} \left(\widetilde{\pi}_x(\alpha, \beta) + \pi(\ell) \right) = 0 \\
\{\mathcal{H}, \widetilde{\mathcal{P}}_y\} = 0 \quad &\Leftrightarrow \quad \mathcal{H}_{\alpha, \beta, \ell} \left(\pi_y(\alpha, \beta) \right) = 0
\end{aligned}
$$

where $\eta(\ell)$ is defined in [(2.3)], $\pi_y(\alpha, \beta), \pi(\ell)$ in [(2.4)], while

$$
\widetilde{\eta}(\alpha, \beta) := \sum_{\vec{\jmath} \in \mathbf{Z}} (\alpha_{\vec{\jmath}} - \beta_{\vec{\jmath}}) \ ,
$$

$$
\widetilde{\pi}_x(\alpha, \beta) := \sum_{\vec{\jmath} = (m, n) \in \mathbf{Z}} m(\alpha_{\vec{\jmath}} - \beta_{\vec{\jmath}}).
$$

5. ELIMINATION OF CUBIC TERMS

If we apply the change $\mathcal{L}^{(0)}$ obtained in Theorem 4.3 to Hamiltonian (3.14), we obtain

$$\mathcal{K}(\lambda; \mathcal{Y}, \theta, \mathbf{a}) := \mathcal{H} \circ \mathcal{L}^{(0)}(\lambda; \mathcal{Y}, \theta, \mathbf{a}) = \omega \cdot \mathcal{Y} + \sum_{\vec{j} \in \mathbb{Z}_N^2 \setminus \mathcal{S}_0} \Omega_{\vec{j}} |a_{\vec{j}}|^2 + \mathcal{K}^{(1)} + \mathcal{K}^{(2)} + \mathcal{K}^{(\geq 3)},$$

(5.1)

$$\mathcal{K}^{(j)} = \mathcal{H}^{(j)} \circ \mathcal{L}^{(0)} \quad (j = 1, 2), \qquad \mathcal{K}^{(\geq 3)} = \mathcal{H}^{(\geq 3)} \circ \mathcal{L}^{(0)}.$$

As a direct consequence of Lemma 3.4 and Theorem 4.3, estimates (3.20) hold also for $\mathcal{K}^{(j)}$, $j = 1, 2$ and $\mathcal{K}^{(\geq 3)}$.

We now perform one step of Birkhoff normal form change of variables which cancels out $\mathcal{K}^{(1)}$ completely. In order to define such a change of variables we need to impose third order Melnikov conditions, which hold true on a subset of the set $\mathcal{C}^{(0)}$ of Theorem 4.3.

Lemma 5.1. *Fix $0 < \varepsilon_1 < \varepsilon_0$ sufficiently small and $\tau_1 > \tau_0$ sufficiently large. There exist constants $\gamma_1 > 0, \mathtt{L}_1 > \mathtt{L}_0$ (with $\mathtt{L}_1$ depending only on $\mathtt{d}$), such that for all $0 < \varepsilon \leq \varepsilon_1$ and for an $\mathtt{L}_1$-generic choice of the set $\mathcal{S}_0$ (in the sense of Definition 1.1), the set*

$$\mathcal{C}^{(1)} := \left\{ \lambda \in \mathcal{C}^{(0)} : |\omega \cdot \ell + \sigma_1 \Omega_{\vec{j}_1}(\lambda, \varepsilon) + \sigma_2 \Omega_{\vec{j}_2}(\lambda, \varepsilon) + \sigma_3 \Omega_{\vec{j}_3}(\lambda, \varepsilon)| \geq \gamma_1 \frac{\varepsilon}{\langle \ell \rangle^{\tau_1}} , \ \forall (\vec{j}, \ell, \sigma) \in \mathfrak{A}_3 \right\},$$

where $\mathfrak{A}_3$ is introduced in Definition 4.1, has positive measure. More precisely, $|\mathcal{C}^{(0)} \setminus \mathcal{C}^{(1)}| \lesssim \varepsilon_1^{\kappa_1}$ for some constant $\kappa_1 > 0$ independent of ε_1.

This lemma is proven in Appendix C of [MP18].

The main result of this section is the following theorem.

Theorem 5.2. *Assume the same hypotheses and use the same notation as in Lemma 5.1. Consider the constants $\mathtt{L}_1$, γ_1, τ_1 given by Lemma 5.1, the associated set $\mathcal{C}^{(1)}$, and the constants ε_0, ρ_0 and r_0 given in Theorem 4.3. There exist $0 < \varepsilon_1 \leq \varepsilon_0$, $0 < \rho_1 \leq \rho_0/64$, $0 < r_1 \leq r_0$ such that the following holds true for all $0 < \varepsilon \leq \varepsilon_1$. For each $\lambda \in \mathcal{C}^{(1)}$ and all $0 < r \leq r_1$, $0 < \rho \leq \rho_1$, there exists a symplectic change of variables $\mathcal{L}^{(1)}$, that is well defined and majorant analytic from $D(\rho/2, r/2) \to D(\rho, r)$ such that applied to Hamiltonian $\mathcal{K}$ in (5.1) leads to*

$$\mathcal{Q} := \mathcal{K} \circ \mathcal{L}^{(1)}(\lambda; \mathcal{Y}, \theta, \mathbf{a}) = \omega \cdot \mathcal{Y} + \sum_{\vec{j} \in \mathbb{Z}_N^2 \setminus \mathcal{S}_0} \Omega_{\vec{j}}(\lambda, \varepsilon) |a_{\vec{j}}|^2 + \mathcal{Q}^{(2)} + \mathcal{Q}^{(\geq 3)} ,$$

(5.2)

where

  (i) *the map $\mathcal{L}^{(1)}$ is the time-1 flow of a cubic hamiltonian $\chi^{(1)}$ such that $|\chi^{(1)}|_{\rho/2, r/2}^{\mathcal{C}^{(1)}} \lesssim \frac{r}{\sqrt{\varepsilon}}$.*
  (ii) *$\mathcal{Q}^{(2)}$ is of degree 2 (in the sense of Definition 2.1) and is given by*

$$\mathcal{Q}^{(2)} = \mathcal{K}^{(2)} + \frac{1}{2} \{ \mathcal{K}^{(1)}, \chi^{(1)} \},$$

(5.3)

  and satisfies $|\mathcal{Q}^{(2)}|_{\rho/2, r/2} \lesssim r^2$.
  (iii) *$\mathcal{Q}^{(\geq 3)}$ is of degree at least 3 and satisfies*

$$|\mathcal{Q}^{(\geq 3)}|_{\rho/2, r/2}^{\mathcal{C}^{(1)}} \lesssim \frac{r^3}{\sqrt{\varepsilon}}.$$

(5.4)

  (iv) *$\mathcal{L}^{(1)}$ satisfies $\widetilde{\mathcal{M}} \circ \mathcal{L}^{(1)} = \widetilde{\mathcal{M}}$ and $\widetilde{\mathcal{P}} \circ \mathcal{L}^{(1)} = \widetilde{\mathcal{P}}$.*
  (v) *$\mathcal{L}^{(1)}$ maps $D(\rho/2, r/2) \cap h^1 \to D(\rho, r) \cap h^1$, and if we denote $(\widetilde{\mathcal{Y}}, \widetilde{\theta}, \widetilde{\mathbf{a}}) = \mathcal{L}^{(1)}(\mathcal{Y}, \theta, \mathbf{a})$, then*

$$\|\widetilde{\mathbf{a}} - \mathbf{a}\|_{\ell^1} \lesssim \|\mathbf{a}\|_{\ell^1}^2.$$

(5.5)

To prove this theorem, we state the following lemma, which is proved in [MP18].

Lemma 5.3. *For every $\rho, r > 0$ the following holds true:*

(i) *Let $h, f \in \mathcal{A}_{\rho,r}^{\mathcal{O}}$. For any $0 < \rho' < \rho$ and $0 < r' < r$, one has*

$$|\{f,g\}|_{\rho',r'}^{\mathcal{O}} \leq \upsilon^{-1} C |f|_{\rho,r}^{\mathcal{O}} |g|_{\rho,r}^{\mathcal{O}} .$$

where $\upsilon := \min\left(1 - \frac{r'}{r}, \rho - \rho'\right)$. If $\upsilon^{-1}|f|_{\rho,r}^{\mathcal{O}} < \zeta$ sufficiently small then the (time-1 flow of the) Hamiltonian vector field X_f defines a close to identity canonical change of variables $\mathcal{T}_f$ such that

$$|h \circ \mathcal{T}_f|_{\rho',r'}^{\mathcal{O}} \leq (1 + C\zeta)|h|_{\rho,r}^{\mathcal{O}} , \qquad \text{for all } 0 < \rho' < \rho , \ \ 0 < r' < r .$$

(ii) *Let $f, g \in \mathcal{A}_{\rho,r}^{\mathcal{O}}$ of minimal degree respectively $\mathtt{d}_f$ and $\mathtt{d}_g$ (see Definition 2.1) and define the function*

(5.6)
$$\widetilde{\mathcal{G}}_{\mathtt{i}}(f; h) = \sum_{l=\mathtt{i}}^{\infty} \frac{(\mathrm{ad}f)^l}{l!} h , \quad \mathrm{ad}(f)h := \{h, f\} .$$

Then $\widetilde{\mathcal{G}}_{\mathtt{i}}(f; g)$ is of minimal degree $\mathtt{d}_f \mathtt{i} + \mathtt{d}_g$ and we have the bound

$$|\widetilde{\mathcal{G}}_{\mathtt{i}}(f; h)|_{\rho',r'}^{\mathcal{O}} \leq C(\rho)\upsilon^{-\mathtt{i}} \left(|f|_{\rho,r}^{\mathcal{O}}\right)^{\mathtt{i}} |g|_{\rho,r}^{\mathcal{O}} , \qquad \forall 0 < \rho' < \rho , \ \ 0 < r' < r .$$

Proof of Theorem 5.2. We look for $\mathcal{L}^{(1)}$ as the time-one-flow of a Hamiltonian $\chi^{(1)}$. With $\widehat{\mathcal{N}} := \omega \cdot \mathcal{Y} + \sum_{\vec{\jmath} \in \mathbb{Z}_N^2 \setminus \mathcal{S}_0} \Omega_{\vec{\jmath}}(\lambda, \varepsilon)|a_{\vec{\jmath}}|^2$ and $\widetilde{\mathcal{G}}_j(\chi^{(1)}; \cdot) = \sum_{k \geq j} \frac{\mathrm{ad}(\chi^{(1)})^{k-1}[\{\cdot, \chi^{(1)}\}]}{k!}$, we have

(5.7)
$$\mathcal{K} \circ \mathcal{L}^{(1)} = \widehat{\mathcal{N}} + \{\widehat{\mathcal{N}}, \chi^{(1)}\} + \mathcal{K}^{(1)}$$

(5.8)
$$+ \widetilde{\mathcal{G}}_2(\chi^{(1)}; \widehat{\mathcal{N}}) + \{\mathcal{K}^{(1)}, \chi^{(1)}\} + \widetilde{\mathcal{G}}_2(\chi^{(1)}; \mathcal{K}^{(1)})$$

(5.9)
$$+ \mathcal{K}^{(2)} + \widetilde{\mathcal{G}}_1(\chi^{(1)}; \mathcal{K}^{(2)}) + \mathcal{K}^{(\geq 3)} \circ \mathcal{L}^{(1)}$$

We choose $\chi^{(1)}$ to solve the homological equation $\{\widehat{\mathcal{N}}, \chi^{(1)}\} + \mathcal{K}^{(1)} = 0$. Thus we set

$$\mathcal{K}^{(1)} = \sum_{\ell,\mathbf{j},\vec{\sigma} \in \mathfrak{A}_3} K_{\ell,\mathbf{j}}^{\vec{\sigma}}(\lambda, \varepsilon) \, e^{\mathrm{i}\theta \cdot \ell} a_{\vec{\jmath}_1}^{\sigma_1} a_{\vec{\jmath}_2}^{\sigma_2} a_{\vec{\jmath}_3}^{\sigma_3} , \qquad \chi^{(1)} = \sum_{\ell,\mathbf{j},\vec{\sigma} \in \mathfrak{A}_3} \chi_{\ell,\mathbf{j}}^{\vec{\sigma}}(\lambda, \varepsilon) \, e^{\mathrm{i}\theta \cdot \ell} a_{\vec{\jmath}_1}^{\sigma_1} a_{\vec{\jmath}_2}^{\sigma_2} a_{\vec{\jmath}_3}^{\sigma_3}$$

with

$$\chi_{\ell,\mathbf{j}}^{\vec{\sigma}}(\lambda, \varepsilon) := \frac{\mathrm{i} K_{\ell,\mathbf{j}}^{\vec{\sigma}}(\lambda, \varepsilon)}{\omega \cdot \ell + \sigma_1 \Omega_{\vec{\jmath}_1}(\lambda, \varepsilon) + \sigma_2 \Omega_{\vec{\jmath}_2}(\lambda, \varepsilon) + \sigma_3 \Omega_{\vec{\jmath}_3}(\lambda, \varepsilon)} .$$

Since $\lambda \in \mathcal{C}^{(1)}$, we have

$$|\chi^{(1)}|_{\frac{\rho}{2},r}^{\mathcal{C}^{(1)}} \lesssim \frac{r}{\sqrt{\varepsilon}},$$

since the terms q_m^{fg} appearing in $\mathcal{H}^{(1)}$ (and hence $\mathcal{K}^{(1)}$) are $O(\sqrt{\varepsilon})$. We come to the terms of line (5.8). First we use the homological equation $\{\widehat{\mathcal{N}}, \chi^{(1)}\} + \mathcal{K}^{(1)} = 0$ to get that

$$\widetilde{\mathcal{G}}_2\left(\chi^{(1)}; \widehat{\mathcal{N}}\right) = \sum_{k \geq 2} \frac{\mathrm{ad}(\chi^{(1)})^{k-1}[\{\widehat{\mathcal{N}}, \chi^{(1)}\}]}{k!} = -\frac{1}{2}\{\mathcal{K}^{(1)}, \chi^{(1)}\} - \sum_{k \geq 2} \frac{\mathrm{ad}(\chi^{(1)})^k[\mathcal{K}^{(1)}]}{(k+1)!}.$$

Therefore, we set $\mathcal{Q}^{(2)}$ as in (5.3) and

$$\mathcal{Q}^{(\geq 3)} = \widetilde{\mathcal{G}}_2(\chi^{(1)}; \mathcal{K}^{(1)}) + \widetilde{\mathcal{G}}_1(\chi^{(1)}; \mathcal{K}^{(2)}) + \mathcal{K}^{(\geq 3)} \circ \mathcal{L}^{(1)} - \sum_{k \geq 2} \frac{\mathrm{ad}(\chi^{(1)})^k[\mathcal{K}^{(1)}]}{(k+1)!}.$$

By Lemma 5.3, $\mathcal{Q}^{(\geq 3)}$ has degree at least 3 and fulfills the quantitative estimate (5.4). To prove (iv), we use the fact that $\{\widetilde{\mathcal{M}}, \chi^{(1)}\} = \{\widetilde{\mathcal{P}}, \chi^{(1)}\} = 0$ follows since $\mathcal{K}^{(1)}$ commutes with $\widetilde{\mathcal{M}}$ and $\widetilde{\mathcal{P}}$, hence its monomials fulfill the selection rules of Remark 4.7. By the explicit formula for $\chi^{(1)}$ above, it follows that the same selection rules hold for $\chi^{(1)}$, and consequently $\mathcal{L}^{(1)}$ preserves $\widetilde{\mathcal{M}}$ and $\widetilde{\mathcal{P}}$.

It remains to show the mapping properties of the operator $\mathcal{L}^{(1)}$. First we show that it maps $D(\rho/2, r/2) \to D(\rho, r)$. Let us denote by $(\widetilde{\mathcal{Y}}, \widetilde{\theta}, \widetilde{\mathbf{a}}) = \mathcal{L}^{(1)}(\mathcal{Y}, \theta, \mathbf{a})$, then $(\widetilde{\mathcal{Y}}, \widetilde{\theta}, \widetilde{\mathbf{a}}) = (\widetilde{\mathcal{Y}}(s), \widetilde{\theta}(s), \widetilde{\mathbf{a}}(s))\big|_{s=1}$ where $(\widetilde{\mathcal{Y}}(s), \widetilde{\theta}(s), \widetilde{\mathbf{a}}(s))$ is the Hamiltonian flow generated by $\chi^{(1)}$ at time $0 \le s \le 1$. Using the identity

$$(\widetilde{\mathcal{Y}}(t), \widetilde{\theta}(t), \widetilde{\mathbf{a}}(t)) = (\mathcal{Y}, \theta, \mathbf{a}) + \int_0^t X_{\chi^{(1)}} \left(\widetilde{\mathcal{Y}}(s), \widetilde{\theta}(s), \widetilde{\mathbf{a}}(s) \right) \mathrm{d}s$$

where $X_{\chi^{(1)}}$ is the Hamiltonian vector field associated with $\chi^{(1)}$ above, and a standard continuity (bootstrap) argument, we conclude that $(\widetilde{\mathcal{Y}}, \widetilde{\theta}, \widetilde{\mathbf{a}}) \in D(\rho, r)$. Similarly, one also concludes estimate (5.5). Finally, to prove that $\mathcal{L}^{(1)}$ maps $D(\rho/2, r/2) \cap h^1 \to h^1$, we note that $\widehat{\mathcal{N}}$ is equivalent to the square of the h^1 norm, and

$$\widehat{\mathcal{N}} \circ \mathcal{L}^{(1)} = \widehat{\mathcal{N}} + \widetilde{\mathbb{G}}_1(\chi^{(1)}; \widehat{\mathcal{N}}) = \widehat{\mathcal{N}} - \sum_{k \ge 0} \frac{\mathrm{ad}(\chi^{(1)})^k[\mathcal{K}^{(1)}]}{(k+1)!} = \widehat{\mathcal{N}} + O(\sqrt{\varepsilon} r^3),$$

and this completes the proof.

$\square$

6. ANALYSIS OF THE QUARTIC PART OF THE HAMILTONIAN

At this stage, we are left with the Hamiltonian $\mathcal{Q}$ given in (5.2). The aim of this section is to eliminate non-resonant terms from $\mathcal{Q}^{(2)}$. First note that $\mathcal{Q}^{(2)}$ contains monomials which have one of the two following forms

$$e^{\mathrm{i}\theta \cdot \ell} a_{\vec{j}_1}^{\sigma_1} a_{\vec{j}_2}^{\sigma_2} a_{\vec{j}_3}^{\sigma_3} a_{\vec{j}_4}^{\sigma_4} \quad \text{or} \quad e^{\mathrm{i}\theta \cdot \ell} \mathcal{Y}^l a_{\vec{j}_1}^{\sigma_1} a_{\vec{j}_2}^{\sigma_2} \quad \text{with} \ |l| = 1.$$

In order to cancel out the terms quadratic in a by a Birkhoff Normal form procedure, we only need the *second Melnikov conditions* imposed in (4.5). In order to cancel out the quartic tems in a we need *fourth Melnikov conditions*, namely to control expressions of the form

$$(6.1) \qquad \omega(\lambda) \cdot \ell + \sigma_1 \Omega_{\vec{j}_1}(\lambda, \varepsilon) + \sigma_2 \Omega_{\vec{j}_2}(\lambda, \varepsilon) + \sigma_3 \Omega_{\vec{j}_3}(\lambda, \varepsilon) + \sigma_4 \Omega_{\vec{j}_4}(\lambda, \varepsilon), \quad \sigma_i = \pm 1.$$

We start by defining the following set $\mathfrak{K}_4 \subset \mathfrak{A}_4$ (see Definition 4.1),

$$(6.2) \qquad \mathfrak{K}_4 := \Big\{ (\mathbf{j}, \ell, \sigma) \colon \ell = 0 \text{ and } \vec{j}_1, \vec{j}_2, \vec{j}_3, \vec{j}_4 \notin \mathscr{S} \text{ form a rectangle}$$

$$\ell = 0 \text{ and } \vec{j}_1, \vec{j}_2 \notin \mathscr{S}, \vec{j}_3, \vec{j}_4 \in \mathscr{S} \text{ form a horizontal rectangle (even degenerate)}$$

$$\ell \ne 0, \ \vec{j}_1, \vec{j}_2, \vec{j}_3 \in \mathscr{S}, \ \vec{j}_4 \notin \mathscr{S} \text{ and } |m_4| < M_0, \text{ where } M_0 \text{ is a universal constant}$$

$$\ell = 0, \ \vec{j}_1, \vec{j}_2, \vec{j}_3, \vec{j}_4 \in \mathscr{S} \text{ form a horizontal trapezoid} \Big\}$$

where $\mathscr{S}$ is the set defined in (4.2). Here a trapezoid (or a rectangle) is said to be *horizontal* if two sides are parallel to the x-axis.

