Commun. Math. Phys. 377, 1681–1759 (2020)
Digital Object Identifier (DOI) https://doi.org/10.1007/s00220-020-03788-z

Communications in
Mathematical Physics

Reducible KAM Tori for the Degasperis–Procesi Equation

Roberto Feola[1], **Filippo Giuliani**[2], **Michela Procesi**[3]

[1] UnivNantes, Nantes, France. E-mail: roberto.feola@univ-nantes.fr
[2] UPC, Barcelona, Spain. E-mail: filippo.giuliani@upc.edu
[3] RomaTRE, Rome, Italy. E-mail: procesi@mat.uniroma3.it

Received: 17 January 2019 / Accepted: 2 April 2020
Published online: 19 June 2020 – © The Author(s) 2020

Abstract: We develop KAM theory close to an elliptic fixed point for quasi-linear Hamiltonian perturbations of the dispersive Degasperis–Procesi equation on the circle. The overall strategy in KAM theory for quasi-linear PDEs is based on Nash–Moser nonlinear iteration, pseudo differential calculus and normal form techniques. In the present case the complicated *symplectic structure*, the *weak dispersive* effects of the linear flow and the presence of *strong resonant interactions* require a novel set of ideas. The main points are to exploit the integrability of the unperturbed equation, to look for special *wave packet* solutions and to perform a very careful algebraic analysis of the resonances. Our approach is quite general and can be applied also to other 1d integrable PDEs. We are confident for instance that the same strategy should work for the Camassa–Holm equation.

Contents

This research was supported by Miur-PRIN 2015 "Variational methods, with applications to problems in mathematical physics and geometry" n. 2015KB9WPT-008, by the ERC Grant "Hamiltonian PDEs and small divisor problems: a dynamical systems approach" n. 306414 under FP7 and by the ERC Project "FAnFArE" n. 637510. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme under Grant Agreement No. 757802.

1. Introduction and Main Result

In this paper we prove existence and stability of Cantor families of quasi-periodic, small amplitude, solutions for quasi-linear Hamiltonian perturbations of the Degasperis–Procesi (DP) equation

$$u_t - u_{xxt} + u_{xxx} - 4u_x - uu_{xxx} - 3u_x u_{xx} + 4uu_x + \mathcal{N}_8(u, u_x, u_{xx}, u_{xxx}) = 0 \quad (1.1)$$

under periodic boundary conditions $x \in \mathbb{T} := \mathbb{R}/2\pi\mathbb{Z}$, where

$$\mathcal{N}_8(u, u_x, u_{xx}, u_{xxx}) := -(4 - \partial_{xx})\partial_x[(\partial_u f)(u)], \quad (1.2)$$

the "Hamiltonian density" f belongs to $C^\infty(\mathbb{R}, \mathbb{R})$ and is such that

$$f(u) = O(u^9), \quad (1.3)$$

where $O(u^9)$ denotes a function with a zero of order at least nine at the origin. The Eq. (1.1) is a Hamiltonian PDE of the form $u_t = J\nabla H(u)$ where ∇H is the $L^2(\mathbb{T}, \mathbb{R})$ gradient and the function

$$H(u) = \int \frac{u^2}{2} - \frac{u^3}{6} + f(u)\,dx, \quad J = (1 - \partial_{xx})^{-1}(4 - \partial_{xx})\partial_x \quad (1.4)$$

is defined on the phase space $H^1_0(\mathbb{T}) := \{u \in H^1(\mathbb{T}, \mathbb{R}) : \int_{\mathbb{T}} u\,dx = 0\}$. The Eq. (1.1) for $f = 0$ is the DP equation which was first proposed in [29] in the form

$$u_t + c_0 u_x + \gamma u_{xxx} - \alpha^2 u_{xxt} = \left(-\frac{2c_1}{\alpha^2}u^2 + c_2(u_x^2 + uu_{xx})\right)_x, \quad (1.5)$$

where $c_0, c_1, c_2, \gamma, \alpha \in \mathbb{R}$, $\alpha \neq 0$. By applying Galilean boosts, translations and time rescaling to (1.5) one obtains Eq. (1.1) with $f = 0$.

The DP equation can be regarded as a model for nonlinear shallow water dynamics and its asymptotic accuracy is the same as for the Camassa–Holm equation and a degree more than the KdV equation [23]. There is a rather large literature on this equation starting form the paper [28] in which the *complete integrability* is proved. The local and global well-posedness, for instance, have been extensively studied as well as existence of wave breaking phenomena (peakons, N-peakons solutions). Without trying to be exhaustive we quote [18,20–22,48,54] and we refer to [32] and references therein for more literature about Degasperis–Procesi equation.

Actually many of these results (notably the wave breaking) are studied in the *dispersionless* case, which corresponds to (1.1) with $f = 0$ and $u \rightsquigarrow u + 1$. In the present paper the presence of the dispersive terms $-4u_x + u_{xxx}$ is fundamental. Our main purpose is to prove existence of quasi-periodic solutions in high Sobolev regularity by following a KAM approach. In this setting a *quasi*-periodic solution with $\nu \in \mathbb{N}$ frequencies is defined by an embedding

$$\mathbb{T}^\nu \ni \varphi \mapsto U(\varphi, x) \in H^1_0(\mathbb{T}, \mathbb{R}) \quad (1.6)$$

and a frequency vector $\omega \in \mathbb{R}^\nu$, with rationally independent entries, such that $u(t,x) = U(\omega t, x)$ is a solution of (1.1) and $U(\varphi, x) \in H^p(\mathbb{T}^{\nu+1}, \mathbb{R})$ for some p sufficiently large.

Notice that, in a neighbourhood of $u = 0$, (1.1) can be seen as a perturbation of the linear PDE

$$v_t - v_{xxt} + v_{xxx} - 4v_x = 0, \tag{1.7}$$

whose bounded solutions have the form

$$v(t,x) = \sum_{j\in\mathbb{Z}} v_j\, e^{\mathrm{i}(\lambda(j)t + jx)}, \qquad \lambda(j) := j\,\frac{4+j^2}{1+j^2} = j + \frac{3j}{1+j^2}, \qquad j \in \mathbb{Z}, \tag{1.8}$$

where $j \mapsto \lambda(j)$ is the *linear* dispersion law. It is easily seen that all solutions of (1.7) with compact Fourier support are periodic, but with period depending on the support. In this context it is natural to investigate whether Eq. (1.1) has periodic or quasi-periodic solutions *close to* to small amplitude linear solutions (1.8). We remark that, since the solutions of (1.8) are all periodic, the existence of quasi-periodic solutions, if any, strongly relies on the presence of the quadratic nonlinearity in (1.1).

In the present paper we construct quasi-periodic solutions mainly supported in Fourier space at $\nu \geq 2$ distinct *tangential sites*

$$S^+ := \{\overline{\jmath}_1, \dots, \overline{\jmath}_\nu\}, \quad S := S^+ \cup (-S^+), \quad \overline{\jmath}_i \in \mathbb{N}\setminus\{0\}, \quad \forall i = 1, \dots, \nu, \tag{1.9}$$

where, without loss of generality, we shall always assume that $\overline{\jmath}_1 = \max_{i=1,\dots,\nu} \overline{\jmath}_i$. We denote by

$$\overline{\omega} := \left(\frac{\overline{\jmath}_1(4+\overline{\jmath}_1^2)}{1+\overline{\jmath}_1^2}, \dots, \frac{\overline{\jmath}_\nu(4+\overline{\jmath}_\nu^2)}{1+\overline{\jmath}_\nu^2}\right) \in \mathbb{Q}^\nu \tag{1.10}$$

the linear frequencies of oscillations related to the tangential sites. More precisely our solutions will have the form

$$u(t,x;\xi) = 2\sum_{i=1}^{\nu} \sqrt{\xi_i}\,\cos(\omega_i t + \overline{\jmath}_i x) + o(\sqrt{|\xi|}), \quad \omega = \overline{\omega} + O(|\xi|), \tag{1.11}$$

where $o(\sqrt{|\xi|})$ is meant in the H^s-topology with s large. It is well know that in looking for quasi-periodic solutions "small divisors" problems arise. To overcome such problems we shall require that S^+ satisfies a *wave packet* condition and that the *unperturbed amplitudes* ξ belong to an appropriate Cantor-like set of positive measure.
The following definition quantifies the *wave packet* condition.

Definition 1.1. For $\mathtt{r} \in (0,1)$, we say that a set of natural numbers $S^+ = \{\overline{\jmath}_1, \dots, \overline{\jmath}_\nu\}$ is in $\mathcal{V}(\mathtt{r})$ if

$$\min_{i=1,\dots,\nu} \overline{\jmath}_i > \frac{1}{\mathtt{r}} \qquad \text{and} \qquad \left|\frac{\overline{\jmath}_i}{\overline{\jmath}_1} - 1\right| \leq \mathtt{r}\,; \tag{1.12}$$

$$\sum_{i=1}^{\nu} \frac{\overline{\jmath}_i}{1+\overline{\jmath}_i^2}\,\ell_i \neq 0, \quad \forall \ell \in \mathbb{Z}^\nu : \quad |\ell| = 4. \tag{1.13}$$

Denoting by $B(0,\varrho)$ the ball centred at the origin of $\mathbb{R}^\nu$ of radius $\varrho > 0$, our result can be stated as follows.

Theorem 1. *Let $\nu \in \mathbb{N}$, $\nu \geq 2$, and consider $f \in C^\infty(\mathbb{R}, \mathbb{R})$ satisfying (1.3). There exists a constant $\mathtt{r}_0 > 0$ such that, for any choice of S^+ in $\mathcal{V}(\mathtt{r})$, with $0 < \mathtt{r} \leq \mathtt{r}_0$, there exist $s \gg 1$, $0 < \varrho \ll 1$ and a positive measure Cantor-like set $\mathfrak{A} \subseteq B(0, \varrho)$ such that the following holds. For any $\xi \in \mathfrak{A}$, the Eq. (1.1) possesses a small amplitude quasi-periodic solution $u(t, x; \xi) = U(\omega t, x; \xi)$ of the form (1.11) where $U(\varphi, x) \in H^s(\mathbb{T}^{\nu+1}, \mathbb{R})$ and $\omega := \omega(\xi) \in \mathbb{R}^\nu$ is a diophantine frequency vector. Moreover for $0 < \varepsilon \leq \sqrt{\varrho}$, the set $\mathfrak{A}$ has asymptotically full relative measure in $[\varepsilon^2, 2\varepsilon^2]^\nu$.*

Moreover we have the following stability result.

Theorem 2 (Linear stability). *The quasi-periodic solutions (1.11) $u(t, x) = U(\omega t, x)$ of Eq. (1.1) are* linearly stable *and* reducible *in the following sense. Consider Eq. (1.1) linearized at the embedded torus $U(\varphi, x)$, then the corresponding operator has purely imaginary spectrum and there exists a change of variables $H^s(\mathbb{T}, \mathbb{R}) \to H^s(\mathbb{T}, \mathbb{R})$, quasi periodic in time with frequency ω, which diagonalizes it in the directions normal to the torus. As a consequence the Cauchy problem of the linearized equation is stable, i.e. the Sobolev norms are uniformly bounded in t.*

Theorems 1, 2 are formulated in the typical style of results on reducible KAM tori for PDEs. For the proof we use the overall strategy of [4], which however has to be substantially developed to deal with (1.1). Let us briefly explain the main new issues.

- The dispersion law is *asymptotically linear* as for the Klein–Gordon equation, studied for instance in [6,7]. As explained in those papers, the fact that the dispersive effects are very weak (essentially time and space play the same role) creates a number of difficulties even in the study of KAM theory for semi-linear PDEs. Of course, since (1.1) is quasi-linear, there are additional serious difficulties coming from the strong perturbative effects of the nonlinearity.
- The DP equation is *resonant* at zero and does not depend on any *external* parameters. This is a fundamental difference w.r.t. the Klein–Gordon equation, where one modulates the mass in order to avoid resonances. Moreover the DP has non-trivial resonances already at order four (see Sect. 1.3), differently from the previous KAM results for quasi-linear PDEs. As a further difficulty the algebraic structure of the resonances is quite complicated. In order to avoid the inherent problems we rely on the presence of "many" (precisely $eight$) approximate constants of motion of (1.1) coming from the integrable structure of the DP equation. Dealing with the problems related to resonances is the core of this paper and requires a set of new ideas and a careful analysis.
- The very strong restriction of the tangential sites S^+ is exploited several times to simplify the problems arising from the rational and asymptotically linear dispersion law. Physically we are looking for solutions mainly supported in Fourier space on modes which are *relatively close* to each other.
  It seems reasonable that such condition could be weakened, but it is not clear to us how to deal with the technical difficulties which would arise.
- As in other resonant cases, the diophantine constant γ is related to the size of the solution one is looking for (see (1.11)). Moreover, due to the linear dispersion law, we are forced to impose very "weak" non-degeneracy conditions on the linear frequencies of oscillations. As a consequence we need a refined bifurcation analysis in order to find a very good first approximate solution and fulfil the smallness conditions required for the Nash–Moser scheme.