Proposition 6.1. *Fix $0 < \varepsilon_2 < \varepsilon_1$ sufficiently small and $\tau_2 > \tau_1$ sufficiently large. There exist positive $\gamma_2 > 0, \mathtt{L}_2 \ge \mathtt{L}_1$ (with $\mathtt{L}_2$ depending only on $\mathbf{d}$), such that for all $0 < \varepsilon \le \varepsilon_2$ and for an $\mathtt{L}_2$-generic choice of the set $\mathcal{S}_0$ (in the sense of Definition 1.1), the set*

$$\mathcal{C}^{(2)} := \left\{ \lambda \in \mathcal{C}^{(1)} : \ |\omega \cdot \ell + \sigma_1 \Omega_{\vec{j}_1}(\lambda, \varepsilon) + \sigma_2 \Omega_{\vec{j}_2}(\lambda, \varepsilon) + \sigma_3 \Omega_{\vec{j}_3}(\lambda, \varepsilon) + \sigma_4 \Omega_{\vec{j}_4}(\lambda, \varepsilon)| \ge \gamma_2 \frac{\varepsilon}{\langle \ell \rangle^{\tau_2}} \ , \ \forall (\vec{j}, \ell, \sigma) \in \mathfrak{A}_4 \setminus \mathfrak{K}_4 \right\}$$

has positive measure and $\left| \mathcal{C}^{(1)} \setminus \mathcal{C}^{(2)} \right| \lesssim \varepsilon_2^{\kappa_2}$ for some $\kappa_2 > 0$ independent of ε_2.

The proof of the proposition, being quite technical, is postponed to Appendix A.

An immediate consequence, following the same strategy as for the proof of Theorem 5.2, is the following result. We define $\Pi_{\mathfrak{K}_4}$ as the projection of a function in $D(\rho, r)$ onto the sum of monomials with indexes in $\mathfrak{K}_4$. Abusing notation, we define analogously $\Pi_{\mathfrak{K}_2}$ as the projection onto monomials $e^{\mathrm{i}\ell \cdot \theta} \mathcal{Y}^l a_{\vec{j}_1}^{\sigma_1} a_{\vec{j}_2}^{\sigma_2}$ with $|l| = 1$ and $(\vec{j}_1, \vec{j}_2, \ell, \sigma_1, \sigma_2) \in \mathfrak{K}_2$.
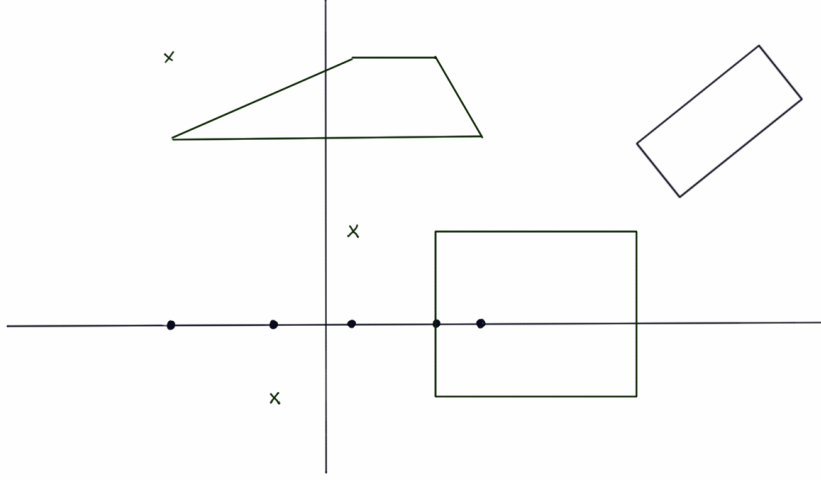
FIGURE 1. The black dots are the points in $\mathcal{S}_0$. The two rectangles and the trapezoid correspond to cases 1,2,4 in $\mathfrak{R}_4$. In order to represent case 3, we have highlighted three points in $\mathcal{S}$. To each such triple we may associate at most one $\ell \neq 0$ and one $\vec{\jmath}_4 \in \mathbf{Z}$, which form a resonance of type 3.

Theorem 6.2. *There exist $0 < r_2 \leq r_1$, $0 < \rho_2 \leq \rho_1$ such that for all $0 < \varepsilon \leq \varepsilon_2$, for all $\lambda \in \mathcal{C}^{(2)}$ and for all $r \in [0, r_2]$, $\rho \in [\frac{\rho_2}{2}, \rho_2]$ there exists a symplectic change of variables $\mathcal{L}^{(2)}$ well defined and majorant analytic from $D(\rho/2, r/2) \to D(\rho, r)$ such that*

$$(6.3) \qquad \mathcal{Q} \circ \mathcal{L}^{(2)}(\mathcal{Y}, \theta, \mathbf{a}) = \omega \cdot \mathcal{Y} + \sum_{\vec{\jmath} \in \mathbb{Z}^2 \setminus \mathcal{S}_0} \Omega_{\vec{\jmath}}(\lambda, \varepsilon) |a_{\vec{\jmath}}|^2 + \mathcal{Q}^{(2)}_{\mathrm{Res}} + \widetilde{\mathcal{Q}}^{(\geq 3)}$$

where

$$(6.4) \qquad \mathcal{Q}^{(2)}_{\mathrm{Res}} = \Pi_{\mathfrak{R}_4} \mathcal{Q}^{(2)} + \Pi_{\mathfrak{R}_2} \mathcal{Q}^{(2)}$$

with $\mathfrak{R}_4$ defined in (6.2), $\mathfrak{R}_2$ defined in (4.4) and

$$|\mathcal{Q}^{(2)}_{\mathrm{Res}}|_{\rho/2, r/2} \lesssim r^2 , \qquad |\widetilde{\mathcal{Q}}^{(\geq 3)}|_{\rho/2, r/2} \lesssim \frac{r^3}{\sqrt{\varepsilon}}.$$

Moreover $\mathcal{L}^{(2)}$ maps $D(\rho/2, r/2) \cap h^1 \to D(\rho, r) \cap h^1$, and if we denote $(\widetilde{\mathcal{Y}}, \widetilde{\theta}, \widetilde{\mathbf{a}}) = \mathcal{L}^{(2)}(\mathcal{Y}, \theta, \mathbf{a})$, then

$$\|\widetilde{\mathbf{a}} - \mathbf{a}\|_{\ell^1} \lesssim \|\mathbf{a}\|_{\ell^1}^3.$$

Proof. The proof is analogous to the one of Theorem 5.2, and we skip it. $\qquad\qquad \square$

7. CONSTRUCTION OF THE TOY MODEL

Once we have performed (partial) Birkhoff normal form up to order 4, we can start applying the ideas developed in [CKS+10] to Hamiltonian (6.3). Note that throughout this section $\varepsilon > 0$ is a fixed parameter. Namely, we do not use its smallness and we do not modify it.

We first perform the (time dependent) change of variables to rotating coordinates

$$(7.1) \qquad a_{\vec{\jmath}} = \beta_{\vec{\jmath}} \, e^{i\Omega_{\vec{\jmath}}(\lambda, \varepsilon) t},$$

to the Hamiltonian (6.3), which leads to the corrected Hamiltonian

$$(7.2) \qquad \mathcal{Q}_{\mathrm{rot}}(\mathcal{Y}, \theta, \beta, t) = \mathcal{Q} \circ \mathcal{L}^{(2)} \left(\mathcal{Y}, \theta, \{\beta_{\vec{\jmath}} \, e^{i\Omega_{\vec{\jmath}}(\lambda, \varepsilon) t}\}_{\vec{\jmath} \in \mathbb{Z}_N^2 \setminus \mathcal{S}_0} \right) - \sum_{\vec{\jmath} \in \mathbb{Z}_N^2 \setminus \mathcal{S}_0} \Omega_{\vec{\jmath}}(\lambda, \varepsilon) |\beta_{\vec{\jmath}}|^2.$$

We split this Hamiltonian as a suitable first order truncation $\mathcal{G}$ plus two remainders,

$$\mathcal{Q}_{\text{rot}}(\mathcal{Y}, \theta, \beta, t) = \mathcal{G}(\mathcal{Y}, \theta, \beta) + \mathcal{J}_1(\mathcal{Y}, \theta, \beta, t) + \mathcal{R}(\mathcal{Y}, \theta, \beta, t)$$

with

$$\mathcal{G}(\mathcal{Y}, \theta, \beta) = \omega \cdot \mathcal{Y} + \mathcal{Q}_{\text{Res}}^{(2)}(\mathcal{Y}, \theta, \beta)$$

(7.3)
$$\mathcal{J}_1(\mathcal{Y}, \theta, \beta, t) = \mathcal{Q}_{\text{Res}}^{(2)} \left(\mathcal{Y}, \theta, \{\beta_{\vec{j}}\, e^{\mathrm{i}\Omega_{\vec{j}}(\lambda, \varepsilon) t}\}_{\vec{j} \in \mathbb{Z}_N^2 \setminus \mathcal{S}_0} \right) - \mathcal{Q}_{\text{Res}}^{(2)}(\mathcal{Y}, \theta, \beta)$$

$$\mathcal{R}(\mathcal{Y}, \theta, \beta, t) = \widetilde{\mathcal{Q}}^{(\geq 3)} \left(\mathcal{Y}, \theta, \{\beta_{\vec{j}}\, e^{\mathrm{i}\Omega_{\vec{j}}(\lambda, \varepsilon) t}\}_{\vec{j} \in \mathbb{Z}_N^2 \setminus \mathcal{S}_0} \right)$$

where $\mathcal{Q}_{\text{Res}}^{(2)}$ and $\widetilde{\mathcal{Q}}^{(\geq 3)}$ are the Hamiltonians introduced in Theorem 6.2.

For the rest of this section we focus our study on the truncated Hamiltonian $\mathcal{G}$. Note that the remainder $\mathcal{J}_1$ is not smaller than $\mathcal{G}$. Nevertheless it will be smaller when evaluated on the particular solutions we consider. The term $\mathcal{R}$ is smaller than $\mathcal{G}$ for small data since it is the remainder of the normal form obtained in Theorem 6.2. Later in Section 8 we show that including the dismissed terms $\mathcal{J}_1$ and $\mathcal{R}$ barely alters the dynamics of the solutions of $\mathcal{G}$ that we analyze.

7.1. The finite set Λ.
We now start constructing special dynamics for the Hamiltonian $\mathcal{G}$ with the aim of treating the contributions of $\mathcal{J}_1$ and $\mathcal{R}$ as remainder terms. Following [CKS+10], we do not study the full dynamics of $\mathcal{G}$ but we restrict the dynamics to invariant subspaces. Indeed, we shall construct a set $\Lambda \subset \mathbf{z} := (\mathbb{Z} \times N\mathbb{Z}) \setminus (\mathcal{S}_0 \cup \mathcal{S})$ for some large N, in such a way that it generates an invariant subspace (for the dynamics of $\mathcal{G}$) given by

(7.4)
$$U_\Lambda := \{\beta_{\vec{j}} = 0 : \vec{j} \notin \Lambda\}.$$

Thus, we consider the following definition.

Definition 7.1 (Completeness). *We say that a set $\Lambda \subset \mathbf{z}$ is* complete *if U_Λ is invariant under the dynamics of $\mathcal{G}$.*

Remark 7.2. *It can be easily seen that if Λ is complete, U_Λ is also invariant under the dynamics of $\mathcal{G} + \mathcal{J}_1$.*

We construct a complete set $\Lambda \subset \mathbf{z}$ (see Definition 7.1) and we study the restriction on it of the dynamics of the Hamiltonian $\mathcal{G}$ in (7.3). Following [CKS+10], we impose several conditions on Λ to obtain dynamics as simple as possible.

The set Λ is constructed in two steps. First we construct a preliminary set $\Lambda_0 \subset \mathbb{Z}^2$ on which we impose numerous geometrical conditions. Later on we scale Λ_0 by a factor N to obtain $\Lambda \subset (N\mathbb{Z} \times N\mathbb{Z}) \subset \mathbf{z}$.

The set Λ_0 is "essentially" the one described in [CKS+10]. The crucial point in that paper is to choose carefully the modes so that each mode in Λ_0 only belongs to two rectangles with vertices in Λ_0. This allows to simplify considerably the dynamics and makes it easier to analyze. Certainly, this requires imposing several conditions on Λ_0. We add some extra conditions to adapt the set Λ_0 to the particular setting of the present paper.

We start by describing them. We split Λ_0 into $\mathfrak{g}$ disjoint generations $\Lambda_0 = \Lambda_{01} \cup \ldots \cup \Lambda_{0\mathfrak{g}}$. We call a quadruplet $(\vec{j}_1, \vec{j}_2, \vec{j}_3, \vec{j}_4) \in \Lambda_0^4$ a *nuclear family* if $\vec{j}_1, \vec{j}_3 \in \Lambda_{0k}$, $\vec{j}_2, \vec{j}_4 \in \Lambda_{0,k+1}$, and the four vertices form a non-degenerate rectangle. Then, we require the following conditions.

- Property I_{Λ_0} (Closure): If $\vec{j}_1, \vec{j}_2, \vec{j}_3 \in \Lambda_0$ are three vertices of a rectangle, then the fourth vertex of that rectangle is also in Λ_0.
- Property II_{Λ_0} (Existence and uniqueness of spouse and children): For each $1 \leq k < \mathfrak{g}$ and every $\vec{j}_1 \in \Lambda_{0k}$, there exists a unique spouse $\vec{j}_3 \in \Lambda_{0k}$ and unique (up to trivial permutations) children $\vec{j}_2, \vec{j}_4 \in \Lambda_{0,k+1}$ such that $(\vec{j}_1, \vec{j}_2, \vec{j}_3, \vec{j}_4)$ is a nuclear family in Λ_0.
- Property III_{Λ_0} (Existence and uniqueness of parents and siblings): For each $1 \leq k < \mathfrak{g}$ and every $\vec{j}_2 \in \Lambda_{0,k+1}$ there exists a unique sibling $\vec{j}_4 \in \Lambda_{0,k+1}$ and unique (up to permutation) parents $\vec{j}_1, \vec{j}_3 \in \Lambda_{0k}$ such that $(\vec{j}_1, \vec{j}_2, \vec{j}_3, \vec{j}_4)$ is a nuclear family in Λ_0.
- Property IV_{Λ_0} (Non-degeneracy): A sibling of any frequency $\vec{j}$ is never equal to its spouse.

- Property V_{Λ_0} (Faithfulness): Apart from nuclear families, Λ_0 contains no other rectangles. In fact, by the closure property I_{Λ_0}, this also means that it contains no right angled triangles other than those coming from vertices of nuclear families.
- Property VI_{Λ_0}: There are no two elements in Λ_0 such that $\vec{j}_1 \pm \vec{j}_2 = 0$. There are no three elements in Λ_0 such that $\vec{j}_1 - \vec{j}_2 + \vec{j}_3 = 0$. If four points in Λ_0 satisfy $\vec{j}_1 - \vec{j}_2 + \vec{j}_3 - \vec{j}_4 = 0$ then either the relation is trivial or such points form a family.
- Property VII_{Λ_0}: There are no points in Λ_0 with one of the coordinates equal to zero i.e.

$$\Lambda_0 \cap \big(\mathbb{Z} \times \{0\} \cup \{0\} \times \mathbb{Z}\big) = \emptyset.$$

- Property $VIII_{\Lambda_0}$: There are no two points in Λ_0 which form a right angle with 0.

Condition I_{Λ_0} is just a rephrasing of the completeness condition introduced in Definition 7.1. Properties II_{Λ_0}, III_{Λ_0}, IV_{Λ_0}, V_{Λ_0} correspond to being a family tree as stated in [CKS$^+$10].

Theorem 7.3. *Fix* $K \gg 1$ *and* $s \in (0,1)$. *Then, there exists* $\mathfrak{g} \gg 1$, $A_0 \gg 1$, $\eta > 0$, *and a set* $\Lambda_0 \subset \mathbb{Z}^2$ *with*

$$\Lambda_0 = \Lambda_{01} \cup \ldots \cup \Lambda_{0\mathfrak{g}},$$

which satisfies conditions I_{Λ_0} *–* $VIII_{\Lambda_0}$ *and also*

$$(7.5) \qquad \frac{\sum_{\vec{j} \in \Lambda_{0,\mathfrak{g}-1}} |\vec{j}|^{2s}}{\sum_{\vec{j} \in \Lambda_{03}} |\vec{j}|^{2s}} \geq \frac{1}{2} 2^{(1-s)(\mathfrak{g}-4)} \geq K^2.$$

Moreover, for any $A \geq A_0$, *there exist* $\mathfrak{g}$ *and a function* $f(\mathfrak{g})$ *satisfying*

$$(7.6) \qquad e^{A^{\mathfrak{g}}} \leq f(\mathfrak{g}) \leq e^{2(1+\eta)A^{\mathfrak{g}}} \qquad \text{for } \mathfrak{g} \text{ large enough,}$$

such that each generation Λ_{0k} *has* $2^{\mathfrak{g}-1}$ *disjoint frequencies* $\vec{j}$ *satisfying*

$$(7.7) \qquad C^{-1} f(\mathfrak{g}) \leq |\vec{j}| \leq C 3^{\mathfrak{g}} f(\mathfrak{g}), \quad \vec{j} \in \Lambda_{0k},$$

and

$$(7.8) \qquad \frac{\sum_{\vec{j} \in \Lambda_{0k}} |\vec{j}|^{2s}}{\sum_{\vec{j} \in \Lambda_{0i}} |\vec{j}|^{2s}} \leq C e^{s\mathfrak{g}}$$

for any $1 \leq i < k \leq \mathfrak{g}$ *and some constant* $C > 0$ *independent of* $\mathfrak{g}$.

The construction of such kind of sets was done first in [CKS$^+$10] (see also [GK15, GK17, Gua14, GHP16]) where the authors construct sets Λ satisfying Properties I_Λ-V_Λ and estimate (7.8). The proof of Theorem 7.3 follows the same lines as the ones in those papers. Indeed, Properties VI_Λ-$VIII_\Lambda$ can be obtained through the same density argument. Finally, the estimate (7.7), even if it is not stated explicitly in [CKS$^+$10], it is an easy consequence of the proof in that paper (in [GK15, GK17, GHP16] a slightly weaker estimate is used).