Some comments on Eq. (1.1) and on Theorems 1, 2 are in order.

The unperturbed DP equation. We look at (1.1) as a perturbation of the linear equation (1.7), in order to fit the typical perturbative setting of KAM for PDEs , we refer to Sect. 1.1 for more details.

Actually, since the Degasperis–Procesi equation is completely integrable (see [28]) it would be very natural to try to construct solutions of (1.1) which bifurcate from quasi-periodic solutions of the unperturbed DP equation

$$u_t - u_{xxt} + u_{xxx} - 4u_x - uu_{xxx} - 3u_x u_{xx} + 4uu_x = 0, \tag{1.14}$$

which corresponds to (1.1) with $f = 0$. Indeed, near zero, the (1.1) can be seen *also* as a perturbation of (1.14). Unfortunately even though algebro-geometric *finite-gap* solutions have been already constructed in literature for the DP equation (see [42]) it is not clear to us whether they are real quasi-periodic solutions in the sense of (1.6). Of course if one were able to bifurcate from *finite-gap* solutions of (1.14) then it would be possible to prove existence of *large* quasi-periodic solutions, by requiring that f is small. Such a strategy has been followed successfully for the KdV and cubic NLS equation on the circle. Actually for those equations one can prove the existence of Birkhoff coordinates [41,43] (the Cartesian version of action-angle variables), which trivialize the dynamics (in the sense that the solutions turn out to be all periodic, quasi-periodic or almost periodic) and provide a fundamental tool for investigating the dynamical consequences of small perturbative effects, also far from the origin, see [14].

For 1d integrable PDEs one would expect this to be the typical scenario at least in a neighborhood of zero, see [5,46]; however, as far as we know, up to now such results are available only for the KdV, the NLS and the Toda system. Theorem 1 provides, again as far as we know, the first existence result of quasi-periodic solutions, in the sense of (1.6), for (1.14).

It would be interesting to apply our KAM approach to the Camassa–Holm equation, which is a well-known integrable PDE with an asymptotically linear dispersion law, but with a different symplectic structure. Even though we have not performed the computations, we expect to be able to prove the equivalent of Theorems 1, 2 also for this equation. We remark that in this case, the finite gap solutions are known to be quasi-periodic tori, see [20].

One could start by comparing them with the solutions predicted by our method and then possibly develop KAM theory close to large finite gap solutions.

Approximate constants of motion of (1.1). Even though we do not fully exploit the integrability of (1.14) it is fundamental for us that (the non integrable) (1.1) has at least *eight* approximate constants of motion (up to an error of order $O(u^9)$). It is interesting to notice that, as shown in [29], no other equation with the same dispersion law, and the same symplectic structure, has *eight* approximate conserved quantities. This means that in (1.1) we cannot consider any quadratic nonlinearity, but we really need the DP structure.

The request of the presence of such approximate conserved quantities it is not only a technical matter. In order to implement a Nash/Moser-KAM algorithm one looks for a family of approximately invariant tori of (1.1) (with a sufficiently good approximation) such that the dynamics on the tori is integrable and non-degenerate, while the dynamics normal to the torus is non-degenerate at the linear level and satisfies the Melnikov conditions. If there are external parameters modulating the linear frequencies, then we can consider as approximate solutions the linear ones. Otherwise the modulation must come from the initial data and, hopefully, this can be achieved by means of Birkhoff

normal form (BNF), see for instance [4,39]. In this case, where the the dispersion law in (1.8) is a rational number and is asymptotically linear, such procedure is very difficult. One has to explicitly compute some potentially dangerous resonant terms in the Hamiltonian and show that they vanish. This is the same type of computations which have been done for water waves, see Craig–Worfolk [27] where the authors verify (by computing them) the vanishing of the coefficients of fourth order resonant interactions, the so called *Benjamin–Feir resonances*. In our case we have to deal with higher order resonances (up to $eight$), so this would be computationally extremely heavy. Our approach is to use the approximate constant of motions. This will be explained more in detail in Sect. 1.3. Once we have constructed the approximate invariant tori we have to impose the non-degeneracy and Melnikov conditions. Differently form the KdV case, this will not be possible for any choice of the tangential set, and it is where we will use the condition $S^+ \in \mathcal{V}(\mathtt{r})$, see Definition 1.1.

Linear stability The linear stability result of Theorem 2 is of course a relevant dynamical information in the study of evolutionary PDEs, but it is also the consequence of a fundamental ingredient of our proof: the reducibility of the linearized equation at any quasi-periodic approximate solution. Reducibility for the Degasperis–Procesi equation linearized at a quasi-periodic function has been obtained in [33], under some appropriate diophantine conditions on the frequencies. Unfortunately, due to the resonances, our case does not fit such hypotheses, and a major point will be to overcome this difficulty. Here we shall use such result (appropriately adapted) inside a nonlinear algorithm to prove the existence of quasi-periodic solutions. This is a classical feature of the literature of KAM theory.

1.1. Some literature. Proving existence and stability for quasi-periodic solutions for PDEs close to an elliptic fixed point is a natural extension of the classical KAM theory for lower dimensional tori [51]. The first results in this direction were for model PDEs on an interval with no derivatives in the nonlinearity and with either Dirichlet, [44,47,51,53] or periodic, [16,19,26], boundary conditions. For extension of KAM theory to higher spatial dimension we mention [8,11,17,25,30,34,52]. While KAM methods for constructing quasi-periodic solutions for PDEs on the circle with no derivatives in the nonlinearity are by now well established, generalizing to cases with derivatives is in general not at all trivial, even in the semi-linear cases (where the derivatives in the nonlinearity are of lower order w.r.t. the linear terms). We mention [45] for the KdV, [49] for the derivative NLS, and [6,7] for the derivative NLW. Recently an innovative strategy was proposed, [3,4] to deal with quasi-linear and fully nonlinear PDEs on the circle. This approach was first developed for the KdV equation but can be applied to many equations of interest in hydrodynamics, such as NLS, [37,38] Kirchhoff [50] or directly the water wave equation [2,15]. While these methods were first thought for PDEs on the circle, of course a very interesting point is the generalization to higher dimensions.

Equation (1.1) is a quasi-linear PDE on the circle and in our study we shall follow the general strategy of [4], extended and adapted to our case. Let us briefly explain the point of view of [4], referring also to [2] for more details.

1.2. The general strategy. We describe the strategy to prove existence and linear stability for small, reducible quasi-periodic solutions of completely resonant quasi-linear PDEs.

(*i*) The starting point is a Nash–Moser theorem of hypothetical conjugation following [9]. The strategy is to construct quadratically convergent sequence of families of approximately invariant (isotropic) tori. Such construction is based on *tame* estimates on the inverse of the operator associated to the Eq. (1.1) linearized at an approximate torus and restricted to the normal direction. This is proved by exploiting the Hamiltonian structure and exhibiting symplectic variables adapted to each approximate invariant torus, which essentially decouple the linearized dynamics. Then the bounds on the inverse are achieved by removing all the "bad" values of the parameters. We mention also [24] for a parallel strategy which does not rely on the Hamiltonian structure.

(*ii*) To construct the sequence of item (*i*) we need a good starting point, i.e. a first family of approximately invariant tori parametrized by real vectors $\xi \in \mathbb{R}^\nu$.

As explained before this is achieved by BNF techniques. In particular, in the quasi-linear context, it is convenient to perform a *Weak* BNF, i.e. to exhibit a change of variables, close to the identity up to a finite rank operator, such that the following holds. The Hamiltonian H transforms to $H_{\text{Birk}} + R$ where R is a small remainder, and

1. The finite dimensional subspace $U_S := \{u_j = 0,\ \forall j \notin S\}$ is invariant for H_{Birk};
2. The Hamiltonian restricted to U_S is integrable and non-degenerate in the sense that the "frequency-to-amplitude" map is invertible.

In order to describe in a simpler way the dynamics in a neighborhood of U_S it is convenient to define *action-angle* variables. This allows to distinguish the tangential and normal dynamics to the approximately invariant tori.

We remark that, for semi-linear PDEs, typically one performs a stronger BNF preliminary step, in order to "normalize" also the linearized dynamics normal to the torus, i.e. the terms in the Hamiltonian which are quadratic in the normal directions. In this case the Birkhoff map is close to the identity up to a bounded operator (at most one-smoothing), see for instance [47,51]. Compared to the latter approach, the weak procedure has the disadvantage that the normal form depends on the angles; on the other hand we do not have to address well-posedness issues, since these changes of coordinates are time-one flow maps of an ODE. Note that the recent papers, [10,35,36] directly study the full Birkhoff normal form for quasi-linear PDES.

(*iii*) The third key point is to study the invertibility of the linearized operator restricted to the normal directions. Thanks to the very "mild" conjugation procedure of item (*ii*) (with a map = identity+finite rank) it turns out that such linear operator is *pseudo differential* (with non constant coefficients) up to a finite rank remainder. This is the most important reason for adopting the weak procedure described in (*ii*).

The invertibility of the linearized operator, with appropriate tame estimates, is based on a *reducibility* argument which is divided into two parts:

(a) A reduction in decreasing order procedure which conjugates the linearized operator to a pseudo differential one with constant coefficients up to a remainder which is a bounded/regularizing term i.e. maps $H^s(\mathbb{T}, \mathbb{R})$ to $H^{s+\rho}(\mathbb{T}, \mathbb{R})$, $\rho \geq 0$. The choice of ρ depends of course on the problem one is studying;
(b) A quadratic KAM scheme (for *bounded* operators) which completely diagonalizes the bounded/smoothing remainder of the previous step.

We want to point out the following:

- The step (a) strongly relies on the pseudo differential structure of the operator;
- The normal form contains angle-dependent terms and some of them turn out to be not perturbative for the KAM scheme (b). The conjugation to constant coefficients of such terms relies on purely algebraic arguments. We refer to this procedure as *linear Birkhoff normal form*;
- As a consequence of having applied the weak and the linear Birkhoff procedure, the normal form around the approximately invariant tori has constant coefficients also in the normal directions.

In order to perform the diagonalization procedure of step (b) one needs the *second Melnikov conditions*, which essentially amount to requiring that the operator has simple eigenvalues with a lower bound on the differences. Once one has diagonalized the operator, the bounds on the inverse follow trivially from lower bounds on the eigenvalues, i.e. *first Melnikov conditions*.

(iv) In the scheme above, at each step we have removed some bad values of the parameters ξ where the Melnikov conditions do not hold. Hence the last (but not least) step is to prove that at the end of the procedure one has still a positive measure set of parameters. Note that often it is more convenient to express such conditions in terms of the frequency of the quasi-periodic solution. This can be done thanks to the invertibility of the frequency-to-amplitude map.

1.3. Main novelties and scheme of the proof. We describe the structure of the paper following Sect. 1.2, and with particular attention to the main novelties.

In Sect. 2 we introduce the Hamiltonian formalism for the DP equation and the functional spaces on which we shall work.

In Sect. 3 we perform the weak Birkhoff normal form explained in item (ii) of the previous section. The result is stated in Proposition 3.2. In order to reach a sufficiently good first approximate solution we need to perform 6-BNF steps. As is well-known, at the n-th step of this procedure one has to take into account the denominators (recall (1.8))

$$\lambda(j_1) + \cdots + \lambda(j_{n+2}). \tag{1.15}$$

We say that a $(n+2)$-uple of integer indices $(j_1, \ldots, j_{n+2})$ is a resonance, and hence may appear in H_{Birk}, if (1.15)$= 0$ and the momentum condition holds, namely $\sum_{i=1}^{n+2} j_i = 0$. We say that a resonance is *trivial* if it has the form $(i, -i, j, -j, \ldots)$ so that the corresponding monomial is integrable.

As mentioned before a major difficulty comes from the fact that the DP equation has many non-trivial resonances (already at order four) and in principle there is no reason why the Birkhoff Hamiltonian restricted to U_S should be integrable. By the fact that the Hamiltonian density f is of order $O(u^9)$ the perturbation does not affect the leading terms of the Birkhoff Hamiltonian and we can exploit the integrability of the DP equation. Indeed the same Birkhoff transformation should normalize simultaneously all the commuting Hamiltonians. This means that a resonant monomial contributes to H_{Birk} if and only if it is resonant for all the constants of motion. This was proved in detail in [32] at the level of formal power series. Here we adapt this result to the Eq. (1.1) which is only approximately integrable (close to the origin) and we reformulate it in a way better suited to the weak Birkhoff normal form context, see Proposition 3.6.

Once we have shown that the H_{Birk}-dynamics restricted to U_S is integrable, in Sect. 4, we prove that it is non-degenerate, i.e. that the frequency to amplitude map is a diffeomorphism. We have a very explicit description of this map and hence this step amounts to

proving that the matrix $\mathbb{A}$ in (4.6) (which depends only on S^+) has determinant bounded away from zero (the so-called twist condition), see Lemma 4.1. A big difference with [4] is that, in our case, the determinant of $\mathbb{A}$ is a rational function of several variables $\overline{\jmath}_i$ that could accumulate to zero as $|\overline{\jmath}_i| \to \infty$. By imposing the wave packet condition we restrict the study of its asymptotic behaviour to regions in which it behaves like a one variable function. Then we use continuity arguments to guarantee the invertibility of $\mathbb{A}$ for every choice of $S^+ \in \mathcal{V}(r)$ (see Definition 1.1) for r small enough. Outside $\mathcal{V}(r)$ the proof of lower bounds for $\det \mathbb{A}$ should rely on purely algebraic arguments and not on perturbative ones.

In Sect. 5 we introduce the Nash Moser hypothetical conjugation theorem (see Theorem 5.4) and in Sect. 6 we explain how to prove the invertibility of the linearized operator at an approximate solution by only studying it in the normal direction. Since there is no difference with [4] we only give a synopsis.

In Sects. 7 and 7.3 we prove the Theorems 7.1 and 7.13 which provide the reducibility of the linearized operator following item (iii) of Sect. 1.2. As we already mentioned, in [33] we provide a reducibility result for the DP equation (1.1) linearized at sufficiently small quasi-periodic functions under appropriate diophantine conditions on the frequencies . Unfortunately in our case the diophantine constant γ is related to the size of the approximate solutions (see (5.3)) and then the smallness and diophantine conditions above cannot be met.

In [4] this issue appears only in the step (b) of the strategy, where it is solved by the linear Birkhoff normal form method. A first difficulty in our case is that this problem appears also in step (a). So that we first need to perform some preliminary steps (see Sect. 7.1), more precisely we need changes of coordinates, preserving the pseudo differential structure, that conjugate the leading order of the linearized operator to a diagonal one plus a correction, which is unbounded but perturbative in the sense of [33]. In such steps the provided changes of coordinates are similar in structure to those of step (a) but they are proved to be well-defined not by using perturbative arguments, but by algebraic computations involving the Birkhoff resonances (see Lemma A.1). These difficulties appear also for the quasi-linear generalized KdV [39], but here we have several further problems due to the complexity of the symplectic structure of the DP equation. The first step, removing terms of order ε, is straightforward. Already at the second step we encounter the difficulties arising form the presence of non-trivial resonances of order 4, and a priori there is no reason why the normal form should be integrable. Here it does not appear simple to apply the strategy of the weak BNF, using the constants of motion. On the other hand, computing the normal form explicitly by hand, as done in [39], is unmanageable. To bypass this problem we take a different point of view, based on an *a posteriori identification argument* of normal forms. More precisely in Theorem 7.9 we prove that the normal form obtained after the weak BNF, the preliminary steps and the linear BNF coincides with the one that we would obtain by performing the full *formal* BNF and then projecting on the quadratic terms in the normal variables. This result strongly relies on the fact that all the resonances contributing to the formal normal form are trivial. A similar identification argument has been used, for instance, in [12,13].

A further point is that, due to the rational dispersion law $\lambda(j)$, it is possible that a *denominator* in the linear BNF is not zero but is still uncontrollably small. In the third step, in order to deal with this problem we need to take into account in the unperturbed Hamiltonian also the integrable terms of order ε^2 coming from the previous steps of linear BNF. For this reason it is important to know the exact expression of the main order of the correction at the eigenvalues given by the perturbation, see for instance (5.5). This is

also needed in the KAM scheme (b), in order to impose the second Melnikov conditions. Computing these corrections by hand would be a very difficult task, but this comes for free from Theorem 7.9.

In the first part of Sect. 8 we show the convergence of the Nash–Moser algorithm (see Theorem 8.1), which requires the ratio between the size of $R = H - H_{\text{Birk}}$ and $\gamma^{7/2}$ to be small (see the smallness condition (8.5)); in the second part we prove that the set of "bad" parameters, i.e. the frequencies which do not meet the first and second Melnikov conditions, has small measure (see (8.25), note that such sets are indexed by three parameters ℓ, j, k).

In Lemma 8.4 we provide the measure of the single bad set. Here we use the algebraic arguments provided by Lemma A.1, which guarantees the non-degeneracy of the leading terms of the small divisors. In Sect. 8.1.2 we deal with the summability of the bad sets in j, k for fixed ℓ.

The key difficulty is that the spectral gap $\lambda(j) - \lambda(k)$ is asymptotically constant, hence there is a bad separation property of the eigenvalues. The same occurs for the wave equation [6,7]. Due to the asymptotically constant spectral gap, these sets are infinitely many. Then the key ingredient is to show that for j, k sufficiently large the second Melnikov conditions are implied by the first ones. This is possible provided that we consider two different diophantine constants. More precisely we have to impose second order Melnikov conditions with $\gamma^{3/2}$ (see (8.6)), which is clearly much smaller than γ. This is why we have to perform many steps of Birkhoff normal form in order to obtain a very good first approximate solution.

We point out that, differently from [2], our Melnikov conditions do not imply a loss of regularity in space. In [2] this loss is acceptable, since in the regularization step ((a) p. 5) the diagonalization is performed up to a very smoothing remainder. In this procedure it is fundamental that the diophantine constant γ is independent of the size of the solution. Of course in our case this is not true and thus in the regularization step we end up with a remainder of order -1, and then in the measure estimates we put some extra efforts to prove second Melnikov conditions without loss of regularity.

2. Functional Setting

Hamiltonian formalism of the Degasperis–Procesi equation For any u, v in the space

$$H^1_0(\mathbb{T}) := \left\{ u \in L^2(\mathbb{T}, \mathbb{R}) : \int_{\mathbb{T}} u \, dx = 0 \right\}$$

we define the non-degenerate symplectic form

$$\Omega(u, v) := \int_{\mathbb{T}} (J^{-1} u)\, v \, dx = (J^{-1} u, v)_{L^2} \tag{2.1}$$

where J is defined in (1.4) and $(\cdot, \cdot)_{L^2}$ is the $L^2(\mathbb{T}, \mathbb{R})$ scalar product. To any C^1 function $H : H^1_0(\mathbb{T}) \to \mathbb{R}$ we associate a vector field X_H by requiring

$$dH(u)[h] = (\nabla H(u), h)_{L^2} = \Omega(X_H(u), h), \quad \forall\, u, h \in H^1_0(\mathbb{T}).$$

The Hamiltonian vector field X_H is uniquely determined since the symplectic form Ω in (2.1) is non-degenerate, in particular $X_H(u) = J \nabla H(u)$. The *Poisson bracket* between two C^1 functions $F, G : H^1_0(\mathbb{T}) \to \mathbb{R}$ is

$$\{F, G\} := \Omega(X_F, X_G) = (\nabla F, J \nabla G)_{L^2}. \tag{2.2}$$

In this way

$$X_{\{F,G\}} = [X_F, X_G], \quad \text{where} \quad [X, Y] := dX[Y] - dY[X]. \tag{2.3}$$

Finally, given a Hamiltonian H we define its *adjoint action* as the operator

$$\mathrm{ad}_H[\cdot] := \{H, \cdot\}. \tag{2.4}$$

Consider now two Hamiltonians H, G and let Φ_G be the time-1 flow map of the vector field X_G. Then we have (formally)

$$\begin{aligned} H \circ \Phi_G &= \sum_{k\geq 0} \frac{(-1)^k}{k!} \mathrm{ad}_G^k[H], \qquad H \circ \Phi_G^{-1} = \sum_{k\geq 0} \frac{1}{k!} \mathrm{ad}_G^k[H], \\ \mathrm{ad}_G^k[H] &:= \mathrm{ad}_G\big[\mathrm{ad}_G^{k-1}[H]\big], \end{aligned} \tag{2.5}$$

where $\mathrm{ad}_G^0 := \mathrm{I}$ is the identity map.

Functional space We consider functions $u(\varphi, x)$ defined on $\mathbb{T}^\nu \times \mathbb{T}$. Passing to the Fourier representation

$$u(\varphi, x) = \sum_{j\in\mathbb{Z}} u_j(\varphi)\, e^{\mathrm{i}jx} = \sum_{\ell\in\mathbb{Z}^\nu, j\in\mathbb{Z}} u_{\ell j}\, e^{\mathrm{i}(\ell\cdot\varphi + jx)}, \quad \overline{u}_j(\varphi) = u_{-j}(\varphi), \quad \overline{u}_{\ell j} = u_{-\ell,-j}. \tag{2.6}$$

We define the scale of Sobolev spaces

$$H^s := \Big\{ u(\varphi, x) \in L^2(\mathbb{T}^{\nu+1}, \mathbb{R}) : \|u\|_s^2 := \sum_{\ell\in\mathbb{Z}^\nu, j\in\mathbb{Z}} |u_{\ell j}|^2 \langle \ell, j\rangle^{2s} < \infty \Big\} \tag{2.7}$$

where $\langle \ell, j\rangle := \max\{1, |\ell|, |j|\}$, $|\ell| := \sum_{i=1}^\nu |\ell_i|$. We shall work on the phase space $H^s \cap H_0^1(\mathbb{T}, \mathbb{R})$. We denote by $\mathfrak{B}_r(0, X)$ the ball of radius r centered at the origin of a Banach space X.

Lipschitz norm Fix $\nu \in \mathbb{N}^* := \mathbb{N}\backslash\{0\}$ and let $\mathcal{O}$ be a compact subset of $\mathbb{R}^\nu$. For a function $u\colon \mathcal{O} \to E$, where $(E, \|\cdot\|_E)$ is a Banach space, we define the sup-norm and the lip-seminorm of u as

$$\begin{aligned} \|u\|_E^{sup} &:= \|u\|_E^{sup,\mathcal{O}} := \sup_{\omega\in\mathcal{O}} \|u(\omega)\|_E, \\ \|u\|_E^{lip} &:= \|u\|_E^{lip,\mathcal{O}} := \sup_{\substack{\omega_1,\omega_2\in\mathcal{O},\\ \omega_1\neq\omega_2}} \frac{\|u(\omega_1) - u(\omega_2)\|_E}{|\omega_1 - \omega_2|}. \end{aligned} \tag{2.8}$$

If E is finite dimensional, for any $\gamma > 0$ we introduce the weighted Lipschitz norm

$$\|u\|_E^{\gamma,\mathcal{O}} := \|u\|_E^{sup,\mathcal{O}} + \gamma \|u\|_E^{lip,\mathcal{O}}. \tag{2.9}$$

If E is a scale of Banach spaces, say $E = H^s$, for $\gamma > 0$ we introduce the weighted Lipschitz norm

$$\|u\|_s^{\gamma,\mathcal{O}} := \|u\|_s^{sup,\mathcal{O}} + \gamma \|u\|_{s-1}^{lip,\mathcal{O}}, \quad \forall s \geq [\nu/2] + 4 \tag{2.10}$$

where we denoted by $[r]$ the integer part of $r \in \mathbb{R}$.

Linear operators Let $A\colon \mathbb{T}^\nu \to \mathcal{L}(L^2(\mathbb{T},\mathbb{R}))$, $\varphi \mapsto A(\varphi)$, be a φ-dependent family of linear operators acting on $L^2(\mathbb{T},\mathbb{R})$. We consider A as an operator acting on $H^s(\mathbb{T}^{\nu+1},\mathbb{R})$ by setting

$$(Au)(\varphi,x) = (A(\varphi)u(\varphi,\cdot))(x).$$

This action is represented in Fourier coordinates as

$$Au(\varphi,x) = \sum_{j,j'\in\mathbb{Z}} A_j^{j'}(\varphi)\, u_{j'}(\varphi)\, e^{\mathrm{i}jx} = \sum_{\ell\in\mathbb{Z}^\nu, j\in\mathbb{Z}} \sum_{\ell'\in\mathbb{Z}^\nu, j'\in\mathbb{Z}} A_j^{j'}(\ell-\ell')\, u_{j'\ell'}\, e^{\mathrm{i}(\ell\cdot\varphi+jx)}. \tag{2.11}$$

Conversely, given a Töpliz in time operator A, namely such that its matrix coefficients (with respect to the Fourier basis in φ, x) satisfy

$$A_{j,\ell}^{j',\ell'} = A_j^{j'}(\ell-\ell') \qquad \forall j,\, j'\in\mathbb{Z},\ \ell,\ell'\in\mathbb{Z}^\nu, \tag{2.12}$$

we can associate it a time dependent family of operators acting on $H^s(\mathbb{T})$ by setting

$$A(\varphi)h = \sum_{j,j'\in\mathbb{Z},\ell\in\mathbb{Z}^\nu} A_j^{j'}(\ell) h_{j'}\, e^{\mathrm{i}jx} e^{\mathrm{i}\ell\cdot\varphi}, \qquad \forall h\in H^s(\mathbb{T},\mathbb{R}).$$

For $m = 1,\dots,\nu$ we define the operators $\partial_{\varphi_m} A(\varphi)$ as

$$(\partial_{\varphi_m} A(\varphi))u(\varphi,x) = \sum_{\ell\in\mathbb{Z}^\nu, j\in\mathbb{Z}} \sum_{\ell'\in\mathbb{Z}^\nu, j'\in\mathbb{Z}} \mathrm{i}(\ell_m-\ell'_m)\, A_j^{j'}(\ell-\ell')\, u_{\ell' j'}\, e^{\mathrm{i}(\ell\cdot\varphi+jx)}. \tag{2.13}$$

We say that A is a *real* operator if it maps real valued functions in real valued functions. For the matrix coefficients this means that

$$\overline{A_j^{j'}(\ell)} = A_{-j}^{-j'}(-\ell).$$

Hamiltonian linear operators In the paper we shall deal with operators which are Hamiltonian according to the following Definition.