Remark 7.4. *Note that* $s \in (0,1)$ *implies that were are constructing a backward cascade orbit (energy is transferred from high to low modes). This means that the modes in each generation of* Λ_0 *are just switched oppositely* $\Lambda_{0j} \leftrightarrow \Lambda_{0,\mathfrak{g}-j+1}$ *compared to the ones constructed in* [CKS$^+$10]. *The second statement of Theorem 1.2 considers* $s > 1$ *and therefore a forward cascade orbit (energy transferred from low to high modes). For this result, we need a set* Λ_0 *of the same kind as that of* [CKS$^+$10], *which thus satisfies*

$$\frac{\sum_{\vec{j} \in \Lambda_{0,\mathfrak{g}-1}} |\vec{j}|^{2s}}{\sum_{\vec{j} \in \Lambda_{03}} |\vec{j}|^{2s}} \geq \frac{1}{2} 2^{(s-1)(\mathfrak{g}-4)} \geq K^2$$

instead of estimate (7.5).

We now scale Λ_0 by a factor N satisfying (4.1) and we denote by $\Lambda := N\Lambda_0$. Note that the listed properties I_{Λ_0} – $VIII_{\Lambda_0}$ are invariant under scaling. Thus, if they are satisfied by Λ_0, they are satisfied by Λ too.

Lemma 7.5. *There exists a set* Λ *satisfying all statements of Theorem 7.3 (with a different* $f(\mathfrak{g})$ *satisfying* (7.6)) *and also the following additional properties.*

(1) *If two points* $\vec{j}_1, \vec{j}_2 \in \Lambda$ *form a right angle with a point* $(m,0) \in \mathbb{Z} \times \{0\}$, *then*

$$|m| \geq \sqrt{f(\mathfrak{g})}.$$

(2) $\Lambda \subset N\mathbb{Z} \times N\mathbb{Z}$ with

$$N = f(\mathfrak{g})^{\frac{4}{5}}.$$

Proof. Consider any of the sets Λ obtained in Theorem 7.3. By property VIII_{Λ_0} one has $m \neq 0$. Define $\vec{j}_3 = (m, 0)$. The condition for orthogonality is either

$$(i) \ (\vec{j}_1 - \vec{j}_2) \cdot (\vec{j}_3 - \vec{j}_2) = 0 \quad \text{or} \quad (ii) \ (\vec{j}_1 - \vec{j}_3) \cdot (\vec{j}_2 - \vec{j}_3) = 0 \ .$$

Taking $\vec{j}_i = (m_i, n_i)$, $i = 1, 2$, condition (i) implies (after some computations) that m is given by

$$m = \frac{(n_1 - n_2)n_2 + (m_1 - m_2)m_2}{m_1 - m_2} \ .$$

Then since $|m_1 - m_2| \leq 2Cf(\mathfrak{g})3^{\mathfrak{g}}$ and the numerator is not zero, we have

$$(7.9) \qquad\qquad |m| \geq \frac{1}{4Cf(\mathfrak{g})3^{\mathfrak{g}}} \geq \frac{1}{(f(\mathfrak{g}))^{3/2}}.$$

Now we consider condition (ii). One gets that m is a root of the quadratic equation

$$m^2 - (m_1 + m_2)m + (m_1 m_2 + n_1 n_2) = 0 \ .$$

First we note that $m_1 m_2 + n_1 n_2 \neq 0$ by property VIII_{Λ_0}, since $m = 0$ cannot be a solution. Now consider the discriminant $\Delta = (m_1 + m_2)^2 - 4(m_1 m_2 + n_1 n_2)$. If $\Delta < 0$, then no right angle is possible. If $\Delta = 0$, then clearly $|m| \geq 1/2$, since once again $m = 0$ is not a solution. Finally let $\Delta > 0$. Then

$$m = \frac{(m_1 + m_2)}{2}\left(1 \pm \sqrt{1 - \frac{4(m_1 m_2 + n_1 n_2)}{(m_1 + m_2)^2}}\right) \ .$$

Denoting by $\gamma := \frac{4(m_1 m_2 + n_1 n_2)}{(m_1 + m_2)^2}$, the condition $\Delta > 0$ implies that $-\infty < \gamma < 1$. Splitting in two cases: $|\gamma| \leq 1$ and $\gamma < -1$ one can easily obtain that either way m satisfies (7.9). Now it only remains to scale the set Λ by a factor $(f(\mathfrak{g}))^4$. Then, taking as new $f(\mathfrak{g})$, $\widetilde{f}(\mathfrak{g}) := (f(\mathfrak{g}))^5$, the obtained set Λ satisfies all statements of Theorem 7.3 and also the statements of Lemma 7.5. $\qquad\square$

7.2. **The truncated Hamiltonian on the finite set Λ and the [CKS$^+$10] toy model.** We use the properties of the set Λ given by Theorem 7.3 and Lemma 7.5 to compute the restriction of the Hamiltonian $\mathcal{G}$ in (7.3) to the invariant subset U_Λ (see (7.4)).

Lemma 7.6. *Consider the set $\Lambda \subset N\mathbb{Z} \times N\mathbb{Z}$ obtained in Theorem 7.3. Then, the set*

$$\mathcal{M}_\Lambda = \{(\mathcal{Y}, \theta, \beta) : \mathcal{Y} = 0, \quad \beta \in U_\Lambda\}$$

is invariant under the flow associated to the Hamiltonian $\mathcal{G}$. Moreover, $\mathcal{G}$ restricted to $\mathcal{M}_\Lambda$ can be written as

$$(7.10) \qquad\qquad \mathcal{G}\big|_{\mathcal{M}_\Lambda}(\theta, \beta) = \mathcal{G}_0(\beta) + \mathcal{J}_2(\theta, \beta)$$

where

$$(7.11) \qquad\qquad \mathcal{G}_0(\beta) = -\frac{1}{2}\sum_{\vec{j} \in \Lambda}|\beta_{\vec{j}}|^4 + \frac{1}{2}\sum_{\substack{(\vec{j}_1, \vec{j}_2, \vec{j}_3, \vec{j}_4) \in \Lambda^4 \\ \vec{j}_i \text{ form a rectangle}}}^{*} \beta_{\vec{j}_1}\bar{\beta}_{\vec{j}_2}\beta_{\vec{j}_3}\bar{\beta}_{\vec{j}_4}$$

and the remainder $\mathcal{J}_2$ satisfies

$$(7.12) \qquad\qquad |\mathcal{J}_2|_{\rho, r} \lesssim r^2(f(\mathfrak{g}))^{-\frac{4}{5}}.$$

Proof. First we note that, since $\mathcal{Y} = 0$ on $\mathcal{M}_\Lambda$,

$$\mathcal{G}\big|_{\mathcal{M}_\Lambda} = \mathcal{Q}^{(2)}_{\text{Res}}\big|_{\mathcal{M}_\Lambda} = \Pi_{\mathfrak{K}_4}\mathcal{Q}^{(2)}\big|_{\mathcal{M}_\Lambda}$$

where $\mathcal{Q}^{(2)}_{\text{Res}}$ is the Hamiltonian defined in Theorem 6.2. We start by analyzing the Hamiltonian $\mathcal{Q}^{(2)}$ introduced in Theorem 5.2, which is defined as

$$\mathcal{Q}^{(2)} = \mathcal{K}^{(2)} + \frac{1}{2}\left\{\mathcal{K}^{(1)}, \chi^{(1)}\right\} \ .$$

We analyze each term. Here it plays a crucial role that $\Lambda \subset N\mathbb{Z} \times N\mathbb{Z}$ with $N = f(\mathfrak{g})^{4/5}$.

In order to estimate $\mathcal{K}^{(2)}$, defined in (5.1), we recall that Λ does not have any mode in the x-axis and therefore the original quartic Hamiltonian has not been modified by the Birkhoff map (1.6) (this is evident from the formula for $\mathcal{H}^{(2)}$ in (3.18)). Thus, it is enough to analyze how the quartic Hamiltonian has been modified by the linear change $\mathcal{L}^{(0)}$ analyzed in Theorems 4.3 and 4.4. Using the smoothing property of the change of coordinates $\mathcal{L}^{(0)}$ given in Statement 5 of Theorem 4.3, one obtains

$$\Pi_{\mathfrak{R}_4}\mathcal{K}^{(2)}\big|_{\mathcal{M}_\Lambda} = -\frac{1}{2}\sum_{\vec{j}\in\Lambda}|a_{\vec{j}}|^4 + \frac{1}{2}\sum_{\text{Rectangles}\subset\Lambda} a_{\vec{j}_1}\bar{a}_{\vec{j}_2}a_{\vec{j}_3}\bar{a}_{\vec{j}_4} + O\left(\frac{r^2}{N}\right).$$

Now we deal with the term $\{\mathcal{K}^{(1)}, \chi^{(1)}\}$. Since we only need to analyze $\Pi_{\mathfrak{R}_4}\{\mathcal{K}^{(1)}, \chi^{(1)}\}\big|_{\mathcal{M}_\Lambda}$, we only need to consider monomials in $\mathcal{K}^{(1)}$ and in $\chi^{(1)}$ which have at least two indexes in Λ. We represent this by setting

$$\chi^{(1)} = \chi^{(1)}_{\#\Lambda\leq 1} + \chi^{(1)}_{\#\Lambda\geq 2}\,,$$

where $\#\Lambda \geq 2$ means that we restrict to those monomials which have at least two indexes in Λ. We then have

$$\{\mathcal{K}^{(1)}, \chi^{(1)}\}\big|_{\mathcal{M}_\Lambda} = \{\mathcal{K}^{(1)}, \chi^{(1)}_{\#\Lambda\geq 2}\}\big|_{\mathcal{M}_\Lambda}.$$

We estimate the size of $\chi^{(1)}_{\#\Lambda\geq 2}$. As explained in the proof of Theorem 5.2, $\chi^{(1)}_{\#\Lambda\geq 2}$ has coefficients

(7.13) $$\chi^{(1)}_{\ell,\mathbf{j},\vec{\sigma}} = \frac{i\mathcal{K}^{(1)}_{\ell,\mathbf{j},\vec{\sigma}}}{\omega\cdot\ell + \sigma_1\Omega_{\vec{j}_1}(\lambda,\varepsilon) + \sigma_2\Omega_{\vec{j}_2}(\lambda,\varepsilon) + \sigma_3\Omega_{\vec{j}_3}(\lambda,\varepsilon)}$$

with $\vec{j}_2, \vec{j}_3 \in \Lambda$.

We first estimate the tails (in ℓ) of $\chi^{(1)}$ and then we analyze the finite number of cases left. For the tails, it is enough to use Theorem 5.2 to deduce the following estimate for any $\rho \leq \rho_1/2$, where ρ_1 is the constant introduced in that theorem,

$$\left|\sum_{|\ell|>\sqrt[4]{N}} \chi^{(1)}_{\ell,\mathbf{j},\vec{\sigma}}\, e^{i\theta\cdot\ell}a^{\sigma_1}_{\vec{j}_1}a^{\sigma_2}_{\vec{j}_2}a^{\sigma_3}_{\vec{j}_3}\right|^{\mathcal{C}^{(1)}}_{\rho,r} \lesssim e^{-(\rho_1-\rho)\sqrt[4]{N}}\left|\chi^{(1)}\right|^{\mathcal{C}^{(1)}}_{\rho_1,r} \leq re^{-(\rho_1-\rho)\sqrt[4]{N}}.$$

We restrict our attention to monomials with $|\ell| \leq \sqrt[4]{N}$. We take $\vec{j}_2, \vec{j}_3 \in \Lambda$ and we consider different cases depending on $\vec{j}_1$ and the properties of the monomial. In each case we show that the denominator of (7.13) is larger than N.

Case 1. Suppose that $\vec{j}_1 \notin \mathscr{S}$. The selection rules are (according to Remark 4.7)

$$\eta(\ell) + \sigma_1 + \sigma_2 + \sigma_3 = 0\,, \quad \vec{m}\cdot\ell + \sigma_1 m_1 + \sigma_2 m_2 + \sigma_3 m_3 = 0\,, \quad \sigma_1 n_1 + \sigma_2 n_2 + \sigma_3 n_3 = 0$$

and the leading term in the denominator of (7.13) is

(7.14) $$\vec{m}^2\cdot\ell + \sigma_1|\vec{j}_1|^2 + \sigma_2|\vec{j}_2|^2 + \sigma_3|\vec{j}_3|^2$$

where $\vec{m}^2 = (m_1^2, \ldots, m_{\mathtt{d}}^2)$. We consider the following subcases:

A1 $\sigma_3 = \sigma_1 = +1$, $\sigma_2 = -1$. In this case $\vec{j}_1 - \vec{j}_2 + \vec{j}_3 - \mathtt{v} = 0$, where $\mathtt{v} := (-\vec{m}\cdot\ell, 0)$. We rewrite (7.14) as

$$\vec{m}^2\cdot\ell + (\vec{m}\cdot\ell)^2 - (\vec{m}\cdot\ell)^2 + |\vec{j}_1|^2 - |\vec{j}_2|^2 + |\vec{j}_3|^2 = \vec{m}^2\cdot\ell + (\vec{m}\cdot\ell)^2 - 2\Big(\mathtt{v} - \vec{j}_3, \vec{j}_3 - \vec{j}_2\Big).$$

Assume first $\vec{j}_2 \neq \vec{j}_3$. Since the set Λ satisfies Lemma 7.5 1. and $|\vec{m}\cdot\ell| \lesssim \sqrt[4]{N} \lesssim f(\mathfrak{g})^{1/5}$, we can ensure that $\vec{j}_2$ and $\vec{j}_3$ do not form a right angle with $\mathtt{v}$, thus

$$\Big(\mathtt{v} - \vec{j}_2, \vec{j}_3 - \vec{j}_2\Big) \in \mathbb{Z}\setminus\{0\}.$$

Actually by the second statement of Lemma 7.5, $\vec{j}_3 - \vec{j}_2 \in N\mathbb{Z}^2$ and hence, using also $|\ell| \leq \sqrt[4]{N}$,

$$\left|\vec{m}^2\cdot\ell + (\vec{m}\cdot\ell)^2 - 2\Big(\mathtt{v} - \vec{j}_3, \vec{j}_3 - \vec{j}_2\Big)\right| \geq 2N - N/8 > N.$$

Now it remains the case $\vec{\jmath}_2 = \vec{\jmath}_3$. Such monomials cannot exist in $\mathcal{H}^{(1)}$ in (3.17) since the monomials with two equal modes have been removed in (3.3) (it does not support degenerate rectangles). Naturally a degenerate rectangle may appear after we apply the change $\mathcal{L}^{(0)}$ introduced in Theorem 4.3. Nevertheless, the map $\mathcal{L}^{(0)}$ is identity plus smoothing (see statement 5 of that theorem), which leads to the needed N^{-1} factor.

B1 $\sigma_3 = \sigma_2 = +1$, $\sigma_1 = -1$. Now the selection rule reads $-\vec{\jmath}_1 + \vec{\jmath}_2 + \vec{\jmath}_3 - \mathbf{v} = \mathbf{0}$, with again $\mathbf{v} = (-\vec{\mathbf{m}} \cdot \ell, 0)$. We rewrite (7.14) as

$$\vec{\mathbf{m}}^2 \cdot \ell + (\vec{\mathbf{m}} \cdot \ell)^2 - (\vec{\mathbf{m}} \cdot \ell)^2 - |\vec{\jmath}_1|^2 + |\vec{\jmath}_2|^2 + |\vec{\jmath}_3|^2 = \vec{\mathbf{m}}^2 \cdot \ell + (\vec{\mathbf{m}} \cdot \ell)^2 - 2\left(\mathbf{v} - \vec{\jmath}_3, \mathbf{v} - \vec{\jmath}_2\right).$$

By the first statement of Lemma 7.5, $\left(\mathbf{v} - \vec{\jmath}_2, \mathbf{v} - \vec{\jmath}_3\right) \neq 0$. By Property VIII$_\Lambda$ and the second statement of Lemma 7.5, one has $|(\vec{\jmath}_2, \vec{\jmath}_3)| \geq N^2$ and estimate (7.7) implies $|\vec{\jmath}_2|, |\vec{\jmath}_3| \leq N^{3/2}$. Then

$$\left|\left(\mathbf{v} - \vec{\jmath}_2, \mathbf{v} - \vec{\jmath}_3\right)\right| \geq |(\vec{\jmath}_2, \vec{\jmath}_3)| - |(\mathbf{v}, \vec{\jmath}_2 + \vec{\jmath}_3)| - |\mathbf{v}|^2 \geq N^2/4$$

and one concludes as in **A1**.

C1 $\sigma_1 = \sigma_3 = \sigma_2 = +1$. The denominator (7.14) satisfies

$$|\vec{\mathbf{m}}^2 \cdot \ell + |\vec{\jmath}_1|^2 + |\vec{\jmath}_2|^2 + |\vec{\jmath}_3|^2| \geq 2N - |\vec{\mathbf{m}}^2 \cdot \ell| \geq 2N - N/8 \geq N.$$

This completes the proof of Case 1.

Case 2. Suppose that $\vec{\jmath}_1 \in \mathscr{S}$. The selection rules are

$$\eta(\ell) + \sigma_2 + \sigma_3 = 0, \quad \vec{\mathbf{m}} \cdot \ell + \sigma_2 m_2 + \sigma_3 m_3 = 0, \quad \sigma_1 n_1 + \sigma_2 n_2 + \sigma_3 n_3 = 0$$

and the leading term in the denominator is

$$(7.15) \qquad\qquad \vec{\mathbf{m}}^2 \cdot \ell + \sigma_1 n_1^2 + \sigma_2 |\vec{\jmath}_2|^2 + \sigma_3 |\vec{\jmath}_3|^2,$$

where $\vec{\mathbf{m}}^2 = (\mathbf{m}_1^2, \ldots, \mathbf{m}_{\mathbf{d}}^2)$. We can reduce Case 2 to Case 1.

B2 $\sigma_2 = \sigma_3 = +1$, $\sigma_1 = -1$. Assume w.l.o.g. that $\vec{\jmath}_1 = (\mathbf{m}_1, n_1)$. Define $\tilde{\ell} = \ell + \mathbf{e}_1$, and obtain from the selection rules and (7.15) that

$$\vec{\mathbf{m}} \cdot \tilde{\ell} - \mathbf{m}_1 + m_2 + m_3 = \vec{\mathbf{m}} \cdot \ell + m_2 + m_3 = 0.$$

Then the leading term in the denominator becomes

$$\vec{\mathbf{m}}^2 \cdot \tilde{\ell} - (\mathbf{m}_1^2 + n_1^2) + |\vec{\jmath}_2|^2 + |\vec{\jmath}_3|^2$$

and one proceeds as in case **B1** with $\tilde{\ell}$ in place of ℓ.

The cases **A2** and **C2** are completely equivalent.

In conclusion we have proved that

$$(7.16) \qquad\qquad \left|\chi_{\#\Lambda \geq 2}^{(1)}\big|_{\mathcal{M}_\Lambda}\right|_{\rho, r}^{\mathcal{C}^{(1)}} \leq rN^{-1}.$$

Item (i) of Lemma 5.3, jointly with estimate (7.16), implies that, for $\rho' \in (0, \rho/2]$ and $r' \in (0, r/2]$

$$\left|\left\{\mathcal{K}^{(1)}, \chi_{\#\Lambda \geq 2}^{(1)}\right\}\big|_{\mathcal{M}_\Lambda}\right|_{\rho', r'}^{\mathcal{C}^{(1)}} \lesssim r^2 N^{-1}.$$

This completes the proof of Lemma 7.6. □

The Hamiltonian $\mathcal{G}_0$ in (7.11) is the Hamiltonian that the I-team derived to construct their toy model. A posteriori we will check that the remainder $\mathcal{J}_2$ plays a small role in our analysis.