Definition 2.1. We say that a linear map is symplectic if it preserves the 2-form Ω in (2.1); similarly we say that a linear operator M is Hamiltonian if Mu is a linear Hamiltonian vector field w.r.t. Ω in (2.1). This means that each $J^{-1}M$ is symmetric respect to the real L^2-scalar product. Similarly, we call a family of maps $\varphi \to A(\varphi)$ symplectic if, for each fixed φ, $A(\varphi)$ is symplectic, same for the Hamiltonians. We shall say that an operator of the form $\omega\cdot\partial_\varphi + M(\varphi)$ is Hamiltonian if $M(\varphi)$ is Hamiltonian.

Notation. We use the notation $A \lesssim B$ to denote $A \le C\,B$ where C is a positive constant possibly depending on fixed parameters given by the problem. We use the notation $A \lesssim_y B$ to denote $A \le C(y)B$ if we wish to highlight the dependence on the variable y of the constant $C(y) > 0$.

Linear Tame operators Here we introduce rigorously the spaces and the classes of operators on which we work.

Definition 2.2 (*σ-Tame operators*). Given $\sigma \geq 0$ we say that a linear operator A is *σ-tame* w.r.t. a non-decreasing sequence $\{\mathfrak{M}_A(\sigma, s)\}_{s=s_0}^{\mathcal{S}}$ (with possibly $\mathcal{S} = +\infty$) if

$$\|Au\|_s \leq \mathfrak{M}_A(\sigma, s)\|u\|_{s_0+\sigma} + \mathfrak{M}_A(\sigma, s_0)\|u\|_{s+\sigma} \qquad u \in H^s, \tag{2.14}$$

for any $s_0 \leq s \leq \mathcal{S}$. We call $\mathfrak{M}_A(\sigma, s)$ a TAME CONSTANT for the operator A. When the index σ is not relevant we write $\mathfrak{M}_A(\sigma, s) = \mathfrak{M}_A(s)$.

Definition 2.3 (*Lip-σ-Tame operators*). Let $\sigma \geq 0$ and $A = A(\omega)$ be a linear operator defined for $\omega \in \mathcal{O} \subset \mathbb{R}^\nu$. Let us define

$$\Delta_{\omega,\omega'} A := \frac{A(\omega) - A(\omega')}{|\omega - \omega'|}, \quad \omega, \omega' \in \mathcal{O}. \tag{2.15}$$

Then A is *Lip-σ-tame* w.r.t. a non-decreasing sequence $\{\mathfrak{M}_A(\sigma, s)\}_{s=s_0}^{\mathcal{S}}$ if the following estimate holds

$$\sup_{\omega\in\mathcal{O}} \|Au\|_s, \gamma \sup_{\omega\neq\omega'} \|(\Delta_{\omega,\omega'} A)\|_{s-1} \leq_s \mathfrak{M}_A^\gamma(\sigma, s)\|u\|_{s_0+\sigma} + \mathfrak{M}_A^\gamma(\sigma, s)\|u\|_{s+\sigma}, \quad u \in H^s. \tag{2.16}$$

We call $\mathfrak{M}_A^\gamma(\sigma, s)$ a LIP- TAME CONSTANT of the operator A. When the index σ is not relevant we write $\mathfrak{M}_A^\gamma(\sigma, s) = \mathfrak{M}_A^\gamma(s)$.

Modulo-tame operators and majorant norms The modulo-tame operators are introduced in Sect. 2.2 of [15]. Note that we are interested only in the Lipschitz variation of the operators respect to the parameters of the problem, whereas in [15] the authors need to control also higher order derivatives.

Definition 2.4. Let $u \in H^s$, $s \geq 0$, we define the majorant function $\underline{u}(\varphi, x) := \sum_{\ell\in\mathbb{Z}^\nu, j\in\mathbb{Z}} |u_{\ell j}| e^{\mathrm{i}(\ell\cdot\varphi + jx)}$. Note that $\|u\|_s = \|\underline{u}\|_s$.

Definition 2.5 (*Majorant operator*). Let $A \in \mathcal{L}(H^s)$ and recall its matrix representation (2.11). We define the majorant matrix $\underline{A}$ as the matrix with entries

$$(\underline{A})_j^{j'}(\ell) := |(A)_j^{j'}(\ell)| \qquad j, j' \in \mathbb{Z}, \ \ell \in \mathbb{Z}^\nu.$$

We consider the majorant operatorial norms

$$\|\underline{M}\|_{\mathcal{L}(H^s)} := \sup_{\|u\|_s \leq 1} \|\underline{M}u\|_s. \tag{2.17}$$

We have a partial ordering relation in the set of the infinite dimensional matrices, i.e. if

$$\begin{aligned} M \preceq N \Leftrightarrow |M_j^{j'}(\ell)| \leq |N_j^{j'}(\ell)| \ \ \forall j, j', \ell \ \Rightarrow \|\underline{M}\|_{\mathcal{L}(H^s)} \leq \|\underline{N}\|_{\mathcal{L}(H^s)}, \\ \|Mu\|_s \leq \|\underline{M}\,\underline{u}\|_s \leq \|\underline{N}\,\underline{u}\|_s. \end{aligned} \tag{2.18}$$

Since we are working on a majorant norm we have the continuity of the projections on monomial subspace, in particular we define the following functor acting on the matrices

$$\Pi_K M := \begin{cases} M_j^{j'}(\ell) & \text{if } |\ell| \leq K, \\ 0 & \text{otherwise} \end{cases} \qquad \Pi_K^\perp := \mathrm{I} - \Pi_K.$$

Finally we define for $\mathfrak{b}_0 \in \mathbb{N}$

$$(\langle\partial_\varphi\rangle^{\mathfrak{b}_0} M)_j^{j'}(\ell) := \langle\ell\rangle^{\mathfrak{b}_0} M_j^{j'}(\ell). \tag{2.19}$$

In the sequel let $1 > \gamma > \gamma^{3/2} > 0$ be fixed constants.

Definition 2.6 (*Lip-σ-modulo tame*). Let $\sigma \geq 0$. A linear operator $A := A(\omega)$, $\omega \in \mathcal{O} \subset \mathbb{R}^\nu$, is Lip-$\sigma$-modulo-tame w.r.t. a non-decreasing sequence $\{\mathfrak{M}_A^{\sharp,\gamma^{3/2}}(\sigma, s)\}_{s=s_0}^{\mathcal{S}}$ if the majorant operators $\underline{A}$, $\underline{\Delta_{\omega,\omega'} A}$ are Lip-σ-tame w.r.t. these constants, i.e. they satisfy the following weighted tame estimates: for $\sigma \geq 0$, for all $s \geq s_0$ and for any $u \in H^s$,

$$\sup_{\omega\in\mathcal{O}} \|\underline{A}u\|_s, \quad \sup_{\omega\neq\omega'\in\mathcal{O}} \gamma^{3/2}\|\underline{\Delta_{\omega,\omega'} A}u\|_s \leq \mathfrak{M}_A^{\sharp,\gamma^{3/2}}(\sigma, s_0)\|u\|_{s+\sigma} + \mathfrak{M}_A^{\sharp,\gamma^{3/2}}(\sigma, s)\|u\|_{s_0+\sigma}. \tag{2.20}$$

The constant $\mathfrak{M}_A^{\sharp,\gamma^{3/2}}(\sigma, s)$ is called the MODULO- TAME CONSTANT of the operator A. When the index σ is not relevant we write $\mathfrak{M}_A^{\sharp,\gamma^{3/2}}(\sigma, s) = \mathfrak{M}_A^{\sharp,\gamma^{3/2}}(s)$.

Definition 2.7. We say that A is Lip--1-modulo tame if $\langle D_x\rangle^{1/2} A \langle D_x\rangle^{1/2}$ is Lip-0-modulo tame. We denote

$$\begin{aligned} \mathfrak{M}_A^{\sharp,\gamma^{3/2}}(-1, s) &:= \mathfrak{M}_{\langle D_x\rangle^{1/2} A \langle D_x\rangle^{1/2}}^{\sharp,\gamma^{3/2}}(0, s), \\ \mathfrak{M}_A^{\sharp,\gamma^{3/2}}(-1, s, a) &:= \mathfrak{M}_{\langle \partial_\varphi\rangle^a \langle D_x\rangle^{1/2} A \langle D_x\rangle^{1/2}}^{\sharp,\gamma^{3/2}}(0, s), \quad a \geq 0. \end{aligned} \tag{2.21}$$

In the following we shall systematically use -1 modulo-tame operators. We refer the reader to the "Appendix" of [33] for the properties of Tame and Modulo-tame operators.
Pseudo differential operators Following [15] we give the following definitions.

Definition 2.8. Let $m \in \mathbb{R}$. A linear operator A is called pseudo differential of order $\leq m$ if its action on any $H^s(\mathbb{T}, \mathbb{R})$ with $s \geq m$ is given by

$$A \sum_{j\in\mathbb{Z}} u_j e^{\mathrm{i}jx} = \sum_{j\in\mathbb{Z}} a(x, j) u_j e^{\mathrm{i}jx},$$

where $a(x, j)$, called the *symbol* of A, is the restriction to $\mathbb{T} \times \mathbb{Z}$ of a complex valued function $a(x, y)$ which is C^∞ smooth on $\mathbb{T} \times \mathbb{R}$, 2π-periodic in x and satisfies

$$|\partial_x^\alpha \partial_y^\beta a(x, y)| \leq C_{\alpha,\beta} \langle y\rangle^{m-\beta}, \quad \forall\, \alpha, \beta \in \mathbb{N}. \tag{2.22}$$

We denote by $A[\cdot] = \mathrm{Op}(a)[\cdot]$ the pseudo operator with symbol $a := a(x, j)$. We call OPS^m the class of the pseudo differential operator of order less or equal to m and $OPS^{-\infty} := \bigcap_m OPS^m$. We define the class S^m as the set of symbols which satisfies (2.22).

We will consider mainly operators acting on $H^s(\mathbb{T}, \mathbb{R})$ with a quasi-periodic time dependence. In the case of pseudo differential operators this corresponds[1] to considering symbols $a(\varphi, x, y)$ with $\varphi \in \mathbb{T}^\nu$. Clearly these operators can be thought as acting on functions $u(\varphi, x) = \sum_{j\in\mathbb{Z}} u_j(\varphi) e^{\mathrm{i}jx}$ in $H^s(\mathbb{T}^{\nu+1}, \mathbb{R})$ in the following sense:

$$(Au)(\varphi, x) = \sum_{j\in\mathbb{Z}} a(\varphi, x, j) u_j(\varphi) e^{\mathrm{i}jx}, \quad a(\varphi, x, j) \in S^m.$$

The symbol $a(\varphi, x, y)$ is C^∞ smooth also in the variable φ. We still denote $A := A(\varphi) = \mathrm{Op}(a(\varphi, \cdot)) = \mathrm{Op}(a)$.

[1] Since ω is diophantine we can replace the time variable with angles $\varphi \in \mathbb{T}^\nu$. The time dependence is recovered by setting $\varphi = \omega t$.

Definition 2.9. Let $a := a(\varphi, x, y) \in S^m$ and set $A := \mathrm{Op}(a) \in OPS^m$,

$$|A|_{m,s,\alpha} := \max_{0\le\beta\le\alpha} \sup_{y\in\mathbb{R}} \|\partial_y^\beta a(\cdot,\cdot,y)\|_s \langle y\rangle^{-m+\beta}. \tag{2.23}$$

We will use also the notation $|a|_{m,s,\alpha} := |A|_{m,s,\alpha}$.

Note that the norm $|\cdot|_{m,s,\alpha}$ is non-decreasing in s and α. Moreover given a symbol $a(\varphi, x)$ independent of y, the norm of the associated multiplication operator $\mathrm{Op}(a)$ is just the H^s norm of the function a. If on the contrary the symbol $a(y)$ depends only on y, then the norm of the corresponding Fourier multipliers $\mathrm{Op}(a(y))$ is just controlled by a constant.

As in formula (2.10), if $A = \mathrm{Op}(a(\omega, \varphi, x, y)) \in OPS^m$ is a family of pseudo differential operators with symbols $a(\omega, \varphi, x, y)$ belonging to S^m and depending in a Lipschitz way on some parameter $\omega \in \mathcal{O} \subset \mathbb{R}^\nu$, we set

$$|A|^{\gamma,\mathcal{O}}_{m,s,\alpha} := \sup_{\omega\in\mathcal{O}} |A|_{m,s,\alpha} + \gamma \sup_{\omega_1,\omega_2\in\mathcal{O}} \frac{|\mathrm{Op}\big(a(\omega_1,\varphi,x,y) - a(\omega_2,\varphi,x,y)\big)|_{m,s-1,\alpha}}{|\omega_1-\omega_2|}. \tag{2.24}$$

For the properties of compositions, adjointness and quantitative estimates of the actions on the Sobolev spaces H^s of pseudo differential operators we refer to "Appendix B" of [33].

3. Weak Birkhoff Normal Form

The aim of this section is to construct a ξ-parameter family of approximately invariant, finite dimensional tori supporting quasi-periodic motions with frequency $\omega(\xi)$. We will impose the map $\xi \mapsto \omega(\xi)$ to be a diffeomorphism and we will consider such approximate solutions as the starting point for the Nash–Moser algorithm.
In order to state the main result of this section, we need some preliminary definitions. We write the DP Hamiltonian in (1.4) in the following way:

$$\begin{aligned} &H(u) = H^{(2)}(u) + H^{(3)}(u) + H^{(\ge 9)}, \\ &H^{(2)}(u) := \frac{1}{2}\int_{\mathbb{T}} u^2\,dx, \quad H^{(3)}(u) := -\frac{1}{6}\int_{\mathbb{T}} u^3\,dx, \quad H^{(\ge 9)}(u) := \int_{\mathbb{T}} f(u)\,dx. \end{aligned} \tag{3.1}$$

Recall S in (1.9) and define $S^c := \mathbb{Z}\backslash\big(S\cup\{0\}\big)$. We decompose the phase space as

$$H^1_0(\mathbb{T}) := H_S \oplus H_S^\perp, \quad H_S := \mathrm{span}\{e^{\mathrm{i}jx} : j\in S\}, \quad H_S^\perp := \mathrm{span}\{e^{\mathrm{i}jx} : j\in S^c\}, \tag{3.2}$$

and we denote by $\Pi_S, \Pi_S^\perp$ the corresponding orthogonal projectors. The subspaces H_S and $H_S^\perp$ are symplectic orthogonal respect to the 2-form Ω (see (2.1)). We write

$$u = v + z, \quad v := \Pi_S u := \sum_{j\in S} u_j e^{\mathrm{i}jx}, \quad z = \Pi_S^\perp u := \sum_{j\in S^c} u_j e^{\mathrm{i}jx}.$$

For a finite dimensional space

$$E := E_C := \mathrm{span}\left\{e^{\mathrm{i}jx} : 0 < |j| \le C\right\}, \quad C > 0, \tag{3.3}$$

let Π_E denote the corresponding L^2-projector on E. The notation $R(v^{k-q}z^q)$ indicates a homogeneous polynomial of degree k in (v, z) of the form

$$R(v^{k-q}z^q) = M[\underbrace{v, \dots, v}_{(k-q)-\text{times}}, \underbrace{z, \dots, z}_{q-\text{times}}], \quad M = k\text{-linear}.$$

We denote with $H^{(n,\geq k)}$, $H^{(n,k)}$, $H^{(n,\leq k)}$ the terms of type $R(v^{n-s}z^s)$, where, respectively, $s \geq k$, $s = k$, $s \leq k$, that appear in the homogeneous polynomial H_n of degree n in the variables (v, z). Given an n-uple $\{j_1, \dots, j_n\} \subset \mathbb{Z}\backslash\{0\}$ and a set $B \subset \mathbb{Z}\backslash\{0\}$ we define

$$\sharp(\{j_1, \dots, j_n\}, B) := \text{number of } j_i \text{ belonging to } B.$$

Now we start the "weak" Birkhoff normal form procedure, i.e. we look for a change of coordinates which normalizes the terms in (3.1) independent and linear in the normal variable z.

As it is well known, one of the main problem of the Birkhoff normal form procedures is to deal with the resonances given by the equations (1.15) = 0 which arise from considering the kernel of the adjoint action $\mathrm{ad}_{H^{(2)}}$ (see (2.4)). It turns out that when $n \geq 2$ there are many non-trivial solutions of (1.15) = 0. A way to deal with this problem is to exploit the integrability of the DP equation.

In [32] the authors construct an infinite number of conserved quantities K_n for the Eq. (1.1) with $f = 0$ starting from the ones given in [28]. By an explicit characterization of the quadratic part of each K_n, they deduce that, at a purely formal level, the Birkhoff normal form of the Degasperis–Procesi equation is action preserving (or integrable). Here we rename these constants of motion in the following way, writing only the quadratic parts (which are fundamental for the study of the Birkhoff resonances at $u = 0$)

$$\begin{aligned} K_0(u) &:= H(u), \qquad K_1(u) := \frac{1}{2}\int_{\mathbb{T}} (J^{-1}u_x)\, u\, dx, \\ K_{n+2} &:= \int_{\mathbb{T}} (\partial_x^n w)^2\, dx + O(u^3), \quad n \geq 0, \end{aligned} \tag{3.4}$$

where we denoted by

$$w := \Lambda^{-1}u := u - u_{xx}, \qquad \Lambda := (1 - \partial_{xx})^{-1}. \tag{3.5}$$

We remark that K_1 is the *momentum* Hamiltonian arising from the translation invariance of the equation.

Definition 3.1. Given a quadratic diagonal Hamiltonian $Q(u) = \sum_j \mathfrak{l}(j)|u_j|^2$, we define $\Pi_{\mathrm{Ker}(Q)}$ as the projection on the kernel of the adjoint action (recall (2.2) and $J = \mathrm{diag}_{j\in\mathbb{Z}}(\lambda(j))$)

$$\begin{aligned} \mathrm{ad}_Q(K) &= \sum_{j_1,\dots,j_n} \Big(\sum_{i=1}^n \mathfrak{l}(j_i)\lambda(j_i)\Big) K_{j_1\dots j_n} u_{j_1} \dots u_{j_n}, \\ K(u) &:= \sum_{j_1,\dots,j_n} K_{j_1\dots j_n} u_{j_1} \dots u_{j_n}. \end{aligned} \tag{3.6}$$

We define the projector on the range of the adjoint action as $\Pi_{\mathrm{Rg}(Q)} := \mathrm{I} - \Pi_{\mathrm{Ker}(Q)}$.

We say that K, as in (3.6), "preserves" momentum if and only if

$$\Big(\sum_{i=1}^{n} j_i\Big) K_{j_1 \dots j_n} = 0 \qquad \forall j_1, \dots, j_n \in \mathbb{Z}\backslash\{0\}.$$

The main result of this section is the following.

Proposition 3.2. *There exist $r > 0$, depending on S (see (1.9)), and an analytic symplectic change of coordinates*

$$\Phi_B \colon \mathcal{B}_r(0, H_0^1(\mathbb{T})) \to H_0^1(\mathbb{T}), \qquad \Phi_B = \mathrm{I} + \Psi, \quad \Psi = \Pi_E \circ \Psi \circ \Pi_E \tag{3.7}$$

where E is a finite dimensional space as in (3.3), such that the Hamiltonian H in (3.1) transforms into

$$\mathcal{H} := H \circ \Phi_B = H^{(2)} + \mathcal{H}^{(4,0)} + \mathcal{H}^{(6,0)} + \mathcal{H}^{(8,0)} + \mathcal{H}^{(\geq 9, \leq 1)} + \mathcal{H}^{(\geq 3, \geq 2)}, \tag{3.8}$$

where

$$\begin{aligned}
\mathcal{H}^{(3,\geq 2)} &:= -\frac{1}{2}\int_{\mathbb{T}} v\, z^2\, dx - \frac{1}{6}\int_{\mathbb{T}} z^3\, dx,\\
\mathcal{H}^{(4,0)} &:= \frac{1}{2}\sum_{j\in S^+} \frac{\lambda(2j)}{2\lambda(j) - \lambda(2j)} |u_j|^4 + \sum_{\substack{j_1, j_2 \in S^+,\\ j_1 - j_2 \neq 0}} \frac{\lambda(j_1 + j_2)}{\lambda(j_1) + \lambda(j_2) - \lambda(j_1 + j_2)} |u_{j_1}|^2 |u_{j_2}|^2\\
&\quad + \sum_{\substack{j_1, j_2 \in S^+,\\ j_1 - j_2 \neq 0}} \frac{\lambda(j_1 - j_2)}{\lambda(j_1) - \lambda(j_2) - \lambda(j_1 - j_2)} |u_{j_1}|^2 |u_{j_2}|^2
\end{aligned} \tag{3.9}$$

and $\mathcal{H}^{(k,0)} = \Pi_{\mathrm{Ker}(H^{(2)})}\mathcal{H}^{(k,0)}$ with $k = 4, 6, 8$ depend only on $|u_j|^2$. The same change of variables Φ_B puts all the Hamiltonians in (3.4) in weak Birkhoff normal form up to order eight as in (3.8). In particular we have $K_1 \circ \Phi_B = K_1$.

In order to prove the Proposition 3.2 above we need some preliminary results proved in detail in [32].

Definition 3.3 (*M-resonances*). Fix $M \in \mathbb{N}$, $M \geq 3$. We recall that the quadratic part of H and K_r, $2 \leq r \leq M$, in (3.4) are

$$K_r^{(2)}(u) := \sum_{j\in\mathbb{Z}\backslash\{0\}} (1 + j^2)^2\, j^{2(r-2)}\, |u_j|^2, \quad H^{(2)}(u) = \sum_{j\geq 1} |u_j|^2.$$

We say that an n-uple $\{j_1, \dots, j_n\} \subset \mathbb{Z}\backslash\{0\}$, with $n \leq M$, is a *M-resonance* of order n for the DP hierarchy if

$$\sum_{i=1}^{n} j_i = 0, \quad \sum_{i=1}^{n} \lambda(j_i) = 0, \quad \sum_{i=1}^{n} (1 + j_i^2)^2\, j_i^{2(r-2)} \lambda(j_i) = 0 \quad \forall r = 2, \dots, M + 1. \tag{3.10}$$

Proposition 3.4. *Fix $M \in \mathbb{N}$, $M \geq 3$. All the M-resonances of the DP equations (3.4) are trivial, namely there are no resonances of odd order and the even ones are, up to permutations, of the form*

$$(i, -i, j, -j, k, -k, p, -p, \dots). \tag{3.11}$$

Proof. Since this Proposition is proved in [32] with different notations, for completeness we restate here a concise proof by induction on M. For $M = 3$ the thesis follows trivially: indeed direct computations show that

$$\sum_{i=1}^{3} j_i = 0, \quad \sum_{i=1}^{3} \lambda(j_i) = 0 \;\Leftrightarrow\; j_1 = -j_2,\ j_3 = 0$$

up to permutations, and this solution is incompatible with $j_i \in \mathbb{Z}\backslash\{0\}$.

Let us now suppose that the thesis is true up to $M - 1 \geq 3$ and prove it for M. We start by noticing that if $n < M$ then (3.10) with $r \leq M - 1$ can hold only if $\{j_1, \dots, j_n\}$ is a $M - 1$ resonance of order n. The inductive hypothesis then says that $\{j_1, \dots, j_n\}$ is trivial. Similarly if $\{j_1, \dots, j_n\}$ contains a trivial resonance, i.e. if $j_{i_1} + j_{i_2} = 0$ for $1 \leq i_1, i_2 \leq n$, then j_{i_1}, j_{i_2} do not appear in (3.10) and hence $\{j_1, \dots, j_n\}$ is an M-resonance of order n if and only if

$$\{j_1, \dots, j_n\}\backslash\{j_{i_1}, j_{i_2}\}, \quad \text{is an } M-2 \text{ resonance of order } n-2.$$

Without loss of generality we assume that $n = M$ and that $j_{i_1} + j_{i_2} \neq 0$ for any $1 \leq i_1, i_2 \leq M$.