The properties of Λ imply that the equation associated to $\mathcal{G}_0$ reads

$$(7.17) \qquad\qquad \mathrm{i}\dot{\beta}_{\vec{\jmath}} = -\beta_{\vec{\jmath}}|\beta_{\vec{\jmath}}|^2 + 2\beta_{\vec{\jmath}_{\text{child}_1}}\beta_{\vec{\jmath}_{\text{child}_2}}\overline{\beta_{\vec{\jmath}_{\text{spouse}}}} + 2\beta_{\vec{\jmath}_{\text{parent}_1}}\beta_{\vec{\jmath}_{\text{parent}_2}}\overline{\beta_{\vec{\jmath}_{\text{sibling}}}}$$

for each $\vec{\jmath} \in \Lambda$. In the first and last generations, the parents and children are set to zero respectively. Moreover, the particular form of this equation implies the following corollary.

Corollary 7.7 ([CKS$^+$10]). *Consider the subspace*

$$\widetilde{U}_\Lambda = \{\beta \in U_\Lambda : \beta_{\vec{j}_1} = \beta_{\vec{j}_2} \text{ for all } \vec{j}_1, \vec{j}_2 \in \Lambda_k \text{ for some } k\},$$

where all the members of a generation take the same value. Then, $\widetilde{U}_\Lambda$ is invariant under the flow associated to the Hamiltonian $\mathcal{G}_0$. Therefore, equation (7.17) restricted to $\widetilde{U}_\Lambda$ becomes

$$(7.18) \qquad\qquad i\dot{b}_k = -b_k^2\bar{b}_k + 2\bar{b}_k\left(b_{k-1}^2 + b_{k+1}^2\right), \ k = 1,\dots\mathfrak{g},$$

where

$$(7.19) \qquad\qquad b_k = \beta_{\vec{j}} \ \text{ for any } \vec{j} \in \Lambda_k.$$

The dimension of $\widetilde{U}_\Lambda$ is $2\mathfrak{g}$, where $\mathfrak{g}$ is the number of generations. In the papers [CKS$^+$10] and [GK15], the authors construct certain orbits of the *toy model* (7.18) which shift its mass from being localized at b_3 to being localized at $b_{\mathfrak{g}-1}$. These orbits will lead to orbits of the original equation (2D-NLS) undergoing growth of Sobolev norms.

Theorem 7.8 ([GK15]). *Fix a large $\gamma \gg 1$. Then for any large enough $\mathfrak{g}$ and $\mu = e^{-\gamma\mathfrak{g}}$, there exist an orbit of system (7.18) and $T_0 > 0$ such that*

$$\begin{aligned}|b_3(0)| &> 1 - \mu \\ |b_i(0)| &< \mu \qquad \text{for } i \neq 3\end{aligned} \quad \text{and} \quad \begin{aligned}|b_{\mathfrak{g}-1}(T_0)| &> 1 - \mu \\ |b_i(T_0)| &< \mu \qquad \text{for } i \neq \mathfrak{g}-1.\end{aligned}$$

Moreover, there exists a constant $C > 0$ independent of $\mathfrak{g}$ such that T_0 satisfies

$$0 < T_0 < C\mathfrak{g}\ln\left(\frac{1}{\mu}\right) = C\gamma\mathfrak{g}^2.$$

This theorem is proven in [CKS$^+$10] without time estimates. The time estimates were obtained in [GK15].

8. The approximation argument

In Sections 4, 5 and 6 we have applied several transformations and in Sections 6 and 7 we have removed certain small remainders. This has allowed us to derive a simple equation, called toy model in [CKS$^+$10]; then, in Section 7, we have analyzed some special orbits of this system. The last step of the proof of Theorem 1.2 is to show that when incorporating back the removed remainders ($\mathcal{J}_1$ and $\mathcal{R}$ in (7.3) and $\mathcal{J}_2$ in (7.10)) and undoing the changes of coordinates performed in Theorems 4.3 and 5.2, in Proposition 6.2 and in (7.1), the toy model orbit obtained in Theorem 7.8 leads to a solution of the original equation (2D-NLS) undergoing growth of Sobolev norms.

Now we analyze each remainder and each change of coordinates. From the orbit obtained in Theorem 7.8 and using (7.19) one can obtain an orbit of Hamiltonian (7.11). Moreover, both the equation of Hamiltonian (7.11) and (7.18) are invariant under the scaling

$$(8.1) \qquad\qquad b^\nu(t) = \nu^{-1}b\left(\nu^{-2}t\right).$$

By Theorem 7.8, the time spent by the solution $b^\nu(t)$ is

$$(8.2) \qquad\qquad T = \nu^2 T_0 \leq \nu^2 C\gamma\mathfrak{g}^2,$$

where T_0 is the time obtained in Theorem 7.8.

Now we prove that one can construct a solution of Hamiltonian (7.2) "close" to the orbit β^ν of Hamiltonian (7.11) defined as

$$(8.3) \qquad \begin{aligned} \beta_{\vec{j}}^\nu(t) &= \nu^{-1}b_k\left(\nu^{-2}t\right) \quad \text{for each} \ \ \vec{j} \in \Lambda_k \\ \beta_{\vec{j}}^\nu(t) &= 0 \ \ \text{for each} \ \ \vec{j} \notin \Lambda, \end{aligned}$$

where $b(t)$ is the orbit given by Theorem 7.8. Note that this implies incorporating the remainders in (7.3) and (7.10).

We take a large ν so that (8.3) is small. In the original coordinates this will correspond to solutions close to the finite gap solution. Taking $\mathcal{J} = \mathcal{J}_1 + \mathcal{J}_2$ (see (7.3) and (7.10)), the equations for β and $\mathcal{Y}$ associated to Hamiltonian (7.2) can be written as

(8.4)
$$
\begin{aligned}
\mathrm{i}\dot{\beta} &= \partial_{\overline{\beta}}\mathcal{G}_0(\beta) + \partial_{\overline{\beta}}\mathcal{J}(\mathcal{Y}, \theta, \beta) + \partial_{\overline{\beta}}\mathcal{R}(\mathcal{Y}, \theta, \beta) \\
\dot{\mathcal{Y}} &= -\partial_\theta \mathcal{J}(\mathcal{Y}, \theta, \beta) - \partial_\theta \mathcal{R}(\mathcal{Y}, \theta, \beta).
\end{aligned}
$$

Now we obtain estimates of the closeness of the orbit of the toy model obtained in Theorem 7.8 and orbits of Hamiltonian (7.2).

Theorem 8.1. *Fix* $0 < s_1 < s_2 < 1$. *Consider a solution* $(\mathcal{Y}, \theta, \beta) = (0, \theta_0, \beta^\nu(t))$ *of Hamiltonian* (7.11) *for any* $\theta_0 \in \mathbb{T}^{\mathrm{d}}$, *where* $\beta^\nu(t) = \{\beta^\nu_{\vec{\jmath}}(t)\}_{\vec{\jmath} \in \mathbb{Z}^2_N \setminus \mathcal{S}_0}$ *is the solution given by* (8.3). *Assume*

(8.5)
$$
f(\mathfrak{g})^{s_1} \leq \nu \leq f(\mathfrak{g})^{s_2}.
$$

Then, there exists σ *(depending on* s_1, s_2 *but independent* $\mathfrak{g}$ *and* γ*) such that any solution* $(\mathcal{Y}(t), \theta(t), \widetilde{\beta}(t))$ *of* (7.2) *with initial condition* $\widetilde{\beta}(0) = \widetilde{\beta}^0 \in \ell^1$, $\mathcal{Y}(0) = \mathcal{Y}^0 \in \mathbb{R}^{\mathrm{d}}$ *with* $\|\widetilde{\beta}^0 - \beta^\nu(0)\|_{\ell^1} \leq \nu^{-1-4\sigma}$ *and* $|\mathcal{Y}^0| \leq \nu^{-2-4\sigma}$ *and any* $\theta(0) = \theta_1 \in \mathbb{T}^{\mathrm{d}}$, *satisfies*

$$
\left\| \widetilde{\beta}_{\vec{\jmath}}(t) - \beta^\nu_{\vec{\jmath}}(t) \right\|_{\ell^1} \leq \nu^{-1-\sigma}, \qquad |\mathcal{Y}(t)| \leq \nu^{-2-\sigma},
$$

for $0 < t < T$, *where* T *is the time defined in* (8.2).

The proof of this theorem is deferred to Section 8.1. Note that the change to rotating coordinates in (7.1) does not alter the ℓ^1 norm and therefore a similar result as this theorem can be stated for orbits of Hamiltonian (6.3) (modulus adding the rotating phase).

Proof of Theorem 1.2. We use Theorem 8.1 to obtain a solution of Hamiltonian (3.14) undergoing growth of Sobolev norms. Then the same property will hold true for the corresponding solution of the Hamiltonian (3.3), by applying the inverse of the Birkhoff map Φ in Theorem 3.1, which leaves untouched the modes $a_{(m,n)}$ with $n \neq 0$. We consider the solution $(\mathcal{Y}^*(t), \theta^*(t), \mathbf{a}^*(t))$ of this Hamiltonian with initial condition

(8.6)
$$
\begin{aligned}
\mathcal{Y}^* &= 0 \\
\theta^* &= \theta_0 \\
a^*_{\vec{\jmath}} &= \nu^{-1} b_k(0) \qquad \text{for each } \vec{\jmath} \in \Lambda_k \\
a^*_{\vec{\jmath}} &= 0 \qquad\qquad\ \text{for each } \vec{\jmath} \notin \Lambda_k
\end{aligned}
$$

for an arbitrary choice of $\theta_0 \in \mathbb{T}^{\mathrm{d}}$. Since this initial condition has finite support, it follows by applying (the inverse of) the Birkhoff map Φ in Theorem 3.1 that the corresponding initial condition in the original coordinates $u_{\vec{\jmath}}$ belongs to h^s for all $s \geq 0$. Then the well-posedness properties of equation (2D-NLS) imply that the solution $u(t,x)$ with this initial condition belongs to $\cap_{s \geq 0} H^s(\mathbb{T}^2, \mathbb{C})$ for all times. Also, note that the distance of $u(0)$ from the torus $\mathcal{T}^I_{\mathcal{S}_0}$ is measured by the amplitude of $\{a^*_{\vec{\jmath}}\}$ (see (3.10)).

We need to prove that Theorem 8.1 applies to this solution with ν satisfying (8.5). To this end, we perform the changes of coordinates given in Theorems 4.3, 5.2 and 6.2, keeping track of the ℓ^1 norm.

For $\mathcal{L}^{(j)}$, $j = 1, 2$, Theorems 5.2 and 6.2 imply the following. Consider $(\mathcal{Y}, \theta, \mathbf{a}) \in D(\rho, r)$ and define $\pi_{\mathbf{a}}(\mathcal{Y}, \theta, \mathbf{a}) := \mathbf{a}$. Then, we have

(8.7)
$$
\left\| \pi_{\mathbf{a}} \mathcal{L}^{(j)}(\mathcal{Y}, \theta, \mathbf{a}) - \mathbf{a} \right\|_{\ell^1} \lesssim \|\mathbf{a}\|^2_{\ell^1}.
$$

This estimate is not true for the change of coordinates $\mathcal{L}^{(0)}$ given in Theorem 4.3. Nevertheless, this change is smoothing (see Statement 5 of Theorem 4.3). This implies that if all $\vec{\jmath} \in \mathrm{supp}\{\mathbf{a}\}$ satisfy $|\vec{\jmath}| \geq J$ then

(8.8)
$$
\left\| \pi_{\mathbf{a}} \mathcal{L}^{(0)}(\mathcal{Y}, \theta, \mathbf{a}) - \mathbf{a} \right\|_{\ell^1} \lesssim J^{-1} \|\mathbf{a}\|_{\ell^1}.
$$

Thanks to Theorem 7.3 (more precisely (7.7)), we can apply this estimate to (8.6) with $J = Cf(\mathfrak{g})$. Using the fact that $\|\mathbf{a}^*\|_{\ell^1} \lesssim \nu^{-1}\mathfrak{g}2^{\mathfrak{g}}$ and the condition on ν in (8.5) (which implies $f(\mathfrak{g})^{-1} \leq \nu^{-1}$), one can check

$$\left\|\pi_{\mathbf{a}}\mathcal{L}^{(0)}(0,\theta^*,\mathbf{a}^*) - \mathbf{a}^*\right\|_{\ell^1} \lesssim \nu^{-1}\mathfrak{g}2^{\mathfrak{g}}f(\mathfrak{g})^{-1} \lesssim \nu^{-3/2}.$$

Therefore, we can conclude

$$\left\|\pi_{\mathbf{a}}\left(\mathcal{L}^{(2)} \circ \mathcal{L}^{(1)} \circ \mathcal{L}^{(0)}(0,\theta^*,\mathbf{a}^*)\right) - \mathbf{a}^*\right\|_{\ell^1} \lesssim \nu^{-3/2}.$$

We define $(\widetilde{\mathcal{Y}}^*, \widetilde{\theta}^*, \widetilde{\mathbf{a}}^*)$ the image of the point (8.6) under the composition of these three changes. We apply Theorem 8.1 to the solution of (7.2) with this initial condition. Note that Theorem 8.1 is stated in rotating coordinates (see (7.1)). Nevertheless, since this change is the identity on the initial conditions, one does not need to make any further modification. Moreover, the change (7.1) leaves invariant both the ℓ^1 and Sobolev norms. We show that such solution $(\widetilde{\mathcal{Y}}^*(t), \widetilde{\theta}^*(t), \widetilde{\mathbf{a}}^*(t))$ expressed in the original coordinates satisfies the desired growth of Sobolev norms.

Define

$$S_i = \sum_{\vec{j} \in \Lambda_i} |\vec{j}|^{2s} \text{ for } i = 1, \ldots, \mathfrak{g}.$$

We first estimate the initial and final Sobolev norms of the solution $(\mathcal{Y}^*(t), \theta^*(t), \mathbf{a}^*(t))$ in terms of the constants S_i. That is, we prove

(8.9) $$\frac{1}{2}\nu^{-2}S_3 \leq \|\mathbf{a}^*(0)\|_{h^s}^2 \leq 2\nu^{-2}S_3 \qquad \text{and} \qquad \|\mathbf{a}^*(T)\|_{h^s}^2 \geq \frac{\nu^{-2}}{4}S_{\mathfrak{g}-1}.$$

The initial condition of the considered orbit given in (8.6) has support Λ (recall that $\mathcal{Y} = 0$). Therefore,

$$\|\mathbf{a}^*(0)\|_{h^s}^2 = \sum_{i=1}^{\mathfrak{g}} \sum_{\vec{j} \in \Lambda_i} |\vec{j}|^{2s}\nu^{-2}|b_i(0)|^2.$$

Then, taking into account Theorem 7.8,

$$\left|\|\mathbf{a}^*(0)\|_{h^s}^2 - \nu^{-2}S_3\right| \leq 3\nu^{-2}\mu S_3 + \nu^{-2}\mu^2 \sum_{i \neq 3} S_i$$

$$\leq \nu^{-2}S_3\left(3\mu + \mu^2 \sum_{i \neq 3} \frac{S_i}{S_3}\right).$$

From Theorem 7.3 we know that for $i \neq 3$,

$$\frac{S_i}{S_3} \lesssim e^{s\mathfrak{g}}.$$

Therefore, to bound these terms we use the definition of μ from Theorem 7.8. Taking $\gamma > \frac{1}{2}$ and taking $\mathfrak{g}$ large enough, we obtain the first estimate in (8.9).

To obtain the second estimate in (8.9) (final Sobolev norm), note that

(8.10) $$\|\mathbf{a}^*(T)\|_{h^s}^2 \geq \sum_{\vec{j} \in \Lambda_{\mathfrak{g}-1}} |\vec{j}|^{2s}\left|a_{\vec{j}}^*(T)\right|^2 \geq S_{\mathfrak{g}-1}\inf_{\vec{j} \in \Lambda_{\mathfrak{g}-1}}\left|a_{\vec{j}}^*(T)\right|^2.$$

Thus, it is enough to obtain a lower bound for $\left|a_{\vec{j}}^*(T)\right|$ for $\vec{j} \in \Lambda_{\mathfrak{g}-1}$. To obtain this estimate we need to express $\mathbf{a}^*$ in normal form coordinates and use Theorem 8.1. We split $|a_{\vec{j}}^*(T)|$ as follows. Define $(\widetilde{\mathcal{Y}}^*(t), \widetilde{\theta}^*(t), \widetilde{\mathbf{a}}^*(t))$ the image of the orbit with initial condition (8.6) under the changes of variables in Theorems 4.3 and 5.2, Proposition 6.2 and in (7.1). Then,

$$\left|a_{\vec{j}}^*(T)\right| \geq \left|\beta_{\vec{j}}^{\nu}(T)\right| - \left|\widetilde{a}_{\vec{j}}^*(T) - \beta_{\vec{j}}^{\nu}(T)e^{i\Omega_{\vec{j}}(\lambda,\varepsilon)T}\right| - \left|\widetilde{a}_{\vec{j}}^*(T) - a_{\vec{j}}^*(T)\right|.$$

The first term, by Theorem 7.8, satisfies $|\beta_{\vec{j}}^{\nu}(T)| \geq \nu^{-1}/2$. For the second one, using Theorem 8.1, we have

$$\left|\widetilde{a}_{\vec{j}}^*(T) - \beta_{\vec{j}}^{\nu}(T)e^{i\Omega_{\vec{j}}(\lambda,\varepsilon)T}\right| \leq \nu^{-1-\sigma}.$$

Finally, taking into account the estimates (8.7) and (8.8), the third one can be bounded as

$$\left|\widetilde{a}_{\vec{\jmath}}^*(T) - a_{\vec{\jmath}}^*(T)\right| \leq \|\widetilde{\mathbf{a}}^*(T) - \mathbf{a}^*(T)\|_{\ell^1} \lesssim \|\mathbf{a}^*(T)\|_{\ell^1}^2 + \frac{\|\mathbf{a}^*(T)\|_{\ell^1}}{|\vec{\jmath}|}.$$

Now, by Theorem 8.1 and Theorem 7.3 (more precisely the fact that $|\vec{\jmath}| \gtrsim f(\mathfrak{g})$ for $\vec{\jmath} \in \Lambda$),

$$\left|\widetilde{a}_{\vec{\jmath}}^*(T) - a_{\vec{\jmath}}^*(T)\right| \leq \nu^{-1-\sigma}.$$

(taking a smaller σ if necessary). Thus, by (8.10), we obtain the second estimate in (8.9).

The last step to prove Theorem 1.2 is to choose suitable ν and $\mathfrak{g}$ in terms of the parameters $K \gg 1$ and $0 < \delta \ll 1$.