Up to a permutation we can assume that for some $M \geq k \geq 1$ and $\alpha_1, \dots, \alpha_k \geq 1$ one has

$$\{j_1, \dots, j_n\} = \{\underbrace{\widehat{j_1}, \dots, \widehat{j_1}}_{\alpha_1}, \dots, \underbrace{\widehat{j_k}, \dots, \widehat{j_k}}_{\alpha_k}\}.$$

Consequently rewrite the third equation in (3.10) as $\sum_{i=1}^{k} \alpha_i (1+\widehat{j_i}^2)^2\, \widehat{j_i}^{2(r-2)} \lambda(\widehat{j_i}) = 0$, $\forall r = 2, \dots, M+1$. Then we can extract k equations from these ones and write them in the form

$$\begin{pmatrix} 1 & \dots & 1 \\ \widehat{j_1}^2 & \dots & \widehat{j_k}^2 \\ \vdots & & \vdots \\ \widehat{j_1}^{2(k-1)} & \dots & \widehat{j_k}^{2(k-1)} \end{pmatrix} \begin{pmatrix} \alpha_1 (1+\widehat{j_1}^2)^2\, \lambda(\widehat{j_1}) \\ \vdots \\ \alpha_k (1+\widehat{j_k}^2)^2\, \lambda(\widehat{j_k}) \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}. \tag{3.12}$$

The determinant of the Vandermonde matrix in (3.12) is $\prod_{i \neq h} (\widehat{j_i}^2 - \widehat{j_h}^2) \neq 0$, since, by hypothesis, $\widehat{j_i} \neq \pm \widehat{j_h}$. Then the only possible solution corresponds to $\widehat{j_i} = 0$ for all i, which is not compatible with $\widehat{j_i} \in \mathbb{Z}\backslash\{0\}$. □

Remark 3.5. Notice that if $j_1, \dots, j_N \in \mathbb{Z}\backslash\{0\}$, $j_1 + \dots + j_N = 0$ and $\#(\{j_1, \dots, j_N\}, S^c) \leq 1$, then $\max_{i=1,\dots,N} |j_i| \leq (N-1)\overline{j}_1$. Thus, the vector field $X_{F^{(N,\leq 1)}}$, generated by the finitely supported Hamiltonian

$$F^{(N,\leq 1)} = \sum_{\substack{j_1 + \dots + j_N = 0 \\ \#(\{j_1, \dots, j_N\}, S^c) \leq 1}} F^{(N,\leq 1)}_{j_1 \dots j_N} u_{j_1} \dots u_{j_N} \tag{3.13}$$

is finite rank, and, in particular, it vanishes outside the finite dimensional subspace $E := E_{(N-1)\overline{j}_1}$ (see (3.3)) and it has the form

$$X_{F^{(N,\leq 1)}}(u) = \Pi_E X_{F^{(N,\leq 1)}}(\Pi_E u).$$

Therefore its flow $\Phi^{(N)}$ is analytic and invertible on the phase space $H^1_0(\mathbb{T})$, provided that $|\Pi_E u|$ is appropriately small.

In order to prove Proposition 3.2 we need the following result.

Proposition 3.6. *Fix $M \in \mathbb{N}$, $M \geq 2$ and consider H in (1.4) and K_m, $m = 1, \dots, M$, in (3.4). Then, for any $N \leq M-2$, there exists $r > 0$ and an analytic symplectic change of coordinates $\Phi_N^{\pm 1} : \mathcal{B}_r(0, H_0^1(\mathbb{T})) \to H_0^1(\mathbb{T})$ of the form*

$$\Phi_N = \mathrm{I} + \Psi_N, \quad \Phi_0 = \mathrm{I}, \quad \Psi_N(u) = \Pi_E \circ \Psi_N \circ \Pi_E, \tag{3.14}$$

where E is a finite dimensional space as in (3.3), such that

$$\begin{aligned} H \circ \Phi_N^{-1} &= H^{(2)} + Z_N^{(\leq N+2,0)} + R_N^{(\geq N+3,\leq 1)} + H_N^{(\geq 3,\geq 2)}, \qquad K_1 \circ \Phi_N^{-1} = K_1, \\ K_m \circ \Phi_N^{-1} &= K_m^{(2)} + W_{m,N}^{(\leq N+2,0)} + Q_{m,N}^{(\geq N+3,\leq 1)} + K_{m,N}^{(\geq 3,\geq 2)}, \quad m = 2, \dots, M, \end{aligned} \tag{3.15}$$

where $Z_N^{(\leq N+2,0)}, W_{m,N}^{(\leq N+2,0)} \in \bigcap_{m=1}^M Ker(K_m^{(2)}) \cap Ker(H^{(2)})$.

Proof. The terms of degree at most 2 in the variable z are not affected by the procedure that we are going to describe.

We argue the result by induction on the number of steps N. For $N = 0$ it is trivial since Φ_0 is the identity map.

Suppose that we have performed N steps. By the fact that $\{H, K_m\} = 0$ then $\{H, K_m\} \circ \Phi_N^{-1} = 0$. For the latter, we are interested in the corresponding equations for the terms of homogeneity at most $N+3$ and degree in the variable z less or equal than one. So we consider the projection $\Pi^{(\leq N+3,\leq 1)}\Big(\{H, K_m\} \circ \Phi_N^{-1}\Big) = 0$ and we get, for any $m = 1, \dots, M$, the following system of equations $\{H^{(2)}, K_m^{(2)}\} = 0$ and

$$\begin{aligned} &\{H^{(2)}, W_{N,m}^{(N+2,0)}\} + \{Z_N^{(N+2,0)}, K_m^{(2)}\} + \Pi^{(\leq N+2)}\{Z_N^{(N+2,0)}, W_{N,m}^{(N+2,0)}\} = 0, \\ &\Pi^{(N+3)}\{Z_N^{(N+2,0)}, W_{N,m}^{(N+2,0)}\} + \{H^{(2)}, Q_{m,N}^{(N+3,\leq 1)}\} + \{R_N^{(N+3,\leq 1)}, K_m^{(2)}\} = 0. \end{aligned}$$

By the inductive hypothesis $W_{N,m}^{(N+2,0)}, Z_N^{(N+2,0)} \in \bigcap_{m=1}^M \mathrm{Ker}(K_m^{(2)}) \cap \mathrm{Ker}(H^{(2)})$, hence

$$\{H^{(2)}, W_{N,m}^{(N+2,0)}\} = \{Z_N^{(N+2,0)}, K_m^{(2)}\} = 0$$

and

$$\{H^{(2)}, Q_{m,N}^{(N+3,\leq 1)}\} + \{R_N^{(N+3,\leq 1)}, K_m^{(2)}\} = 0, \quad m = 1, \dots, M, \tag{3.16}$$

since $\{H^{(2)}, Q_{m,N}^{(N+3,\leq 1)}\} \in \mathrm{Rg}(H^{(2)})$ and $\{R_N^{(N+3,\leq 1)}, K_m^{(2)}\} \in \mathrm{Rg}(K_m^{(2)})$.
We note the following fact, which derives from the Jacobi identity: if $f \in \mathrm{Ker}(H^{(2)})$ then $\{f, K_m^{(2)}\} \in \mathrm{Ker}(H^{(2)})$.
Then we have that $\{\Pi_{\mathrm{Ker}(H^{(2)})} R_N^{(N+3,\leq 1)}, K_m^{(2)}\} \in \mathrm{Ker}(H^{(2)})$ and by (3.16)

$$\begin{aligned} \{\Pi_{\mathrm{Ker}(H^{(2)})} R_N^{(N+3,\leq 1)}, K_m^{(2)}\} &= -\{\Pi_{\mathrm{Rg}(H^{(2)})} R_N^{(N+3,\leq 1)}, K_m^{(2)}\} \\ &+\{H^{(2)}, Q_{N,m}^{(N+3,\leq 1)}\} \in \mathrm{Rg}(H^{(2)}). \end{aligned}$$

Thus $\{\Pi_{\mathrm{Ker}(H^{(2)})} R_N^{(N+3,\leq 1)}, K_m^{(2)}\} = 0$ and

$$\Pi_{\mathrm{Ker}(H^{(2)})} R_N^{(N+3,\leq 1)} = \Pi_{\mathrm{Ker}(H^{(2)})} \Pi_{\mathrm{Ker}(K_m^{(2)})} R_N^{(N+3,\leq 1)}.$$

By symmetry $\Pi_{\mathrm{Ker}(K_m^{(2)})} Q_{m,N}^{(N+3,\leq 1)} = \Pi_{\mathrm{Ker}(H^{(2)})}\Pi_{\mathrm{Ker}(K_m^{(2)})} Q_{m,N}^{(N+3,\leq 1)}$. Hence

$$\Pi_{\mathrm{Rg}(H^{(2)})}\Pi_{\mathrm{Ker}(K_m^{(2)})} Q_{m,N}^{(N+3,\leq 1)} = \Pi_{\mathrm{Rg}(K_m^{(2)})}\Pi_{\mathrm{Ker}(H^{(2)})} R_N^{(N+3,\leq 1)} = 0, \quad m = 1,\dots,M. \tag{3.17}$$

In order to obtain the Birkhoff normal form at order $N+3$ we consider a Birkhoff transformation $\Phi_{F^{(N+3,\leq 1)}}$ with generator $F^{(N+3),\leq 1}$ of the form (3.13) (with $N \rightsquigarrow N+3$) and we define $\Phi_{N+1} := \Phi_{F^{(N+3,\leq 1)}} \circ \Phi_N$. By Remark 3.5 the flow $\Phi_{F^{(N+3,\leq 1)}}$ is well defined in an appropriately small ball and it has the form Identity plus a finite rank operator. Note that, since $F^{(N+3,\leq 1)}$ is Fourier supported on $(j_1,\dots,j_{N+3})$ such that $j_1+\cdots+j_{N+3}=0$, the Hamiltonian K_1 commutes with $F^{(N+3,\leq 1)}$ and, by the inductive hypothesis, $K_1 \circ \Phi_{N+1}^{-1} = K_1$. The function $F^{(N+3,\leq 1)}$ is chosen in order to solve the homological equation

$$\{H^{(2)}, F^{(N+3,\leq 1)}\} = \Pi_{\mathrm{Rg}(H^{(2)})} R_N^{(N+3,\leq 1)} \overset{(3.17)}{=} \Pi_{\mathrm{Rg}(K_m^{(2)})}\Pi_{\mathrm{Rg}(H^{(2)})} R_N^{(N+3,\leq 1)}.$$

We now show that $F^{(N+3,\leq 1)}$ solves also the homological equation for the commuting Hamiltonians $K_m \circ \Phi_N^{-1}$, $m=1,\dots,M$. Indeed, by the fact that $\mathrm{ad}_{H^{(2)}}^{-1}$ commutes with $\mathrm{ad}_{K_m^{(2)}}$ on the intersection $\mathrm{Rg}(H^{(2)}) \cap \mathrm{Rg}(K_m^{(2)})$, we have

$$\{K_m^{(2)}, F^{(N+3,\leq 1)}\} = \mathrm{ad}_{H^{(2)}}^{-1}\{K_m^{(2)}, \Pi_{Rg(K_m^{(2)})}\Pi_{Rg(H^{(2)})} R_N^{(N+3,\leq 1)}\},$$

and by (3.16), (3.17) we get

$$\{K_m^{(2)}, \Pi_{\mathrm{Rg}(K_m^{(2)})}\Pi_{\mathrm{Rg}(H^{(2)})} R_N^{(N+3,\leq 1)}\} = \{H^{(2)}, \Pi_{\mathrm{Rg}(K_m^{(2)})}\Pi_{\mathrm{Rg}(H^{(2)})} Q_{m,N}^{(N+3,\leq 1)}\}.$$

By (3.17) we have that the resonant term $Z_{N+1}^{(N+3,\leq 1)} := \Pi_{\mathrm{Ker}(H^{(2)})} R_N^{(N+3,\leq 1)}$ belongs to the intersection of the kernels and by Proposition 3.4 these terms are supported only on n-ples of indices of the form $(i,-i,j,-j,k,-k,\dots)$. By the symmetry of the tangential set S this is possible for a set of indices with at most one outside S if and only if all the indices belong to S. Hence $Z_{N+1}^{(N+3,1)} = 0$ and we define $Z_{N+1}^{(\leq N+3,0)} := Z_{N+1}^{(N+3,0)} + Z_N^{(\leq N+2,0)}$. We do not compute explicitly the radius r of the ball in which we can perform the Birkhoff change of variables, however one can easily check that $r \to 0$ as $N \to \infty$ or as $\varkappa \to 0$ in Definition 1.9. □