To measure the growth of Sobolev norms, note that (8.9) implies

$$\frac{\|\mathbf{a}^*(T)\|_{h^s}^2}{\|\mathbf{a}^*(0)\|_{h^s}^2} \geq \frac{S_{\mathfrak{g}-1}}{8S_3} \geq \frac{1}{16} 2^{(1-s)(\mathfrak{g}-4)}.$$

Thus, taking $\mathfrak{g} \sim \ln(K/\delta)$, one obtains the growth of Sobolev norm

$$\frac{\|\mathbf{a}^*(T)\|_{h^s}^2}{\|\mathbf{a}^*(0)\|_{h^s}^2} \gtrsim \frac{K^2}{\delta^2}.$$

To control the initial Sobolev norm, we need that $\|\mathbf{a}^*(0)\|_{h^s}^2 \sim \nu^{-2}S_3 \sim \delta^2$. Note that this estimate and the ones just obtained imply $\|\mathbf{a}^*(T)\|_{h^s}^2 \gtrsim K^2$. To estimate $\|\mathbf{a}^*(0)\|_{h^s}^2$, it is enough to choose a suitable ν (as a function of $\mathfrak{g}$). To this end, note that by Theorem 7.3, S_3 satisfies

$$C^{-1}2^{\mathfrak{g}}(f(\mathfrak{g}))^{2s} \leq S_3 \leq C2^{\mathfrak{g}}3^{\mathfrak{g}}(f(\mathfrak{g}))^{2s}.$$

Thus, one can take $\nu^2 \sim \delta^{-2}S_3$, which satisfies

$$f(\mathfrak{g})^s \lesssim \nu \lesssim \delta^{-1}2^{\mathfrak{g}/2}3^{\mathfrak{g}/2}f(\mathfrak{g})^s$$

Then, choosing the parameters s_1, s_2 introduced in Theorem 8.1 such that $0 < s_1 < s < s_2 < 1$, one has that, taking $\mathfrak{g}$ large enough (recall that we have chosen $\mathfrak{g} \sim \ln(K/\delta)$, which we can take arbitrarily big by, if needed, enlarging K), the chosen ν belongs to the range admitted for ν in Theorem 8.1. This gives

$$C^{-1}\delta^2 \leq \|\mathbf{a}^*(0)\|_{h^s}^2 \leq C\delta^2$$

for some $C > 0$ independent of δ. Note that in the reasoning above to obtain small initial Sobolev norm h^s it has been crucial that $s \in (0, 1)$.

Remark 8.2. *In case we ask only the ℓ^2 norm of $\mathbf{a}^*(0)$ to be small we can drop the condition $s < 1$. Indeed $\|\mathbf{a}^*(0)\|_{\ell^2} \lesssim \nu^{-1}2^{\mathfrak{g}}\mathfrak{g}$ which can be made arbitrary small by simply taking $\mathfrak{g}$ large enough (and ν as in (8.5)).*

The time estimates can be easily deduced by (8.2), (8.5), (7.6) and Theorem 7.8, which concludes the proof of the first statement of Theorem 1.2.

For the proof of the second statement of Theorem 1.2 it is enough to point out that the condition $s < 1$ has only been used in imposing that the initial Sobolev norm is small. The estimate for the ℓ^2 norm can be obtained as explained in Remark 8.2.

$\square$

8.1. **Proof of Theorem 8.1.** To prove Theorem 8.1, we define

$$\xi = \beta - \beta^{\nu}(t).$$

We use the equations in (8.4) to deduce an equation for ξ. It can be written as

$$(8.11) \qquad \mathrm{i}\dot{\xi} = \mathcal{Z}_0(t) + \mathcal{Z}_1(t)\xi + \mathcal{Z}_1'(t)\overline{\xi} + \mathcal{Z}_1''(t)\mathcal{Y} + \mathcal{Z}_2(\xi, \mathcal{Y}, t),$$

where

$$
\begin{aligned}
\mathcal{Z}_0(t) =& \partial_{\overline{\beta}}\mathcal{J}(0,\theta,\beta^\nu) + \partial_{\overline{\beta}}\mathcal{R}(0,\theta,\beta^\nu) \\
\mathcal{Z}_1(t) =& \partial_{\beta\overline{\beta}}\mathcal{G}_0(\beta^\nu) + \partial_{\beta\overline{\beta}}\mathcal{J}(0,\theta,\beta^\nu) \\
\mathcal{Z}_1'(t) =& \partial_{\overline{\beta}\overline{\beta}}\mathcal{G}_0(\beta^\nu) + \partial_{\overline{\beta}\overline{\beta}}\mathcal{J}(0,\theta,\beta^\nu) \\
\mathcal{Z}_1''(t) =& \partial_{\mathcal{Y}\overline{\beta}}\mathcal{G}_0(\beta^\nu) + \partial_{\mathcal{Y}\overline{\beta}}\mathcal{J}(0,\theta,\beta^\nu) \\
\mathcal{Z}_2(t) =& \partial_{\overline{\beta}}\mathcal{G}_0(\beta^\nu + \xi) - \partial_{\overline{\beta}}\mathcal{G}_0(\beta^\nu) - \partial_{\beta\overline{\beta}}\mathcal{G}_0(\beta^\nu)\xi - \partial_{\overline{\beta}\overline{\beta}}\mathcal{G}_0(\beta^\nu)\overline{\xi} \\
& \partial_{\overline{\beta}}\mathcal{J}(\mathcal{Y},\theta,\beta^\nu+\xi) - \partial_{\overline{\beta}}\mathcal{J}(0,\theta,\beta^\nu) - \partial_{\beta\overline{\beta}}\mathcal{J}(0,\theta,\beta^\nu)\xi - \partial_{\overline{\beta}\overline{\beta}}\mathcal{J}(0,\theta,\beta^\nu)\overline{\xi} \\
& - \partial_{\mathcal{Y}\overline{\beta}}\mathcal{J}(0,\theta,\beta^\nu)\mathcal{Y} + \partial_{\overline{\beta}}\mathcal{R}(\mathcal{Y},\theta,\beta^\nu+\xi) - \partial_{\overline{\beta}}\mathcal{R}(0,\theta,\beta^\nu).
\end{aligned}
$$

(8.12)

We analyze the equations for ξ in (8.11) and $\mathcal{Y}$ in (8.4).

Lemma 8.3. *Assume that $(\beta^\nu, \mathcal{Y})$, $(\beta^\nu + \xi, \mathcal{Y}) \in D(r_2)$ (see (2.1)) where r_2 has been given by Theorem 6.2. Then, the function $\|\xi\|_{\ell^1}$ satisfies*

$$
\begin{aligned}
\frac{d}{dt}\|\xi\|_{\ell^1} \leq& C\nu^{-4}\mathfrak{g}^4 2^{4\mathfrak{g}} + C\nu^{-3}\mathfrak{g}^3 2^{3\mathfrak{g}}\left(f(\mathfrak{g})^{-\frac{4}{5}} + tf(\mathfrak{g})^{-2}\right) \\
& + C\nu^{-2}\mathfrak{g}^2 2^{2\mathfrak{g}}\|\xi\|_{\ell^1} + C\nu^{-1}2^{\mathfrak{g}}|\mathcal{Y}| + C\nu^{-1}2^{\mathfrak{g}}\|\xi\|_{\ell^1}^2 + C\|\xi\|_{\ell^1}|\mathcal{Y}| + C|\mathcal{Y}|^2
\end{aligned}
$$

for some constant $C > 0$ independent of ν.

Proof. We compute estimates for each term in (8.12). For $\mathcal{Z}_0$, we use the fact that the definition of $\mathcal{R}$ in (7.3) and Theorem 6.2 imply $\|\partial_{\overline{\beta}}\mathcal{R}(0,\theta,\beta^\nu)\|_{\ell^1} \leq \mathcal{O}(\|\beta^\nu\|_{\ell^1}^4)$. Thus, it only remains to use the results in Theorems 7.8 (using (8.1)) and Theorem 7.3, to obtain

$$\|\partial_{\overline{\beta}}\mathcal{R}(0,\theta,\beta^\nu)\|_{\ell^1} \leq C\nu^{-4}\mathfrak{g}^4 2^{4\mathfrak{g}}.$$

To bound $\partial_{\overline{\beta}}\mathcal{J}(0,\theta,\beta^\nu)$, the other term in $\mathcal{Z}_0$, recall that $\mathcal{J} = \mathcal{J}_1 + \mathcal{J}_2$ (see (7.3) and (7.10)). Then, we split into two terms $\partial_{\overline{\beta}}\mathcal{J}(0,\theta,\beta^\nu) = \partial_{\overline{\beta}}\mathcal{J}_1(0,\theta,\beta^\nu) + \partial_{\overline{\beta}}\mathcal{J}_2(\theta,\beta^\nu)$ as

$$
\begin{aligned}
\partial_{\overline{\beta}}\mathcal{J}_1(0,\theta,\beta^\nu) &= \partial_{\overline{\beta}}\left\{\mathcal{G}\left(0,\theta,(\beta_{\vec{j}}^\nu e^{\mathrm{i}\Omega_{\vec{j}}(\lambda,\varepsilon)t})_{\vec{j}\in\mathbb{Z}_N^2\setminus S_0}\right) - \mathcal{G}(0,\theta,\beta^\nu)\right\} \\
&= \partial_{\overline{\beta}}\left\{\mathcal{Q}_{\mathrm{Res}}^{(2)}\left(0,\theta,(\beta_{\vec{j}}^\nu e^{\mathrm{i}\Omega_{\vec{j}}(\lambda,\varepsilon)t})_{\vec{j}\in\mathbb{Z}_N^2\setminus S_0}\right) - \mathcal{Q}_{\mathrm{Res}}^{(2)}(0,\theta,\beta^\nu)\right\}
\end{aligned}
$$

(8.13)

$$\partial_{\overline{\beta}}\mathcal{J}_2(\theta,\beta^\nu) = \partial_{\overline{\beta}}\left\{\mathcal{G}\left(0,\theta,(\beta_{\vec{j}}^\nu e^{\mathrm{i}\Omega_{\vec{j}}(\lambda,\varepsilon)t})_{\vec{j}\in\mathbb{Z}_N^2\setminus S_0}\right) - \mathcal{G}_0\left((\beta_{\vec{j}}^\nu e^{\mathrm{i}\Omega_{\vec{j}}(\lambda,\varepsilon)t})_{\vec{j}\in\mathbb{Z}_N^2\setminus S_0}\right)\right\}$$

(8.14)

To bound (8.13), recall that $\mathcal{Q}_{\mathrm{Res}}^{(2)}$ defined in (6.4) is the sum of two terms. Since $\Pi_{\mathfrak{K}_2}\mathcal{Q}^{(2)}$ is action preserving, the only terms contributing to (8.13) are the ones coming from $\Pi_{\mathfrak{K}_4}\mathcal{Q}^{(2)}$. Since β^ν is supported on Λ, it follows from (6.2) that

$$
\partial_{\overline{\beta}}\mathcal{J}_1(0,\theta,\beta^\nu) = \left(\sum_{\substack{\vec{j}_1,\vec{j}_2,\vec{j}_3\in\Lambda \\ |\vec{j}_1|^2 - |\vec{j}_2|^2 - |\vec{j}_3|^2 - |\vec{j}|^2 = 0}} \left(e^{\mathrm{i}t(\Omega_{\vec{j}_1} - \Omega_{\vec{j}_2} + \Omega_{\vec{j}_3} - \Omega_{\vec{j}})} - 1\right) \mathcal{J}_{\vec{j}_1\vec{j}_2\vec{j}_3\vec{j}} \beta_{\vec{j}_1}^\nu \overline{\beta_{\vec{j}_2}^\nu} \beta_{\vec{j}_3}^\nu\right)_{\vec{j}\in\Lambda}.
$$

(8.15)

In order to bound the oscillating factor, we use the formula for the eigenvalues given in Theorem 4.4, to obtain that, for $\vec{j}_1, \vec{j}_2, \vec{j}_3, \vec{j} \in \Lambda$, one has

$$\left|e^{\mathrm{i}t(\Omega_{\vec{j}_1} - \Omega_{\vec{j}_2} + \Omega_{\vec{j}_3} - \Omega_{\vec{j}})} - 1\right| \lesssim |t|\,|\Omega_{\vec{j}_1} - \Omega_{\vec{j}_2} + \Omega_{\vec{j}_3} - \Omega_{\vec{j}}| \lesssim \frac{|t|}{f(\mathfrak{g})^2}.$$

Hence, for $t \in [0, T]$, using the estimate for $\mathcal{Q}_{\mathrm{Res}}^{(2)}$ given by Theorem 6.2,

$$\|\partial_{\overline{\beta}}\mathcal{J}_1(0,\theta,\beta^\nu)\|_{\ell^1} \leq Ctf(\mathfrak{g})^{-2}\|\beta^\nu\|_{\ell^1}^3 \leq Ct\nu^{-3}\mathfrak{g}^3 2^{3\mathfrak{g}}f(\mathfrak{g})^{-2}.$$

To bound (8.14), it is enough to use (7.12) and (7.6) to obtain

$$\|\partial_{\overline{\beta}}\mathcal{J}_2(\theta,\beta^\nu)\|_{\ell^1} \leq Cf(\mathfrak{g})^{-\frac{4}{5}}\|\beta^\nu\|_{\ell^1}^3 \leq C\nu^{-3}\mathfrak{g}^3 2^{3\mathfrak{g}}f(\mathfrak{g})^{-\frac{4}{5}}.$$

For the linear terms, one can easily see that

$$\|\mathcal{Z}_1(t)\xi\|_{\ell^1} \leq C\|\beta^\nu\|_{\ell^1}^2\|\xi\|_{\ell^1} \leq C\nu^{-2}\mathfrak{g}^2 2^{2\mathfrak{g}}\|\xi\|_{\ell^1}$$

and the same for $\left\|\mathcal{Z}_1'(t)\overline{\xi}\right\|_{\ell^1}$. Analogously,

$$\|\mathcal{Z}_1''(t)\mathcal{Y}\|_{\ell^1} \leq C\|\beta^\nu\|_{\ell^1}|\mathcal{Y}| \leq C\nu^{-1}\mathfrak{g}2^{\mathfrak{g}}|\mathcal{Y}|.$$

Finally, it is enough to use the definition of $\mathcal{Z}_2$, the definition of $\mathcal{R}$ in (7.3) and Theorem 6.2 to show

$$\|\mathcal{Z}_2\| \leq \|\beta^\nu\|_{\ell^1}\|\xi\|_{\ell^1}^2 + \|\beta^\nu\|_{\ell^1}^2|\mathcal{Y}| + \|\xi\|_{\ell^1}|\mathcal{Y}| + \|\beta^\nu\|_{\ell^1}^3\|\xi\|_{\ell^1} + |\mathcal{Y}|^2$$
$$\leq C\nu^{-1}\mathfrak{g}2^{\mathfrak{g}}|\mathcal{Y}|\|\xi\|_{\ell^1}^2 + C\nu^{-2}\mathfrak{g}^2 2^{2\mathfrak{g}}\|\xi\|_{\ell^1}|\mathcal{Y}| + C\nu^{-3}\mathfrak{g}^3 2^{3\mathfrak{g}}\|\xi\|_{\ell^1} + |\mathcal{Y}|^2.$$

$\square$

Lemma 8.4. *Assume that* $(\beta^\nu, \mathcal{Y}), (\beta^\nu + \xi, \mathcal{Y}) \in D(r_2)$ *(see* (2.1)*) where r_2 has been given by Theorem 6.2. Then, the function* $|\mathcal{Y}|$ *satisfies*

$$\frac{d}{dt}|\mathcal{Y}| \leq C\nu^{-5}\mathfrak{g}^5 2^{5\mathfrak{g}} + C\nu^{-3}\mathfrak{g}^3 2^{3\mathfrak{g}}\|\xi\|_{\ell^1}^2$$
$$+ C\nu^{-1}\mathfrak{g}2^{\mathfrak{g}}\|\xi\|_{\ell^1}^3 + C\|\xi\|_{\ell^1}|\mathcal{Y}| + C\nu^{-3}\mathfrak{g}^3 2^{3\mathfrak{g}}|\mathcal{Y}|^2$$

for some constant $C > 0$ independent of ν.

Proof. Proceeding as for $\dot{\xi}$, we write the equation for $\dot{\mathcal{Y}}$ as

(8.16) $$\dot{\mathcal{Y}} = \mathcal{X}_0(t) + \mathcal{X}_1(t)\xi + \mathcal{X}_1'(t)\overline{\xi} + \mathcal{X}_1''(t)\mathcal{Y} + \mathcal{X}_2(\xi, \mathcal{Y}, t),$$

with

$$\mathcal{X}_0(t) = -\partial_\theta\mathcal{J}(0, \theta, \beta^\nu) - \partial_\theta\mathcal{R}(0, \theta, \beta^\nu).$$
$$\mathcal{X}_1(t) = \partial_{\beta\theta}\mathcal{J}(0, \theta, \beta^\nu)$$
$$\mathcal{X}_1'(t) = \partial_{\overline{\beta}\theta}\mathcal{J}(0, \theta, \beta^\nu)$$
$$\mathcal{X}_1''(t) = \partial_{\mathcal{Y}\theta}\mathcal{J}(0, \theta, \beta^\nu)$$
$$\mathcal{X}_2(t) = -\partial_\theta\mathcal{J}(\mathcal{Y}, \theta, \beta^\nu + \xi) + \partial_\theta\mathcal{J}(0, \theta, \beta^\nu)$$
$$- \partial_\theta\mathcal{R}(\mathcal{Y}, \theta, \beta^\nu + \xi) + \partial_\theta\mathcal{R}(0, \theta, \beta^\nu).$$

We claim that $\mathcal{X}_1(t)$ and $\mathcal{X}_1'(t)$ are identically zero. Then, proceeding as in the proof of Lemma 8.3, one can bound each term and complete the proof of Lemma 8.4.

To explain the absence of linear terms, consider first $\partial_{\beta\theta}\mathcal{J}(0, \theta, \beta^\nu)$. It contains two types of monomials: those coming from $\mathfrak{K}_2$ (see (4.4)) which however do not depend on θ, and those coming from $\mathfrak{K}_4$ (see (6.2)). But also these last monomials do not depend on θ once they are restricted on the set Λ (indeed the only monomials of $\mathfrak{K}_4$ which are θ dependent are those of the third line of (6.2), which are supported outside Λ). Therefore $\partial_{\beta\theta}\mathcal{J}(0, \theta, \beta^\nu) \equiv 0$ (and so $\partial_{\overline{\beta}\theta}\mathcal{J}(0, \theta, \beta^\nu)$ and $\partial_{\mathcal{Y}\theta}\mathcal{J}(0, \theta, \beta^\nu)$).

$\square$

We define

$$M = \|\xi\|_{\ell^1} + \nu|\mathcal{Y}|.$$

As a conclusion of these two lemmas, we can deduce that

$$\dot{M} \leq C\left(\nu^{-4}\mathfrak{g}^4 2^{4\mathfrak{g}} + \nu^{-3}\mathfrak{g}^3 2^{3\mathfrak{g}}\left(f(\mathfrak{g})^{-\frac{4}{5}} + tf(\mathfrak{g})^{-2}\right)\right) + C\nu^{-2}\mathfrak{g}^2 2^{2\mathfrak{g}}M + \nu^{-1}\mathfrak{g}2^{\mathfrak{g}}M^2.$$

Now we apply a bootstrap argument. Assume that for some $T^* > 0$ and $0 < t < T^*$ we have

$$M(t) \leq C\nu^{-1-\sigma/2}$$

(the constant σ will be determined later).

Recall that, by assumption, for $t = 0$ we know that it is already satisfied since

(8.17) $$M(0) \leq \nu^{-1-4\sigma}.$$

A posteriori we will show that the time T in (8.2) satisfies $0 < T < T^*$ and therefore the bootstrap assumption holds. Note that, taking $\mathfrak{g}$ large enough (and recalling (7.6) and (8.5)), the bootstrap estimate implies that $(\beta^\nu, \mathcal{Y})$ and $(\beta^\nu + \xi, \mathcal{Y})$ satisfy the assumption of Lemmas 8.3 and 8.4. With the boostrap assumption then, we have

$$\dot{M} \leq C\left(\nu^{-4}\mathfrak{g}^4 2^{4\mathfrak{g}} + \nu^{-3}\mathfrak{g}^3 2^{3\mathfrak{g}}\left(f(\mathfrak{g})^{-\frac{4}{5}} + tf(\mathfrak{g})^{-2}\right)\right) + C\nu^{-2}\mathfrak{g}^2 2^{2\mathfrak{g}}M.$$

Applying Gronwall inequality on obtains,

$$M \leq C\left(M(0) + \nu^{-4}\mathfrak{g}^4 2^{4\mathfrak{g}}t + \nu^{-3}\mathfrak{g}^3 2^{3\mathfrak{g}}\left(tf(\mathfrak{g})^{-\frac{4}{5}} + t^2 f(\mathfrak{g})^{-2}\right)\right)e^{\nu^{-2}\mathfrak{g}^2 2^{2\mathfrak{g}}t}$$

Thus, using (8.2), the estimates for T_0 in Theorem 7.8, (8.17) and (8.5) (note that it implies $f(\mathfrak{g})^{-1} \leq \nu^{-1}$),

$$M \leq C\left(M(0) + \nu^{-2}\mathfrak{g}^6 2^{4\mathfrak{g}} + \nu^{-1}\mathfrak{g}^5 2^{3\mathfrak{g}}f(\mathfrak{g})^{-\frac{4}{5}} + \nu\mathfrak{g}^7 2^{3\mathfrak{g}}f(\mathfrak{g})^{-2}\right)e^{C\mathfrak{g}^4 2^{2\mathfrak{g}}}$$

$$\leq C\left(\nu^{-1-4\sigma} + \nu^{-2}\mathfrak{g}^6 2^{4\mathfrak{g}} + \nu^{-9/5}\mathfrak{g}^5 2^{3\mathfrak{g}} + \nu^{-3}\mathfrak{g}^7 2^{3\mathfrak{g}}\right)e^{C\mathfrak{g}^4 2^{2\mathfrak{g}}}$$

Now, taking A large enough (see Theorem 7.3), there exists $\sigma > 0$ such that for $t \in [0, T]$, provided $\mathfrak{g}$ is sufficiently large,

$$M(t) \leq \nu^{-1-\sigma}.$$

This implies that $T \leq T^*$. That is, the bootstrap assumption was valid. This completes the proof.

APPENDIX A. PROOF OF PROPOSITION 6.1

We split the proof in several steps. We first perform an algebraic analysis of the nonresonant monomials.

A.1. **Analysis of monomials of the form** $e^{i\theta\cdot\ell}a_{\vec{j}_1}^{\sigma_1}a_{\vec{j}_2}^{\sigma_2}a_{\vec{j}_3}^{\sigma_3}a_{\vec{j}_4}^{\sigma_4}$. We analyze the small divisors (6.1) related to these monomials. Taking advantage of the asymptotics of the eigenvalues given in Theorem 4.4, we consider a "good" first order approximation of the small divisor given by

$$(A.1) \qquad \omega(\lambda)\cdot\ell + \sigma_1\widetilde{\Omega}_{\vec{j}_1}(\lambda,\varepsilon) + \sigma_2\widetilde{\Omega}_{\vec{j}_2}(\lambda,\varepsilon) + \sigma_3\widetilde{\Omega}_{\vec{j}_3}(\lambda,\varepsilon) + \sigma_4\widetilde{\Omega}_{\vec{j}_4}(\lambda,\varepsilon).$$

Note that this is an affine function in ε and therefore it can be written as

$$(A.1) \equiv \mathtt{K}_{\mathbf{j},\ell}^\sigma + \varepsilon\mathtt{F}_{\mathbf{j},\ell}^\sigma(\lambda).$$

We say that a monomial is Birkhoff non-resonant if, for any $\varepsilon > 0$, this expression is not 0 as a function of λ.

Lemma A.1. *Assume that the $\mathtt{m}_k$'s do not solve any of the linear equations defined in (A.5) (this determines $\mathtt{L}_2$ in the statement of Theorem 6.1). Consider a monomial of the form $e^{i\theta\cdot\ell}a_{\vec{j}_1}^{\sigma_1}a_{\vec{j}_2}^{\sigma_2}a_{\vec{j}_3}^{\sigma_3}a_{\vec{j}_4}^{\sigma_4}$ with $(\mathbf{j},\ell,\sigma) \in \mathfrak{A}_4$. If $(\mathbf{j},\ell,\sigma) \notin \mathfrak{R}_4$, then it is Birkhoff non resonant.*

Proof. We write explicitly the functions $\mathtt{K}_{\mathbf{j},\ell}^\sigma$ and $\mathtt{F}_{\mathbf{j},\ell}^\sigma(\lambda)$ as

$$(A.2) \qquad \mathtt{K}_{\mathbf{j},\ell}^\sigma := \omega^{(0)}\cdot\ell + \sigma_1\widetilde{\Omega}_{\vec{j}_1}(\lambda,0) + \sigma_2\widetilde{\Omega}_{\vec{j}_2}(\lambda,0) + \sigma_3\widetilde{\Omega}_{\vec{j}_3}(\lambda,0) + \sigma_4\widetilde{\Omega}_{\vec{j}_3}(\lambda,0)$$

$$\mathtt{F}_{\mathbf{j},\ell}^\sigma(\lambda) := \partial_\varepsilon\left(\omega(\lambda)\cdot\ell + \sigma_1\widetilde{\Omega}_{\vec{j}_1}(\lambda,\varepsilon) + \sigma_2\widetilde{\Omega}_{\vec{j}_2}(\lambda,\varepsilon) + \sigma_3\widetilde{\Omega}_{\vec{j}_3}(\lambda,\varepsilon) + \sigma_4\widetilde{\Omega}_{\vec{j}_4}(\lambda,\varepsilon)\right)\Big|_{\varepsilon=0}$$

$$(A.3) \qquad = -\lambda\cdot\ell + \sigma_1\vartheta_{\vec{j}_1}(\lambda) + \sigma_2\vartheta_{\vec{j}_2}(\lambda) + \sigma_3\vartheta_{\vec{j}_3}(\lambda) + \sigma_4\vartheta_{\vec{j}_4}(\lambda)$$

As in [MP18], $\mathtt{K}_{\mathbf{j},\ell}^\sigma$ is an integer while the functions $\vartheta_{\vec{j}}(\lambda)$ belong to the finite list of functions $\vartheta_{\vec{j}}(\lambda) \in \left\{0, \{\mu_i(\lambda)\}_{1\leq i\leq\mathtt{d}}\right\}$ defined in Theorem 4.4. Clearly to prove that the resonance (A.1) not to hold identically, it is enough to ensure that

$$(A.4) \qquad \mathtt{K}_{\mathbf{j},\ell}^\sigma = 0 \quad \text{and} \quad \mathtt{F}_{\mathbf{j},\ell}^\sigma(\lambda) \equiv 0$$

cannot occur for $(\mathbf{j},\ell,\sigma) \in \mathfrak{A}_4 \setminus \mathfrak{R}_4$. We study all the possible combinations, each time we assume that (A.4) holds and we deduce a contradiction.

(1) $\vec{\jmath}_i \in \mathbf{Z}$ for any $1 \le i \le 4$. In case $\ell \ne 0$, then $\mathtt{F}^{\sigma}_{\mathbf{j},\ell}(\lambda) = -\lambda \cdot \ell$ is not identically 0. Now take $\ell = 0$. By conservation of $\widetilde{\mathcal{P}}_x, \widetilde{\mathcal{P}}_y$ we have that $\sum_{i=1}^{4} \sigma_i \vec{\jmath}_i = 0$ and $\mathtt{K}^{\sigma}_{\mathbf{j},\ell} = 0$ implies $\sum_{i=1}^{4} \sigma_i |\vec{\jmath}_i|^2 = 0$. Then, using mass conservation (see Remark 4.7), since $\ell = 0$, one has $\sum_{i=1}^{4} \sigma_i = 0$ and therefore the $\vec{\jmath}_i$'s form a rectangle (and thus $(\mathbf{j}, 0, \sigma)$ belongs to $\mathfrak{K}_4$).

(2) $\vec{\jmath}_1, \vec{\jmath}_2, \vec{\jmath}_3 \in \mathbf{Z}$, $\vec{\jmath}_4 \in \mathscr{S}$. Then $\mathtt{F}^{\sigma}_{\mathbf{j},\ell}(\lambda) = -\lambda \cdot \ell + \sigma_3 \mu_i(\lambda)$ for some $1 \le i \le \mathtt{d}$. If $\mathtt{F}^{\sigma}_{\mathbf{j},\ell}(\lambda) \equiv 0$ then $\mu_i(\lambda) = \sigma_3 \lambda \cdot \ell$ is a root in $\mathbb{Z}[\lambda]$ of the polynomial $P(t, \lambda)$ defined in Theorem 4.4; however $P(t, \lambda)$ is irreducible over $\mathbb{Q}(\lambda)[t]$, thus leading to a contradiction.

(3) $\vec{\jmath}_1, \vec{\jmath}_2 \in \mathbf{Z}$, $\vec{\jmath}_3, \vec{\jmath}_4 \in \mathscr{S}$. W.l.o.g. let $\vec{\jmath}_3 = (\mathtt{m}_i, n_3)$, $\vec{\jmath}_4 = (\mathtt{m}_k, n_4)$ for some $1 \le i, k \le \mathtt{d}$. Then

$$\mathtt{F}^{\sigma}_{\mathbf{j},\ell}(\lambda) = -\lambda \cdot \ell + \sigma_3 \mu_i(\lambda) + \sigma_4 \mu_k(\lambda) .$$

Case $\ell \ne 0$. Then $\mathtt{F}^{\sigma}_{\mathbf{j},\ell}(\lambda) \equiv 0$ iff $\mu_i(\lambda) \equiv -\sigma_3 \sigma_4 \mu_k(\lambda) + \sigma_3 \lambda \cdot \ell$. This means that $\mu_k(\lambda)$ is a common root of $P(t, \lambda)$ and $P(-\sigma_3 \sigma_4 t + \sigma_3 \lambda \cdot \ell, \lambda)$. However this last polynomial is irreducible as well, being the translation of an irreducible polynomial. Hence the two polynomials must be equal (or opposite). A direct computation shows that this does not happen (see Lemma 6.1 of [MP18] for details).

Case $\ell = 0$. Then $\mathtt{F}^{\sigma}_{\mathbf{j},\ell}(\lambda) \equiv 0$ iff $\mu_i(\lambda) \equiv -\sigma_3 \sigma_4 \mu_k(\lambda)$.
- If $i \ne k$ and $\sigma_3 \sigma_4 = -1$, then $P(t, \lambda)$ would have a root with multiplicity 2. But $P(t, \lambda)$, being an irreducible polynomial, has no multiple roots.
- If $i \ne k$ and $\sigma_3 \sigma_4 = 1$, then $P(t, \lambda)$ and $P(-t, \lambda)$ would have $\mu_k(\lambda)$ as a common root. However $P(-t, \lambda)$ is irreducible on $\mathbb{Z}[\lambda]$ as well, and two irreducible polynomials sharing a common root must coincide (up to a sign), i.e. $P(t, \lambda) \equiv \pm P(-t, \lambda)$. A direct computation using the explicit expression of $P(t, \lambda)$ shows that this is not true.
- If $i = k$ and $\sigma_3 \sigma_4 = 1$ then $\mu_i(\lambda) \equiv 0$ would be a root of $P(t, \lambda)$. But $P(t, \lambda)$ is irreducible over $\mathbb{Z}[\lambda]$, does it cannot have 0 as a root.
- If $i = k$ and $\sigma_3 \sigma_4 = -1$ (w.l.o.g. assume $\sigma_3 = 1$, $\sigma_4 = -1$), by mass conservation one has $\sigma_1 + \sigma_2 = 0$ and by conservation of $\widetilde{\mathcal{P}}_x$ one has $\sigma_1 m_1 + \sigma_2 m_2 = 0$, thus $m_1 = m_2$. Then by conservation of $\widetilde{\mathcal{P}}_y$ we get $n_1 - n_2 + n_3 - n_4 = 0$, which together with $0 = \mathtt{K}^{\sigma}_{\mathbf{j},\ell} = n_1^2 - n_2^2 + n_3^2 - n_4^2$ gives $\{n_1, n_3\} = \{n_2, n_4\}$. One verifies easily that in such a case the sites $\vec{\jmath}_r$'s form a horizontal rectangle (that could be even degenerate), and therefore they belong to $\mathfrak{K}_4$.

(4) $\vec{\jmath}_1, \vec{\jmath}_2, \vec{\jmath}_3 \in \mathscr{S}$, $\vec{\jmath}_4 \in \mathbf{Z}$. W.l.o.g. let $\vec{\jmath}_1 = (\mathtt{m}_{i_1}, n_1)$, $\vec{\jmath}_2 = (\mathtt{m}_{i_2}, n_2)$, $\vec{\jmath}_3 = (\mathtt{m}_{i_3}, n_3)$ for some $1 \le i_1, i_2, i_3 \le \mathtt{d}$ and $n_1, n_2, n_3 \ne 0$. Then

$$\mathtt{F}^{\sigma}_{\mathbf{j},\ell}(\lambda) = -\lambda \cdot \ell + \sigma_1 \mu_{i_1}(\lambda) + \sigma_2 \mu_{i_2}(\lambda) + \sigma_3 \mu_{i_3}(\lambda) .$$

By conservation of mass $\eta(\ell) + \sigma_4 = 0$, hence $\ell \ne 0$. Assume $\mathtt{F}^{\sigma}_{\mathbf{j},\ell}(\lambda) \equiv 0$. This can only happen for (at most) a unique choice of $\ell^{(\mathbf{i},\sigma)} \in \mathbb{Z}^{\mathtt{d}}$ uniquely, $\mathbf{i} := (i_1, i_2, i_3)$. By conservation of $\widetilde{\mathcal{P}}_x$ we have $\sum_k \mathtt{m}_k \ell_k^{(\mathbf{i},\sigma)} + \sigma_4 m_4 = 0$. These two conditions fix $m_4 \equiv m_4^{(\mathbf{i},\sigma)}$ uniquely. In particular if m_4 is sufficiently large, we have a contradiction.

(5) $\vec{\jmath}_r \in \mathscr{S}$, $\forall 1 \le r \le 4$. Then

$$\mathtt{F}^{\sigma}_{\mathbf{j},\ell}(\lambda) = -\lambda \cdot \ell + \sigma_1 \mu_{i_1}(\lambda) + \sigma_2 \mu_{i_2}(\lambda) + \sigma_3 \mu_{i_3}(\lambda) + \sigma_4 \mu_{i_4}(\lambda) .$$

If $\ell \ne 0$, the condition $\mathtt{F}^{\sigma}_{\mathbf{j},\ell}(\lambda) \equiv 0$ fixes $\ell^{(\mathbf{i},\sigma)} \in \mathbb{Z}^{\mathtt{d}}$ uniquely, $\mathbf{i} := (i_1, i_2, i_3, i_4)$. By conservation of $\widetilde{\mathcal{P}}_x$ we have the condition

(A.5)
$$\sum_k \mathtt{m}_k \ell_k^{(\mathbf{i},\sigma)} = 0$$

defining a hyperplane, which can be excluded by suitably choosing the tangential sites $\mathtt{m}_k$ (recall that the functions $\mu_i(\lambda)$ are independent of this choice, see Remark 4.5).

If $\ell = 0$, we have $\sum_r \sigma_r n_r = \sum_r \sigma_r n_r^2 = 0$. Then $\{n_1, n_3\} = \{n_2, n_4\}$. One verifies easily that in such case the sites $\vec{\jmath}_r$'s form a horizontal trapezoid (that could be even degenerate).

$\square$

A.2. **Analysis of monomials of the form** $e^{i\theta\cdot\ell}\mathcal{Y}^l a_{j_1}^{\sigma_1} a_{j_2}^{\sigma_2}$. In this case, since the factor $\mathcal{Y}^l$ does not affect the Poisson brackets, admissible monomials (in the sense of Definition 4.1) are non-resonant provided they do not belong to the set $\mathfrak{K}_2$ introduced in Definition 4.2.

Lemma A.2. *Any monomial of the form* $e^{i\theta\cdot\ell} a_{j_1}^{\sigma_1} a_{j_2}^{\sigma_2}\mathcal{Y}_i$ *with* $(\mathbf{j}, \ell, \sigma) \notin \mathfrak{K}_2$ *admissible in the sense of Definition 4.1 is Birkhoff non-resonant.*

Proof. We skip the proof since it is analogous to Lemma 6.1 of [MP18]. $\qquad\square$

A.3. **Quantitative measure estimate.** We are now in a position to prove our quantitative non-resonance estimate. Recall that, by Theorem 4.3, the frequencies $\Omega_{\bar{j}}(\lambda, \varepsilon)$ of Hamiltonian (5.1) have the form (4.8). Expanding $\Omega_{\bar{j}}(\lambda, \varepsilon)$ in Taylor series in powers of ε we get that

$$(A.6) \qquad \omega(\lambda)\cdot\ell + \sigma_1\Omega_{\bar{j}_1}(\lambda, \varepsilon) + \sigma_2\Omega_{\bar{j}_2}(\lambda, \varepsilon) + \sigma_3\Omega_{\bar{j}_3}(\lambda, \varepsilon) + \sigma_4\Omega_{\bar{j}_4}(\lambda, \varepsilon) = \mathtt{K}_{\mathbf{j},\ell}^{\sigma} + \varepsilon\, \mathtt{F}_{\mathbf{j},\ell}^{\sigma}(\lambda) + \varepsilon^2\, \mathtt{G}_{\mathbf{j},\ell}^{\sigma}(\lambda, \varepsilon)\ ,$$

where $\mathtt{K}_{\mathbf{j},\ell}^{\sigma}$ is defined in (A.2) and $\mathtt{F}_{\mathbf{j},\ell}^{\sigma}(\lambda)$ is defined in (A.3). We wish to prove that the set of $\lambda \in \mathcal{C}_{\varepsilon}^{(2)}$ such that

$$(A.7) \qquad |\omega(\lambda)\cdot\ell + \sigma_1\Omega_{\bar{j}_1}(\lambda, \varepsilon) + \sigma_2\Omega_{\bar{j}_2}(\lambda, \varepsilon) + \sigma_3\Omega_{\bar{j}_3}(\lambda, \varepsilon) + \sigma_4\Omega_{\bar{j}_4}(\lambda, \varepsilon)| \geq \varepsilon\frac{\gamma_2}{\langle\ell\rangle^{\tau_2}}\ , \qquad \forall\, (\mathbf{j}, \ell, \sigma) \in \mathfrak{A}_4 \setminus \mathfrak{K}_4$$

has positive measure for γ_2 and ε small enough and τ_2 large enough. We treat separately the cases $|\ell| \leq 4\mathtt{M}_0$ and $|\ell| > 4\mathtt{M}_0$.

A.3.1. *Case $|\ell| \leq 4\mathtt{M}_0$*. We start with the following lemma.

Lemma A.3. *There exist* $\mathtt{k} \in \mathbb{N}$, *such that for any* $\gamma_c > 0$ *sufficiently small, there exists a compact domain* $\mathcal{C}_c \subset \mathcal{O}_0$, *with* $|\mathcal{O}_0 \setminus \mathcal{C}_c| \sim \gamma_c^{1/\mathtt{k}}$ *and*

$$\min\left\{ \left| \mathtt{F}_{\mathbf{j},\ell}^{\sigma}(\lambda) \right| : \lambda \in \mathcal{C}_c, (\ell, \mathbf{j}, \sigma) \in \mathfrak{A}_4 \setminus \mathfrak{K}_4, |\ell| \leq 4\mathtt{M}_0\ , \mathtt{K}_{\mathbf{j},\ell}^{\sigma} = 0 \right\} \geq \gamma_c > 0\ .$$

Proof. See Lemma 6.4 of [MP18]. The estimate on the measure follows from classical results on sublevels of analytic functions. $\qquad\square$

We can now prove the following result.