Proof of Proposition 3.2. We apply Proposition 3.6 with $N=6$ and $M=8$ and we obtain (3.7), (3.8) by setting $\Phi_B := \Phi_N^{-1}$. To prove (3.9) we have to show explicitly the computations of the first step of Birkhoff normal form.
First we remove the cubic terms independent of z and linear in z from the Hamiltonian

$$H^{(3)} = -\frac{1}{6}\int_{\mathbb{T}} u^3\,dx = -\frac{1}{6}\int_{\mathbb{T}} v^3\,dx - \frac{1}{2}\int_{\mathbb{T}} v^2 z\,dx - \frac{1}{2}\int_{\mathbb{T}} v z^2\,dx - \frac{1}{6}\int_{\mathbb{T}} z^3\,dx. \tag{3.18}$$

We consider $\Phi_1 := (\Phi^t_{F^{(3,\leq 1)}})_{|t=1}$ as the time-1 flow map generated by the Hamiltonian vector field $X_{F^{(3,\leq 1)}}$, with an auxiliary Hamiltonian $F^{(3,\leq 1)}$ of the form (3.13) with $N=3$. The transformed Hamiltonian is $H_1 := H \circ \Phi_1^{-1} = H^{(2)} + H_1^{(3)} + H_1^{(4)} + H_1^{(\geq 5)}$ with

$$H_1^{(3)} := \{F^{(3,\leq 1)}, H^{(2)}\} + H^{(3)}, \quad H_1^{(4)} := \frac{1}{2}\{F^{(3,\leq 1)}, \{F^{(3,\leq 1)}, H^{(2)}\}\} + \{F^{(3,\leq 1)}, H^{(3)}\}, \tag{3.19}$$

and where $H_1^{(\geq 5)}$ collects all the terms of order at least five in (v, z). We choose $F^{(3,\leq 1)}$ such that the following homological equation holds

$$\{F^{(3,\leq 1)}, H^{(2)}\} + H^{(3)} = H^{(3,\geq 2)} \quad \Leftrightarrow \quad \{H^{(2)}, F^{(3,\leq 1)}\} = \Pi_{\mathrm{Rg}(H^{(2)})} H^{(3,\leq 1)}. \tag{3.20}$$

Recalling (2.2) and (3.18), the solution of the Eq. (3.20) is given by $F^{(3,\leq 1)}$ as in (3.13) with $N = 3$ with coefficients defined as

$$F^{(3,\leq 1)}_{j_1 j_2 j_3} := \begin{cases} \dfrac{1}{6\,\mathrm{i}\,(\lambda(j_1) + \lambda(j_2) + \lambda(j_3))} & \text{if } \sharp(\{j_1, j_2, j_3\}, S^c) \leq 1,\ j_1 + j_2 + j_3 = 0, \\ 0 & \text{otherwise.} \end{cases} \tag{3.21}$$

The Hamiltonian $F^{(3,\leq 1)}$ is well defined since, by Proposition 3.4, there are no non-trivial 3-resonances of order 3. Since $\Pi_{\mathrm{Rg}(H^{(2)})} H^{(3,\leq 1)} = H^{(3,\leq 1)}$ we get (see (3.19), (3.20))

$$H_1^{(3)} = H^{(3,\geq 2)}, \qquad H_1^{(4)} = \frac{1}{2}\{F^{(3,\leq 1)}, H^{(3,\leq 1)}\} + \{F^{(3,\leq 1)}, H^{(3,\geq 2)}\}. \tag{3.22}$$

In the second step we normalize the terms of total degree 4 and ≤ 1 in the variable z. The term $\Pi_{\mathrm{Ker}(H^{(2)})} H_1^{(4,\leq 1)}$ is Fourier supported on the set of 4-resonances of order 4, which are trivial by Proposition 3.4. By Proposition 3.6 $\Pi_{\mathrm{Ker}(H^{(2)})} H_1^{(4,1)} = 0$. Thus we have to compute only $\Pi_{\mathrm{Ker}(H^{(2)})} H_1^{(4,0)}$. We have

$$Z_2^{(4,0)} = \Pi_{\mathrm{Ker}(H^{(2)})} H_1^{(4,0)} = \frac{1}{8} \sum_{\substack{j_1, j_2, j_3, j_4 \in S, \\ j_1 + j_2 + j_3 + j_4 = 0 \\ j_1 + j_2 \neq 0,\ j_3 + j_4 \neq 0, \\ \sum_{k=1}^4 \lambda(j_k) = 0}} \frac{\lambda(j_1 + j_2)}{\lambda(j_1) + \lambda(j_2) - \lambda(j_1 + j_2)}\, u_{j_1} u_{j_2} u_{j_3} u_{j_4}. \tag{3.23}$$

The remaining steps of this procedure do not affect the terms with degree of homogeneity less or equal than 4. Hence by (3.23), the fact that $\lambda(-j) = -\lambda(j)$ (see (1.8)) and the symmetry of S we obtain (3.9). □

4. Action-Angle Variables

On the submanifold $\{z = 0\}$ we put the following action-angle variables

$$\mathbb{T}^\nu \times [0, \infty)^\nu \longrightarrow \{z = 0\}, \qquad (\theta, I) \longmapsto v = \sum_{j \in S} \sqrt{I_j}\, e^{\mathrm{i}\theta_j}\, e^{\mathrm{i}jx}. \tag{4.1}$$

Note that this change of coordinates is real if and only if $I_{-j} = I_j$ and $\theta_{-j} = -\theta_j$. The symplectic form in (2.1) restricted to the subspace H_S transforms into the 2-form $\sum_{j \in S^+} \frac{1}{\lambda(j)} d\theta_j \wedge dI_j$. We have that the Hamiltonian $\mathcal{H}^{(\leq 8)}(\theta, I, 0) = \sum_{j \in S^+} I_j + \mathcal{H}^{(4,0)}(I) + \mathcal{H}^{(6,0)}(I) + \mathcal{H}^{(8,0)}(I)$ depends only by the actions I and its equations of motion read as

$$\begin{cases} \dot\theta_j = \lambda(j)\ \partial_{I_j} \mathcal{H}^{(\leq 8)}(\theta, I, 0), & j \in S^+, \\ \dot I_j = -\lambda(j) \partial_{\theta_j} \mathcal{H}^{(\leq 8)}(\theta, I, 0) = 0, & j \in S^+, \end{cases} \tag{4.2}$$

where, by (3.9),

$$\partial_{I_j}\mathcal{H}^{(\leq 8)}(\theta, I, 0) = 1 + \frac{\lambda(2j)}{4\lambda(j) - 2\lambda(2j)}\, I_j + \sum_{k\in S^+, k\neq j} \mathfrak{b}_{jk}\, I_k + O(I^2)\,, \quad j \in S^+\,, \tag{4.3}$$

$$\mathfrak{b}_{jk} := \frac{2}{3}\frac{(1+k^2)(1+j^2)(2+k^2+j^2)}{(3+k^2+j^2+kj)(3+k^2+j^2-kj)}. \tag{4.4}$$

In order to highlight the fact that we are working close to zero, we introduce a small parameter $\varepsilon > 0$ and we rescale $I \mapsto \varepsilon^2 I$, so that the *frequency–amplitude map* can be written as

$$\omega(I) \;\rightsquigarrow\; \alpha(I) = \overline{\omega} + \varepsilon^2 \mathbb{A}\, I + O(\varepsilon^4), \tag{4.5}$$

where $\overline{\omega}$ is the vector of the linear frequencies (see (1.10)),

$$\mathbb{A} := \frac{1}{2}\mathbb{D}\ \mathrm{diag}\left(\frac{\lambda(2j)}{2\lambda(j)-\lambda(2j)}\right)_{j\in S^+} + \mathbb{D}\ \mathbb{B}, \quad \mathbb{D} := \mathrm{diag}\big(\lambda(j)\big)_{j\in S^+},$$
$$\mathbb{B}^k_j := \begin{cases} \mathfrak{b}_{jk} & \text{if } \ j \neq k, \\ 0 & \text{if } \ j = k. \end{cases} \tag{4.6}$$

The submanifold $\{z = 0\}$ is foliated by tori, parameterized by the actions, supporting small amplitude quasi-periodic solutions for the truncated system with Hamiltonian $\mathcal{H}^{(\leq 8)}$. We shall select some of them as starting point of the Nash–Moser scheme, by fixing $I = \xi$ (here ξ is a parameter), so that appropriate non-resonance conditions on the frequency $\alpha(I)$ hold.

In order to work in a small neighbourhood of the prefixed torus $\{I \equiv \xi\}$ it is advantageous to introduce a set of coordinates $(\theta, y, z) \in \mathbb{T}^\nu \times \mathbb{R}^\nu \times H^\perp_S$ adapted to it, defined by

$$\begin{aligned} u = A_\varepsilon(\theta, y, z) &= \varepsilon v_\varepsilon(\theta, y) + \varepsilon^b z \\ \iff \quad & \begin{cases} u_j := \varepsilon\sqrt{\xi_j + \varepsilon^{2b-2}|\lambda(j)|y_j}\, e^{\mathrm{i}\theta_j}\, e^{\mathrm{i}\, j\, x}, & j \in S, \\ u_j := \varepsilon^b z_j, & j \in S^c, \end{cases} \end{aligned} \tag{4.7}$$

with $b > 1$ and where (recall $\overline{u}_j = u_{-j}$)

$$\xi_{-j} = \xi_j, \quad \xi_j > 0, \quad y_{-j} = y_j, \quad \theta_{-j} = -\theta_j, \quad \theta_j \in \mathbb{T},\ y_j \in \mathbb{R}, \quad \forall j \in S.$$

The parameter b will be chosen close to one, to this purpose we shall set

$$a := 2b - 2, \tag{4.8}$$

and fix $a > 0$ appropriately small. For the tangential sites $S^+ := \{\overline{\jmath}_1, \dots, \overline{\jmath}_\nu\}$ we will also denote $\theta_{\overline{\jmath}_i} := \theta_i$, $y_{\overline{\jmath}_i} := y_i$, $\xi_{\overline{\jmath}_i} := \xi_i$, $i = 1, \dots, \nu$. The symplectic 2-form Ω in (2.1), up to rescaling of time, becomes

$$\mathcal{W} := \sum_{i=1}^{\nu} d\theta_i \wedge dy_i + \frac{1}{2}\sum_{j\in S^c}\frac{1}{\mathrm{i}\lambda(j)}\, dz_j \wedge dz_{-j} = \Big(\sum_{i=1}^{\nu} d\theta_i \wedge dy_i\Big) \oplus \Omega_{S^\perp}, \tag{4.9}$$

where $\Omega_{S^\perp}$ is the symplectic form Ω in (2.1) restricted to the subspace $H_S^\perp$ in (3.2). The Hamiltonian system generated by $\mathcal{H}$ in (3.8) becomes

$$H_\varepsilon := \varepsilon^{-2b}\,\mathcal{H}\circ A_\varepsilon. \tag{4.10}$$

In the following lemma we prove that, under an appropriate choice of the tangential set (1.9), the function (4.5) is a diffeomorphism for ε small enough and then the system (4.2) is integrable and non-isochronous.

Lemma 4.1 (Twist condition). *There exist* $\mathtt{r}_0, \mathtt{c}_* > 0$ *such that, for any choice of the tangential sites* $S^+ \in \mathcal{V}(\mathtt{r})$ *with* $0 < \mathtt{r} \le \mathtt{r}_0$ *(see Definition 1.1), one has* $|\det \mathbb{A}| \ge \mathtt{c}_*\, \overline{\jmath}_1^{3\nu}$.

Proof. The proof is postponed in "Appendix A". □

As a consequence of the non-degeneracy condition in Lemma 4.1 the map in (4.5) is invertible and we denote

$$\xi := \xi(\omega) := \alpha^{(-1)}(\omega) = \varepsilon^{-2}\mathbb{A}^{-1}(\omega - \overline{\omega}) + O(\varepsilon^2). \tag{4.11}$$

5. The Nonlinear Functional Setting

We write the Hamiltonian in (4.10) (possibly eliminating constant terms depending only on ξ which are irrelevant for the dynamics) as

$$\begin{aligned} &H_\varepsilon = \mathcal{N} + P,\\ &\mathcal{N}(\theta, y, z) = \omega\cdot y + \frac{1}{2}(N(\theta)z, z)_{L^2},\quad \frac{1}{2}(N(\theta)z, z)_{L^2} := \frac{1}{2}((\mathtt{D}_z\nabla_z H_\varepsilon)(\theta, 0, 0)[z], z)_{L^2}, \end{aligned} \tag{5.1}$$

where $\mathcal{N}$ describes the linear dynamics normal to the torus, and $P := H_\varepsilon - \mathcal{N}$ collects the nonlinear perturbative effects. Note that both N and P depend on ω through the map $\omega \mapsto \xi(\omega)$.