Proposition A.4. *There exits* $\varepsilon_c > 0$ *and a set* $\mathcal{C}_c \subset \mathcal{O}_0$ *such that for any* $\varepsilon \leq \varepsilon_c$, *any* $\lambda \in \mathcal{C}_c$, *one has*

$$(A.8) \qquad \left| \omega(\lambda)\cdot\ell + \sum_{l=1}^{4} \sigma_l\Omega_{\bar{j}l}(\lambda, \varepsilon) \right| \geq \frac{\gamma_c\varepsilon}{2}\ , \qquad \forall\, (\mathbf{j}, \ell, \sigma) \in \mathfrak{A}_4 \setminus \mathfrak{K}_4\ , \quad |\ell| \leq 4\mathtt{M}_0.$$

Moreover, one has that $|\mathcal{O}_0 \setminus \mathcal{C}_c| \leq \alpha\varepsilon_c^{\kappa}$ *where* α, κ *do not depend on* ε_c.

Proof. By the very definition of $\mathtt{M}_0$ in (4.9) and the estimates on the eigenvalues given in Theorem 4.4, one has $\sup_{\lambda \in \mathcal{O}_0} |\mathtt{F}_{\mathbf{j},\ell}^{\sigma}(\lambda)| \leq 8\,\mathtt{M}_0$ and $\sup_{\lambda \in \mathcal{O}_0} |\mathtt{G}_{\mathbf{j},\ell}^{\sigma}(\lambda)| \leq 4\,\mathtt{M}_0$. Assume first that $\mathtt{K}_{\mathbf{j},\ell}^{\sigma} \in \mathbb{Z} \setminus \{0\}$, then if ε_c is sufficiently small and for $\varepsilon < \varepsilon_c$ one has

$$|(A.6)| \geq |\mathtt{K}_{\mathbf{j},\ell}^{\sigma}| - \varepsilon 8\mathtt{M}_0 - \varepsilon^2 4\mathtt{M}_0 \geq \frac{1}{2}.$$

Hence, for such ℓ's, (A.8) is trivially fulfilled $\forall\lambda \in \mathcal{O}_0$. If instead $\mathtt{K}_{\mathbf{j},\ell}^{\sigma} = 0$, we use Lemma A.3 with $\gamma_c = 10\mathtt{M}_0\varepsilon_c$ to obtain a set $\mathcal{C}_c \subset \mathcal{O}_0$, such that for any $\lambda \in \mathcal{C}_c$ and any $(\mathbf{j}, \ell, \sigma) \in \mathfrak{A}_4 \setminus \mathfrak{K}_4$ with $|\ell| \leq 4\mathtt{M}_0$

$$|(A.6)| \geq \varepsilon\gamma_c - \varepsilon^2 4\mathtt{M}_0 \geq \frac{\varepsilon\gamma_c}{2} =: C\varepsilon\ .$$

$\qquad\square$

A.3.2. *Case $|\ell| > 4\mathtt{M}_0$.* In this case we prove the following result.

Proposition A.5. *Fix $\varepsilon_\star > 0$ sufficiently small and $\tau_\star > 0$ sufficiently large. For any $\varepsilon < \varepsilon_\star$, there exists a set $\mathcal{C}_\star \subset \mathcal{O}_0$ such that $|\mathcal{O}_0 \setminus \mathcal{C}_\star| \lesssim \varepsilon_\star^\kappa$ (with α, κ independent of $\varepsilon_\star$), and for any $\lambda \in \mathcal{C}_\star$ and $|\ell| > 4\mathtt{M}_0$ one has*

$$
(A.9) \qquad \left| \omega(\lambda) \cdot \ell + \sum_{l=1}^4 \sigma_l \Omega_{\vec{j}l}(\lambda, \varepsilon) \right| \geq \gamma_\star \frac{\varepsilon}{\langle \ell \rangle^{\tau_\star}} \ .
$$

for some constant $\gamma_\star$ depending on $\varepsilon_\star$.

To prove the proposition, first define, for $1 \leq i \leq \mathtt{d}$ and $0 \leq k \leq \mathtt{d}$, the functions

$$
\widehat{\mathtt{F}}_{i,k}(\lambda) = \begin{cases} \varepsilon \mu_i(\lambda) & \text{if } k = 0 \\ \varepsilon \mu_{i,k}^+(\lambda) & \text{if } 1 \leq i < k \leq \mathtt{d} \\ \varepsilon \mu_{i,k}^-(\lambda) & \text{if } 1 \leq k < i \leq \mathtt{d} \\ 0 & \text{if } 1 \leq i = k \leq \mathtt{d} \end{cases}
$$

The right hand side of (A.6) is always of the form

$$
\begin{aligned}
(A.10) \qquad & \omega(\lambda) \cdot \ell + K + \eta_1 \widehat{\mathtt{F}}_{i_1,k_1}(\lambda) + \eta_2 \widehat{\mathtt{F}}_{i_2,k_2}(\lambda) + \eta_3 \widehat{\mathtt{F}}_{i_3,k_3}(\lambda) + \eta_4 \widehat{\mathtt{F}}_{i_4,k_4}(\lambda) \\
& + \eta_{11} \frac{\Theta_{m_1}(\lambda,\varepsilon)}{\langle m_1 \rangle^2} + \eta_{12} \frac{\Theta_{m_2}(\lambda,\varepsilon)}{\langle m_2 \rangle^2} + \eta_{13} \frac{\Theta_{m_3}(\lambda,\varepsilon)}{\langle m_3 \rangle^2} + \eta_{14} \frac{\Theta_{m_4}(\lambda,\varepsilon)}{\langle m_4 \rangle^2} \\
& + \eta_{21} \frac{\Theta_{m_1,n_1}(\lambda,\varepsilon)}{\langle m_1 \rangle^2 + \langle n_1 \rangle^2} + \eta_{22} \frac{\Theta_{m_2,n_2}(\lambda,\varepsilon)}{\langle m_2 \rangle^2 + \langle n_2 \rangle^2} + \eta_{23} \frac{\Theta_{m_3,n_3}(\lambda,\varepsilon)}{\langle m_3 \rangle^2 + \langle n_3 \rangle^2} + \eta_{24} \frac{\Theta_{m_4,n_4}(\lambda,\varepsilon)}{\langle m_4 \rangle^2 + \langle n_4 \rangle^2} \\
& + \eta_{31} \frac{\varpi_{m_1}(\lambda,\varepsilon)}{\langle m_1 \rangle} + \eta_{32} \frac{\varpi_{m_2}(\lambda,\varepsilon)}{\langle m_2 \rangle} + \eta_{33} \frac{\varpi_{m_3}(\lambda,\varepsilon)}{\langle m_3 \rangle} + \eta_{34} \frac{\varpi_{m_4}(\lambda,\varepsilon)}{\langle m_4 \rangle}
\end{aligned}
$$

for a particular choice of $K \in \mathbb{Z}$, $m_i \in \mathbb{Z}$, $n_i \in N\mathbb{Z} \setminus \{0\}$ and $\eta_r, \eta_{jj'} \in \{-1, 0, 1\}$. Therefore it is enough to show (A.9) where the left hand side is replaced by (A.10).

Proof of Proposition A.5. If the integer K is sufficiently large, namely $|K| \geq 4|\ell| \max_{1 \leq i \leq \mathtt{d}}(\mathtt{m}_i^2)$, then the quantity in the left hand side of (A.9) is far from zero. More precisely one has

$$
\begin{aligned}
|(A.10)| &\geq |K| - |\omega(\lambda)| \, |\ell| - \sum_{r=1}^4 \left| \widehat{\mathtt{F}}_{i_r,k_r} \right|^{\mathcal{O}_1} - \sum_{r=1}^4 \frac{|\Theta_{m_r}(\cdot,\varepsilon)|^{\mathcal{O}_1}}{\langle m_r \rangle^2} - \sum_{r=1}^4 \frac{|\Theta_{m_r,n_r}(\cdot,\varepsilon)|^{\mathcal{O}_1}}{\langle m_r \rangle^2 + \langle n_r \rangle^2} - \sum_{r=1}^4 \frac{|\varpi_{m_r}(\cdot,\varepsilon)|^{\mathcal{O}_1}}{\langle m_r \rangle} \\
&\geq 4 \max_{1 \leq i \leq \mathtt{d}}(\mathtt{m}_i^2) |\ell| - \max_{1 \leq i \leq \mathtt{d}}(\mathtt{m}_i^2) |\ell| - \varepsilon |\ell| - 4\varepsilon \mathtt{M}_0 - 4\varepsilon^2 \mathtt{M}_0 \geq \mathtt{M}_0 \ .
\end{aligned}
$$

So from now on we restrict ourselves to the case $|K| \leq 4|\ell| \max_{1 \leq i \leq \mathtt{d}}(\mathtt{m}_i^2)$. We will repeatedly use the following result, which is an easy variant of Lemma 5 of [Pös96].

Lemma A.6. *Fix arbitrary $K \in \mathbb{Z}$, $m_i \in \mathbb{Z}$, $n_i \in \mathbb{Z} \setminus \{0\}$, $\eta_j, \eta_{jj'} \in \{-1, 0, 1\}$. For any $\alpha > 0$ one has*

$$
\mathrm{meas}(\{\lambda \in \mathcal{O}_0 : |(A.10)| < \varepsilon \alpha\}) < 16\alpha |\ell|^{-1} \ .
$$

The proof relies on the fact that all the functions appearing in (A.10) are Lipschitz in λ, for full details see e.g. Lemma C.2 of [MP18].

Now, let us fix

$$
(A.11) \qquad \gamma_\star = \frac{\varepsilon_\star \mathtt{M}_0}{100}.
$$

We construct the set $\mathcal{C}_\star$ by induction on the number n defined by

$$
n := |\eta_{1,1}| + \cdots + |\eta_{3,4}| \leq 12
$$

which is nothing but the number of nonzero coefficients in (A.10) . For every $0 \leq n \leq 12$ we construct (i) a positive increasing sequence τ_n and (ii) a sequence of nested sets $\mathcal{C}^n = \mathcal{C}^n(\gamma_\star, \tau_n)$ such that

(1) There exists $C > 0$, independent of ε and $\gamma_\star$, s.t.

$$(A.12) \qquad \operatorname{meas}(\mathcal{O}_0 \setminus \mathcal{C}^0) \leq C\gamma_\star , \quad \operatorname{meas}(\mathcal{C}^n \setminus \mathcal{C}^{n+1}) \leq C\gamma_\star$$

(2) For $\lambda \in \mathcal{C}^n$ and $|\ell| \geq 4\mathtt{M}_0$ one has

$$(A.13) \qquad \left| (A.10) \right| \geq \frac{\varepsilon\, \gamma_\star}{\langle \ell \rangle^{\tau_n}} .$$

Then the proposition follows by taking $\mathcal{C}_\star := \mathcal{C}^{12}$, $\tau_\star = \tau_{12}$, so that one has $|\mathcal{O}_0 \setminus \mathcal{C}_\star| \leq 13C\gamma_\star \sim \gamma_\star$, provided $\gamma_\star$ is small enough.

Case $n = 0$: Define the set

$$G^0_{K,\mathbf{i},\mathbf{k},\eta,\ell}(\gamma_\star, \tau_0) := \left\{ \lambda \in \mathcal{O}_0 \ : \ |(A.10)| \leq \frac{\varepsilon\, \gamma_\star}{\langle \ell \rangle^{\tau_0}} \ \text{ and } \ \eta_{jj'} = 0 \quad \forall j, j' \right\} ,$$

where $K \in \mathbb{Z}$ with $|K| \leq 4 \max\limits_{1 \leq i \leq \mathtt{d}}(\mathtt{m}_i^2)\, |\ell|$, $\mathbf{i} = (i_1, i_2, i_3, i_4) \in \{1, \dots \mathtt{d}\}^4$, $\mathbf{k} = (k_1, k_2, k_3, k_4) \in \{0, \dots, \mathtt{d}\}^4$, $\ell \in \mathbb{Z}^\mathtt{d}$ with $|\ell| \geq 4\mathtt{M}_0$, $\eta = (\eta_1, \eta_2, \eta_3, \eta_4) \in \{-1, 0, 1\}^4$. By Lemma A.6 with $\alpha = \gamma_\star \langle \ell \rangle^{-\tau_0}$ we have

$$\operatorname{meas}\left(G^0_{K,\mathbf{i},\mathbf{k},\eta,\ell}(\gamma_\star, \tau_0)\right) \leq \frac{16\gamma_\star}{\langle \ell \rangle^{\tau_0+1}} .$$

Taking the union over all the possible values of $K, \mathbf{i}, \mathbf{k}, \eta, \ell$ one gets that

$$\operatorname{meas}\left(\bigcup_{\substack{|\ell| \geq 4\mathtt{M}_0, \ \mathbf{i}, \mathbf{k}, \eta \\ |K| \leq 4 \max_i(\mathtt{m}_i^2)\, |\ell|}} G^0_{K,\mathbf{i},\mathbf{k},\eta,\ell}(\gamma_\star, \tau_0) \right) \leq C(\mathtt{d})\, \gamma_\star \sum_{|\ell| \geq 4\mathtt{M}_0} \frac{1}{\langle \ell \rangle^{\tau_0}} \leq C\gamma_\star ,$$

which is finite provided $\tau_0 \geq \mathtt{d} + 1$. Letting

$$\mathcal{C}^0 := \mathcal{O}_0 \setminus \bigcup_{\substack{|\ell| \geq 4\mathtt{M}_0, \ \mathbf{i}, \mathbf{k}, \eta \\ |K| \leq 4 \max_i(\mathtt{m}_i^2)\, |\ell|}} G^0_{K,\mathbf{i},\mathbf{k},\eta,\ell}(\gamma_\star, \tau_0)$$

one has clearly that $\operatorname{meas}(\mathcal{O}_0 \setminus \mathcal{C}^0) \leq C\gamma_\star$ and for $\lambda \in \mathcal{C}^0$ we have

$$\left| \omega(\lambda) \cdot \ell + K + \sum_{r=1}^4 \eta_j \widehat{\mathbb{F}}_{i_r, k_r}(\lambda) \right| \geq \frac{\varepsilon\, \gamma_\star}{\langle \ell \rangle^{\tau_0}}$$

for any admissible choice of $\ell, K, \mathbf{i}, \mathbf{k}, \eta$. This proves the inductive step for $n = 0$.

Case $n \rightsquigarrow n+1$: Assume that (A.13) holds for any possible choice of $\eta_{11}, \dots, \eta_{34}$ s.t. $|\eta_{11}| + \cdots + |\eta_{34}| \leq n \leq 11$ for some $(\tau_j)_{j=1}^n$. We prove now the step $n+1$. Let us fix $\tau_{n+1} \geq \mathtt{d} + 1 + 6\tau_n$. We shall show that for each $|\ell| \geq 4\mathtt{M}_0$, the set

$$(A.14) \qquad G^{n+1}_\ell := \left\{ \lambda \in \mathcal{C}^n : |(A.10)| \leq \frac{\varepsilon\gamma_\star}{\langle \ell \rangle^{\tau_{n+1}}} , \quad |\eta_{11}| + \dots + |\eta_{34}| = n+1 \right\}$$

has measure $\leq \dfrac{C(\mathtt{d})\gamma_\star}{\langle \ell \rangle^{\mathtt{d}+1}}$. Thus defining

$$\mathcal{C}^{n+1} := \mathcal{C}^n \setminus \bigcup_{|\ell| \geq 4\mathtt{M}_0} G^{n+1}_\ell(\gamma_\star, \tau_{n+1}).$$

we obtain the claimed estimates (A.12) and (A.13). To estimate the measure of (A.14) we split in three cases.

Case 1: Assume that

$$\exists\, m_i \ \text{s.t.} \ |m_i| \geq \langle \ell \rangle^{\tau_n}$$

(of course we also assume that one of the coefficients $\eta_{1i}, \eta_{2i}, \eta_{3i}$ is not null). W.l.o.g. assume it is m_4. Then we treat all the terms in (A.10) which contain m_4 as perturbations, and we estimate all the other terms using the inductive assumption. Here the details: first we have

$$\left|\frac{\Theta_{m_4}(\lambda,\varepsilon)}{\langle m_4\rangle^2}\right| + \left|\frac{\Theta_{m_4,n_4}(\lambda,\varepsilon)}{\langle m_4\rangle^2 + \langle n_4\rangle^2}\right| + \left|\frac{\varpi_{m_4}(\lambda,\varepsilon)}{\langle m_4\rangle}\right| \leq \frac{\mathtt{M}_0\,\varepsilon^2}{\langle\ell\rangle^{\tau_n}}.$$

By the inductive assumption (A.13) and (A.11), for any $\lambda \in \mathcal{C}^n$ one has

$$\left|(A.10)\right| \geq \left|\omega(\lambda)\cdot\ell + K + \sum_{j=1}^4 \eta_i \widehat{\mathtt{F}}_{i_j,k_j}(\lambda) + \sum_{j=1}^3 \eta_{1j}\frac{\Theta_{m_r}(\lambda,\varepsilon)}{\langle m_j\rangle^2} + \sum_{j=1}^3 \eta_{2j}\frac{\Theta_{m_j,n_j}(\lambda,\varepsilon)}{\langle m_j\rangle^2 + \langle n_j\rangle^2} + \sum_{j=1}^3 \eta_{3j}\frac{\varpi_{m_j}(\lambda,\varepsilon)}{\langle m_j\rangle}\right| - \frac{\mathtt{M}_0\,\varepsilon^2}{\langle\ell\rangle^{\tau_n}}$$

$$\geq \frac{\varepsilon\,\gamma_\star}{\langle\ell\rangle^{\tau_n}} - \frac{\mathtt{M}_0\,\varepsilon^2}{\langle\ell\rangle^{\tau_n}} \geq \frac{\varepsilon\,\gamma_\star}{2\langle\ell\rangle^{\tau_n}} \geq \frac{\varepsilon\,\gamma_\star}{\langle\ell\rangle^{\tau_{n+1}}}$$

provided $\tau_{n+1} \geq \tau_n + 1$. Therefore, in this case, there are no λ's contributing to the set (A.14).

Case 2: Assume that

$$\exists\, n_i \text{ s.t. } |n_i|^2 \geq \langle\ell\rangle^{\tau_n}$$

(and again we also assume that one of the coefficients η_{2i} is not null). W.l.o.g. assume it is n_4. Similarly to the previous case, we treat the term in (A.10) which contains n_4 as a perturbation, and we estimate all the other terms using the inductive assumption. More precisely we have

$$\left|\frac{\Theta_{m_4,n_4}(\lambda,\varepsilon)}{\langle m_4\rangle^2 + \langle n_4\rangle^2}\right| \leq \frac{\mathtt{M}_0\,\varepsilon^2}{\langle\ell\rangle^{\tau_n}},$$

so by the inductive assumption (A.13) and (A.11)

$$\left|(A.10)\right| \geq \left|\omega(\lambda)\cdot\ell + K + \sum_{j=1}^4 \eta_i \widehat{\mathtt{F}}_{i_j,k_j}(\lambda) + \sum_{j=1}^4 \eta_{1j}\frac{\Theta_{m_j}(\lambda,\varepsilon)}{\langle m_j\rangle^2} + \sum_{j=1}^3 \eta_{2j}\frac{\Theta_{m_j,n_j}(\lambda,\varepsilon)}{\langle m_j\rangle^2 + \langle n_j j\rangle^2} + \sum_{j=1}^4 \eta_{3j}\frac{\varpi_{m_j}(\lambda,\varepsilon)}{\langle m_j\rangle}\right| - \frac{\mathtt{M}_0\,\varepsilon^2}{\langle\ell\rangle^{\tau_n}}$$

$$\geq \frac{\varepsilon\,\gamma_\star}{2\langle\ell\rangle^{\tau_n}} \geq \frac{\varepsilon\,\gamma_\star}{\langle\ell\rangle^{\tau_{n+1}}}$$

provided $\tau_{n+1} \geq \tau_n + 1$. Also in this case, there are no λ's contributing to the set (A.14).

Case 3: We have

$$|m_i|, |n_i|^2 \leq \langle\ell\rangle^{\tau_n}$$

for all the m_i, n_i that appear in (A.10) with nonzero coefficients. Furthermore, recall that we are considering just the case $|K| \leq 4\max_i(\mathtt{m}_i^2)|\ell|$. Thus we are left with a finite number of cases and we can impose a finite number of Melnikov conditions. So define the sets

$$G^{n+1}_{K,\mathbf{i},\mathbf{k},\eta,\ell,\mathbf{m},\mathbf{n}}(\gamma_\star,\tau_{n+1}) := \left\{\lambda \in \mathcal{C}^n \;:\; |(A.10)| \leq \frac{\varepsilon\,\gamma_\star}{\langle\ell\rangle^{\tau_{n+1}}}\;,\qquad |\eta_{11}| + \cdots + |\eta_{34}| = n+1\right\}.$$

By Lemma A.6 with $\alpha = \gamma/\langle\ell\rangle^{\tau_{n+1}}$ we have

$$(A.15) \qquad \operatorname{meas}\left(G^{n+1}_{K,\mathbf{i},\mathbf{k},\eta,\ell,\mathbf{m},\mathbf{n}}(\gamma_\star,\tau_{n+1})\right) \leq \frac{16\gamma_\star}{\langle\ell\rangle^{\tau_{n+1}+1}}\;,$$

and taking the union over the possible values of $K, \mathbf{i}, \mathbf{k}, \eta, \mathbf{m}, \mathbf{n}$ one gets that

$$G^{n+1}_\ell \equiv \bigcup_{\mathbf{i},\mathbf{k},\eta}\; \bigcup_{\substack{|m_i|,\,|n_i|^2 \leq \langle\ell\rangle^{\tau_n} \\ |K| \leq 4\max_i(\mathtt{m}_i^2)|\ell|}} G_{K,\mathbf{i},\mathbf{k},\eta,\ell,\mathbf{m},\mathbf{n}}(\gamma_\star,\tau_{n+1}).$$

Estimate (A.15) gives immediately

$$\operatorname{meas}\left(G^{n+1}_\ell\right) \leq C(\mathtt{d})\,\gamma_\star\,\frac{\langle\ell\rangle^{1+6\tau_n}}{\langle\ell\rangle^{\tau_{n+1}+1}} \leq \frac{C(\mathtt{d})\,\gamma_\star}{\langle\ell\rangle^{\mathtt{d}+1}}$$

which is what we claimed.

$\square$

We can finally prove Proposition 6.1.

Proof of Proposition 6.1. Fix $\gamma_c = \gamma_\star =: \gamma_2$ sufficiently small, and put $\varepsilon_2 := \min(\varepsilon_c, \varepsilon_\star)$, $\tau_2 := \tau_\star$ and $\mathcal{C}^{(2)} := \mathcal{C}_c \cap \mathcal{C}_\star$. Propositions A.4 and A.5 guarantee that for any $\lambda \in \mathcal{C}^{(2)}$, estimate (A.7) is fulfilled. Finally one has $|\mathcal{C}^{(1)} \setminus \mathcal{C}^{(2)}| \lesssim \gamma_2^{1/\mathbf{k}} + \gamma_2 \sim \gamma_2^{1/\mathbf{k}}$.

$\square$

APPENDIX B. LIST OF NOTATIONS

We give a list of notations and parameters. We also specify the relations between the parameters needed to prove the first statement of Theorem 1.2.

- $\mathtt{d} \in \mathbb{N}$ – Dimension of the torus $\mathcal{T}_{\mathcal{S}_0}^I$
- $\mathcal{S}_0 \subset \mathbb{Z} \times \{0\}$ – Set of (Birkhoff) modes where the torus $\mathcal{T}_{\mathcal{S}_0}^I$ is supported. It has cardinality $\mathtt{d}$.
- $I = (I_{\mathtt{m}_1}, \ldots, I_{\mathtt{m}_\mathtt{d}}) \in \mathbb{R}_{>0}^\mathtt{d}$ – Actions which define the torus $\mathcal{T}_{\mathcal{S}_0}^I$.
- $\varepsilon \in \mathbb{R}$ – Size of the actions I.
- $s \in (0,1)$ – The index for the Sobolev norm H^s in the first statement of Theorem 1.2.
- $\delta \ll 1$ – It measures the initial distance from the torus $\mathcal{T}_{\mathcal{S}_0}^I$ (in the Sobolev norm H^s) in the first statement of Theorem 1.2.
- $K \gg 1$ – It measures the final Sobolev norm in the first statement of Theorem 1.2.
- $r \in \mathbb{R}$ – Size of the neighborhood of 0 where the several steps of Birkhoff normal form are performed (see (2.1)).
- $\rho \in \mathbb{R}$ – Width of the analyticity domain in the angles θ (see (2.1)).
- $\lambda \in \left(\frac{1}{2}, 1\right)^\mathtt{d}$ – Parameter used to modulate the actions I (see Lemma 3.3).
- $\omega(\lambda) \in \mathbb{R}^\mathtt{d}$ – Tangential frequencies of the torus (see Lemma 3.3).
- $N \in \mathbb{N}$ – It is introduced in (4.1) and defines the set lattice $\mathbb{Z} \times N\mathbb{Z}$ where equation (2D-NLS) is restricted. It will be chosen depending on $\mathfrak{g}$ (see below).
- $\Omega_{\bar{j}}(\lambda, \varepsilon)$ – Normal frequencies of the torus $\mathcal{T}_{\mathcal{S}_0}^I$ (see Theorem 4.3).
- $\Lambda \subset \mathbb{Z} \times N\mathbb{Z}$ – Set where the solution undergoing growth of Sobolev norms is essentially supported (see Theorem 7.3).
- $\mathfrak{g} \in \mathbb{N}$ – Number of generations of the set Λ. In Section 8 it is chosen $\mathfrak{g} \sim \ln(K/\delta)$.
- $f(\mathfrak{g})$ – It gives the size of the modes in Λ (see (7.6) and (7.7)).
- $\mu \in \mathbb{R}$ – It measures the errors in the toy model orbit (see Theorem 7.8). It satisfies $\mu = e^{-\gamma \mathfrak{g}}$ for some $\gamma \gg 1$.
- $T_0 \in \mathbb{R}$ – Transition time for the toy model orbit (see Theorem 7.8). It satisfies $T_0 \sim \mathfrak{g}^2$.
- $\nu \in \mathbb{R}$ – Scaling applied to the toy model solution. In the case $s \in (0,1)$ (first part of the statement of Theorem 1.2) it satisfies $f(\mathfrak{g})^{s_1} \leq \nu \leq f(\mathfrak{g})^{s_2}$ for some $0 < s_1 < s < s_2 < 1$.
- $T \in \mathbb{R}$ – Transition time for the orbit in the first statement of Theorem 1.2 (see also (8.2)). It satisfies $T \sim \nu^2 \mathfrak{g}^2 \lesssim e^{\left(\frac{K}{\delta}\right)^\beta}$ for some $\beta > 1$.

REFERENCES

[BB11]   M. Berti and L. Biasco. Branching of Cantor manifolds of elliptic tori and applications to PDEs. *Comm. Math. Phys.*, 305(3):741–796, 2011.

[BB13]   M. Berti and Ph. Bolle. Quasi-periodic solutions with Sobolev regularity of NLS on $\mathbb{T}^d$ with a multiplicative potential. *J. Eur. Math. Soc. (JEMS)*, 15(1):229–286, 2013.

[BBHM18] P. Baldi, M. Berti, E. Haus, and R. Montalto. Time quasi-periodic gravity water waves in finite depth. *Invent. Math.*, 214(2):739–911, 2018.

[BCP15]   Massimiliano Berti, Livia Corsi, and Michela Procesi. An abstract Nash-Moser theorem and quasi-periodic solutions for NLW and NLS on compact Lie groups and homogeneous manifolds. *Comm. Math. Phys.*, 334(3):1413–1454, 2015.

[BM16]   D. Bambusi and A. Maspero. Birkhoff coordinates for the Toda lattice in the limit of infinitely many particles with an application to FPU. *J. Funct. Anal.*, 270(5):1818–1887, 2016.

[Bou93]   J. Bourgain. Fourier transform restriction phenomena for certain lattice subsets and application to nonlinear evolution equations. i. schrödinger equations. *Geom. Funct. Anal.*, 3(2):107–156, 1993.

[Bou96] J. Bourgain. On the growth in time of higher Sobolev norms of smooth solutions of Hamiltonian PDE. *Internat. Math. Res. Notices*, 6:277–304, 1996.

[Bou98] J. Bourgain. Quasi-periodic solutions of Hamiltonian perturbations of 2D linear Schrödinger equations. *Ann. of Math. (2)*, 148(2):363–439, 1998.

[Bou00] J. Bourgain. Problems in Hamiltonian PDE's. *Geom. Funct. Anal.*, Special Volume, Part I:32–56, 2000. GAFA 2000 (Tel Aviv, 1999).

[Bou04] J. Bourgain. Remarks on stability and diffusion in high-dimensional Hamiltonian systems and partial differential equations. *Ergodic Theory Dynam. Systems*, 24(5):1331–1357, 2004.

[CDKS01] J. E. Colliander, J.-M. Delort, C. E. Kenig, and G. Staffilani. Bilinear estimates and applications to 2D NLS. *Trans. Amer. Math. Soc.*, 353(8):3307–3325 (electronic), 2001.

[CKO12] J. Colliander, S. Kwon, and T. Oh. A remark on normal forms and the "upside-down" I-method for periodic NLS: growth of higher Sobolev norms. *J. Anal. Math.*, 118(1):55–82, 2012.

[CKS$^+$10] J. Colliander, M. Keel, G. Staffilani, H. Takaoka, and T. Tao. Transfer of energy to high frequencies in the cubic defocusing nonlinear Schrödinger equation. *Invent. Math.*, 181(1):39–113, 2010.

[CW10] F. Catoire and W.-M. Wang. Bounds on Sobolev norms for the defocusing nonlinear Schrödinger equation on general flat tori. *Commun. Pure Appl. Anal.*, 9(2):483–491, 2010.

[CZ13] C. Q. Cheng and J. Zhang. Asymptotic trajectories of KAM torus. Preprint available at http://arxiv.org/abs/1312.2102, 2013.

[Del10] J.-M. Delort. Growth of Sobolev norms of solutions of linear Schrödinger equations on some compact manifolds. *Int. Math. Res. Not. IMRN*, (12):2305–2328, 2010.

[Dou88] R. Douady. Stabilité ou instabilité des points fixes elliptiques. *Ann. Sci. École Norm. Sup. (4)*, 21(1):1–46, 1988.

[DSPSN07] D. De Silva, N. Pavlovic, G. Staffilani, and Tzirakisc N. Global well-posedness for a periodic, nonlinear schrödinger equation in 1d and 2d. *Discrete & Continuous Dynamical Systems – A*, 19(1):37–65, 2007.

[EK10] L. H. Eliasson and S. B. Kuksin. KAM for the nonlinear Schrödinger equation. *Ann. of Math. (2)*, 172(1):371–435, 2010.

[FGL14] E. Faou, L. Gauckler, and C. Lubich. Plane wave stability of the split-step Fourier method for the nonlinear Schrödinger equation. *Forum Math. Sigma*, 2:e5, 45, 2014.

[GG10] P. Gérard and S. Grellier. The cubic Szegő equation. *Ann. Sci. Éc. Norm. Supér. (4)*, 43(5):761–810, 2010.

[GG12] P. Gérard and S. Grellier. Effective integrable dynamics for a certain nonlinear wave equation. *Anal. PDE*, 5(5):1139–1155, 2012.

[GG15] P. Gérard and S. Grellier. An explicit formula for the cubic Szegő equation. *Trans. Amer. Math. Soc.*, 367(4):2979–2995, 2015.

[GHP16] M. Guardia, E. Haus, and M. Procesi. Growth of Sobolev norms for the analytic NLS on $\mathbb{T}^2$. *Adv. Math.*, 301:615–692, 2016.

[GK14a] B. Grébert and T. Kappeler. *The defocusing NLS equation and its normal form*. EMS Series of Lectures in Mathematics. European Mathematical Society (EMS), Zürich, 2014.

[GK14b] M. Guardia and M. Kaloshin. Orbits of nearly integrable systems accumulating to KAM tori. Preprint available at https://arxiv.org/abs/1412.7088, 2014.

[GK15] M. Guardia and V. Kaloshin. Growth of Sobolev norms in the cubic defocusing nonlinear Schrödinger equation. *J. Eur. Math. Soc. (JEMS)*, 17(1):71–149, 2015.

[GK17] M. Guardia and V. Kaloshin. Erratum to "Growth of Sobolev norms in the cubic defocusing nonlinear Schrödinger equation" [MR3312404]. *J. Eur. Math. Soc. (JEMS)*, 19(2):601–602, 2017.

[Gua14] M. Guardia. Growth of Sobolev norms in the cubic nonlinear Schrödinger equation with a convolution potential. *Comm. Math. Phys.*, 329(1):405–434, 2014.

[GXY11] J. Geng, X. Xu, and J. You. An infinite dimensional KAM theorem and its application to the two dimensional cubic Schrödinger equation. *Adv. Math.*, 226(6):5361–5402, 2011.

[Han14] Z. Hani. Long-time instability and unbounded Sobolev orbits for some periodic nonlinear Schrödinger equations. *Arch. Ration. Mech. Anal.*, 211(3):929–964, 2014.

[HP15] E. Haus and M. Procesi. Growth of Sobolev norms for the quintic NLS on T^2. *Anal. PDE*, 8(4):883–922, 2015.

[HPTV15] Z. Hani, B. Pausader, N. Tzvetkov, and N. Visciglia. Modified scattering for the cubic Schrödinger equation on product spaces and applications. *Forum Math. Pi*, 3:e4, 63, 2015.

[KMMT16] T. Kappeler, A. Maspero, J. Molnar, and P. Topalov. On the convexity of the KdV Hamiltonian. *Comm. Math. Phys.*, 346(1):191–236, 2016.

[KMV04] V. Kaloshin, J. N. Mather, and E. Valdinoci. Instability of resonant totally elliptic points of symplectic maps in dimension 4. *Astérisque*, (297):79–116, 2004. Analyse complexe, systèmes dynamiques, sommabilité des séries divergentes et théories galoisiennes. II.

[KP96] S.B. Kuksin and J. Pöschel. Invariant Cantor manifolds of quasi-periodic oscillations for a nonlinear Schrödinger equation. *Ann. of Math. (2)*, 143(1):149–179, 1996.

[KST17] T. Kappeler, B. Schaad, and P. Topalov. Scattering-like phenomena of the periodic defocusing NLS equation. *Math. Res. Lett.*, 24(3):803–826, 2017.

[Kuk96] S.B. Kuksin. Growth and oscillations of solutions of nonlinear Schrödinger equation. *Comm. Math. Phys.*, 178(2):265–280, 1996.

[Kuk97a] S. B. Kuksin. On turbulence in nonlinear Schrödinger equations. *Geom. Funct. Anal.*, 7(4):783–822, 1997.

[Kuk97b] S.B. Kuksin. Oscillations in space-periodic nonlinear Schrödinger equations. *Geom. Funct. Anal.*, 7(2):338–363, 1997.

[LCD83] P. Le Calvez and R. Douady. Exemple de point fixe elliptique non topologiquement stable en dimension 4. *C.R.Acad.Sci. Paris*, 296:895–898, 1983.

[Mas18a] A. Maspero. Lower bounds on the growth of Sobolev norms in some linear time dependent Schrödinger equations. *Math. Res. Lett.*, In press, 2018.

[Mas18b] A. Maspero. Tame majorant analyticity for the Birkhoff map of the defocusing Nonlinear Schrödinger equation on the circle. *Nonlinearity*, 31(5):1981–2030, 2018.

[MP18] A. Maspero and M. Procesi. Long time stability of small finite gap solutions of the cubic Nonlinear Schrödinger equation on $\mathbb{T}^2$. *J. Diff. Eq.*, 265(7):3212–3309, 2018.

[Poc11] O. Pocovnicu. Explicit formula for the solution of the Szegö equation on the real line and applications. *Discrete Contin. Dyn. Syst.*, 31(3):607–649, 2011.

[Poc13] O. Pocovnicu. First and second order approximations for a nonlinear wave equation. *J. Dynam. Differential Equations*, 25(2):305–333, 2013.

[Pös96] J. Pöschel. A KAM-theorem for some nonlinear partial differential equations. *Ann. Scuola Norm. Sup. Pisa Cl. Sci. (4)*, 23(1):119–148, 1996.

[PP12] M. Procesi and C. Procesi. A normal form for the Schrödinger equation with analytic non-linearities. *Comm. Math. Phys.*, 312(2):501–557, 2012.

[PP15] C. Procesi and M. Procesi. A KAM algorithm for the resonant non-linear Schrödinger equation. *Adv. Math.*, 272:399–470, 2015.

[PTV17] F. Planchon, N. Tzvetkov, and N. Visciglia. On the growth of Sobolev norms for NLS on 2- and 3-dimensional manifolds. *Anal. PDE*, 10(5):1123–1147, 2017.

[PX13] M. Procesi and X. Xu. Quasi-Töplitz Functions in KAM Theorem. *SIAM J. of Math. Anal.*, 45(4):2148 – 2181, 2013.

[RT11] F. Rousset and N. Tzvetkov. Transverse instability of the line solitary water-waves. *Invent. Math.*, 184(2):257–388, 2011.

[RT12] F. Rousset and N. Tzvetkov. Stability and instability of the KdV solitary wave under the KP-I flow. *Comm. Math. Phys.*, 313(1):155–173, 2012.

[Soh11a] V. Sohinger. Bounds on the growth of high Sobolev norms of solutions to nonlinear Schrödinger equations on $\mathbb{R}$. *Indiana Univ. Math. J.*, 60(5):1487–1516, 2011.

[Soh11b] V. Sohinger. Bounds on the growth of high Sobolev norms of solutions to nonlinear Schrödinger equations on S^1. *Differential Integral Equations*, 24(7-8):653–718, 2011.

[Soh12] V. Sohinger. Bounds on the growth of high Sobolev norms of solutions to 2D Hartree equations. *Discrete Contin. Dyn. Syst.*, 32(10):3733–3771, 2012.

[Sta97] G. Staffilani. Quadratic forms for a 2-D semilinear Schrödinger equation. *Duke Math. J.*, 86(1):79–107, 1997.

[Wan16] W.-M. Wang. Energy supercritical nonlinear Schrödinger equations: quasiperiodic solutions. *Duke Math. J.*, 165(6):1129–1192, 2016.

[Way90] C.E. Wayne. Periodic and quasi-periodic solutions of nonlinear wave equations via kam theory. *Comm. Math. Phys.*, 127(3):479–528, 1990.

[Zho08] S. Zhong. The growth in time of higher Sobolev norms of solutions to Schrödinger equations on compact Riemannian manifolds. *J. Differential Equations*, 245(2):359–376, 2008.