We consider H_ε as a (ω, ε)-parameter family of Hamiltonians and we note that, for $P = 0$, H_ε possess an invariant torus at the origin with frequency ω, which we want to continue to an invariant torus for the full system.
We will select the frequency parameters from the following set (recall (4.11))

$$\Omega_\varepsilon := \{\omega \in \mathbb{R}^\nu \,:\, \xi(\omega) \in [1, 2]^\nu\}. \tag{5.2}$$

Setting (see (4.7))

$$\gamma = \varepsilon^{2b},\quad \tau := 2\nu + 6, \tag{5.3}$$

we define the non-resonant sets

$$\mathcal{G}_0^{(0)} := \left\{\omega \in \Omega_\varepsilon : |\omega\cdot\ell| \ge \gamma\,\langle\ell\rangle^{-\tau},\ \forall \ell \in \mathbb{Z}^\nu\backslash\{0\}\right\}, \tag{5.4}$$

$$\begin{aligned} \mathcal{G}_0^{(1)} := \Big\{&\omega \in \Omega_\varepsilon : |\overline{\omega}\cdot\ell + \varepsilon^2\mathbb{A}\xi(\omega)\cdot\ell + \lambda(j') - \lambda(j) + \varepsilon^2(\lambda(j')\mathfrak{l}_{j'} - \lambda(j)\mathfrak{l}_j)| > C\gamma,\\ &\sum_{i=1}^{\nu}\overline{\jmath}_i\ell_i + j' - j = 0,\ \forall |\ell| \le 3,\ \ell \in \mathbb{Z}^\nu\backslash\{0\}\ j, j' \in S^c,\ \ (\ell, j, j') \ne (0, j, j)\Big\}, \end{aligned} \tag{5.5}$$

for some constant C depending on S, where $\mathbb{A}$ is defined in (4.6) and

$$\mathfrak{l}_j := \frac{2}{3} \sum_{j_2 \in S^+} \frac{(1+j_2^2)(1+j^2)(2+j_2^2+j^2)}{(3+j_2^2-j_2 j+j^2)(3+j_2^2+j_2 j+j^2)} \xi_{j_2}(\omega). \tag{5.6}$$

We require that

$$\omega \in \mathcal{G}_0 := \mathcal{G}_0^{(0)} \cap \mathcal{G}_0^{(1)}. \tag{5.7}$$

Lemma 5.1. *We have that* $|\Omega_\varepsilon \backslash \mathcal{G}_0| \leq C_* \varepsilon^{2(\nu-1)} \gamma$ *for some* $C_* = C_*(S) > 0$.

Proof. The proof is postponed in "Appendix A". □

Remark 5.2. The diophantine condition $\omega \in \mathcal{G}_0^{(0)}$ is typical of KAM scheme. The lower bound in $\mathcal{G}_0^{(1)}$ involves resonances of order five with two normal modes. As explained in the introduction, in order to impose such lower bounds we need to take into account also the corrections of order ε^2. The matrix $\mathbb{A}$ comes from the weak BNF of Sect. 3. The terms $\mathfrak{l}_j$ come from the *linear* BNF procedure of Sect. 7.2. In particular they are evaluated explicitly using the identification argument of Theorem 7.9.

Remark 5.3. Note that the definition of γ in (5.3) is slightly stronger than the minimal condition for which is possible to prove that $\mathcal{G}_0^{(0)}$ has large measure, namely $\gamma \leq c\,\varepsilon^2$, with $c > 0$ small enough. Our choice turns out to be useful for proving that the Cantor set of frequencies of the expected quasi-periodic solutions has asymptotically full measure (as $\varepsilon \to 0$).

We look for an embedded invariant torus

$$i : \mathbb{T}^\nu \to \mathbb{T}^\nu \times \mathbb{R}^\nu \times H_S^\perp, \quad \varphi \mapsto i(\varphi) := (\theta(\varphi), y(\varphi), z(\varphi)) \tag{5.8}$$

of the Hamiltonian vector field X_{H_ε} (see (5.1)) supporting quasi-periodic solutions with diophantine frequency $\omega \in \mathcal{G}_0$.

For technical reason, it is useful to consider the modified Hamiltonian

$$H_{\varepsilon,\zeta}(\theta, y, z) := H_\varepsilon(\theta, y, z) + \zeta \cdot \theta, \quad \zeta \in \mathbb{R}^\nu. \tag{5.9}$$

More precisely, we introduce ζ in order to control the average in the y-component in our Nash Moser scheme. The vector ζ has no dynamical consequences since an invariant torus for the Hamiltonian vector field $X_{H_{\varepsilon,\zeta}}$ is actually invariant for X_{H_ε} itself.

Thus, we look for zeros of the nonlinear operator $\mathcal{F}(i, \zeta) \equiv \mathcal{F}(i, \zeta, \omega, \varepsilon) := \omega \cdot \partial_\varphi i(\varphi) - X_{\mathcal{N}}(i(\varphi)) - X_P(i(\varphi)) + (0, \zeta, 0)$ defined as

$$\mathcal{F}(i, \zeta) = \begin{pmatrix} \omega \cdot \partial_\varphi \Theta(\varphi) - \partial_y P(i(\varphi)) \\ \omega \cdot \partial_\varphi y(\varphi) + \frac{1}{2} \partial_\theta (N(\theta(\varphi)) z(\varphi))_{L^2(\mathbb{T})} + \partial_\theta P(i(\varphi)) + \zeta \\ \omega \cdot \partial_\varphi z(\varphi) - J N(\theta(\varphi))\, z(\varphi) - J \nabla_z P(i(\varphi)) \end{pmatrix} \tag{5.10}$$

where $\Theta(\varphi) := \theta(\varphi) - \varphi$ is $(2\pi)^\nu$-periodic. We define the Sobolev norm of the periodic component of the embedded torus

$$\mathfrak{I}(\varphi) := i(\varphi) - (\varphi, 0, 0) := (\Theta(\varphi), y(\varphi), z(\varphi)) \quad \|\mathfrak{I}\|_s := \|\Theta\|_s + \|y\|_s + \|z\|_s, \tag{5.11}$$

where $z \in H^s_{S^\perp} := H^s \cap H^\perp_S$ (recall (3.2)) with norm defined in (2.7) and with abuse of notation, we are denoting by $\|\cdot\|_s$ the Sobolev norms of functions in $H^s(\mathbb{T}^\nu, \mathbb{R}^\nu)$. From now on we fix $s_0 := [\nu/2] + 4$.

Notice that in the coordinates (4.7), a quasi-periodic solution corresponds to an embedded invariant torus (5.8). Therefore we can reformulate the main Theorem 1 as follows.

Theorem 5.4. *There exists a small constant $\mathtt{r} > 0$ such that, for any $S^+ \in \mathcal{V}(\mathtt{r})$ (see (1.9) and Definition 1.1), there exists $\varepsilon_0 > 0$, small enough, such that the following holds. For all $\varepsilon \in (0, \varepsilon_0)$ there exist positive constants $C = C(\nu)$, $\mu = \mu(\nu)$ and a Cantor-like set $\mathcal{C}_\varepsilon \subseteq \Omega_\varepsilon$ (see (5.2)), with asymptotically full measure as $\varepsilon \to 0$, namely*

$$\lim_{\varepsilon \to 0} \frac{|\mathcal{C}_\varepsilon|}{|\Omega_\varepsilon|} = 1, \tag{5.12}$$

such that, for all $\omega \in \mathcal{C}_\varepsilon$, there exists a solution $i_\infty(\varphi) := i_\infty(\omega, \varepsilon)(\varphi)$ of the equation $\mathcal{F}(i_\infty, 0, \omega, \varepsilon) = 0$ (see (5.10)). Hence the embedded torus $\varphi \mapsto i_\infty(\varphi)$ is invariant for the Hamiltonian vector field X_{H_ε}, and it is filled by quasi-periodic solutions with frequency ω. The torus i_∞ satisfies

$$\|i_\infty(\varphi) - (\varphi, 0, 0)\|^{\gamma, \mathcal{C}_\varepsilon}_{s_0+\mu} \leq C\, \varepsilon^{9-2b}\, \gamma^{-1}.$$

Moreover the torus i_∞ is linearly stable.

We can deduce Theorem 1 from Theorem 5.4, indeed the quasi-periodic solution u in (1.11) is

$$u(t, x) = \Big(\Phi_B \circ A_\varepsilon\Big) i_\infty(\omega t)$$

for $\omega = \omega(\xi) \in \mathcal{C}_\varepsilon$, where $\omega(\xi)$ is the frequency amplitude map (4.5).

The rest of the paper is devoted to the proof of Theorem 5.4.

5.1. Tame estimates of the nonlinear vector field. We give tame estimates for the composition operator induced by the Hamiltonian vector fields $X_\mathcal{N}$ and X_P in (5.10). Since the functions $y \to \sqrt{\xi + \varepsilon^{2(b-1)} y}$, $\theta \to e^{\mathrm{i}\theta}$ are analytic for ε small enough and $|y| \leq C$, classical composition results (see for instance Lemma 6.2 in [3]) imply that, for all $\|\mathfrak{I}\|^{\gamma, \mathcal{O}}_{s_0} \leq 1$,

$$\|A_\varepsilon(\theta(\varphi), y(\varphi), z(\varphi))\|^{\gamma, \mathcal{O}}_s \lesssim_s \varepsilon(1 + \|\mathfrak{I}\|^{\gamma, \mathcal{O}}_s).$$

In the following lemma we collect tame estimates for the Hamiltonian vector fields $X_\mathcal{N}$, X_P, X_{H_ε}, see (5.1). These bounds rely on tame estimates for composition operators and their proof is completely analogous to the one in Sect. 5 of [4].

Lemma 5.5. *Let $\mathfrak{I}(\varphi)$ in (5.11) satisfy $\|\mathfrak{I}\|^{\gamma, \mathcal{O}}_{s_0+1} \lesssim \varepsilon^{9-2b}\gamma^{-1}$. Then we have*

$$\begin{aligned}
\|\partial_y P(i)\|^{\gamma, \mathcal{O}}_s &\lesssim_s \varepsilon^7 + \varepsilon^{2b} \|\mathfrak{I}\|^{\gamma, \mathcal{O}}_{s+1}, \\
\|\partial_\theta P(i)\|^{\gamma, \mathcal{O}}_s &\lesssim_s \varepsilon^{9-2b}(1 + \|\mathfrak{I}\|^{\gamma, \mathcal{O}}_{s+1}), \\
\|\nabla_z P(i)\|^{\gamma, \mathcal{O}}_s &\lesssim_s \varepsilon^{8-b} + \varepsilon^{9-b}\gamma^{-1} \|\mathfrak{I}\|^{\gamma, \mathcal{O}}_{s+1},
\end{aligned}$$

$$\begin{aligned}
\|X_P(i)\|_s^{\gamma,\mathcal{O}} &\lesssim_s \varepsilon^{9-2b} + \varepsilon^{2b}\|\mathfrak{I}\|_{s+1}^{\gamma,\mathcal{O}},\\
\|\partial_\theta \partial_y P(i)\|_s^{\gamma,\mathcal{O}} &\lesssim_s \varepsilon^{7} + \varepsilon^{8}\gamma^{-1}\|\mathfrak{I}\|_{s+1}^{\gamma,\mathcal{O}},\\
\|\partial_y \nabla_z P(i)\|_s^{\gamma,\mathcal{O}} &\lesssim_s \varepsilon^{6+b} + \varepsilon^{2b-1}\|\mathfrak{I}\|_{s+1}^{\gamma,\mathcal{O}},\\
\|\partial_{yy} P(i) - \frac{\varepsilon^{2b}}{2}\mathbb{A}\Omega\|_s^{\gamma,\mathcal{O}} &\lesssim_s \varepsilon^{5+2b} + \varepsilon^{6+2b}\gamma^{-1}\|\mathfrak{I}\|_{s+1}^{\gamma,\mathcal{O}},
\end{aligned} \tag{5.13}$$

and for all $\widehat{\imath} := (\widehat{\Theta}, \widehat{y}, \widehat{z})$,

$$\begin{aligned}
\|\partial_y \mathtt{D}_i X_P(i)[\widehat{\imath}]\|_s^{\gamma,\mathcal{O}} &\lesssim_s \varepsilon^{2b-1}(\|\widehat{\imath}\|_{s+1}^{\gamma,\mathcal{O}} + \|\mathfrak{I}\|_{s+1}^{\gamma,\mathcal{O}}\|\widehat{\imath}\|_{s_0+1}^{\gamma,\mathcal{O}}),\\
\|\mathtt{D}_i X_{H_\varepsilon}(i)[\widehat{\imath}] + (0,0,J\widehat{z})\|_s^{\gamma,\mathcal{O}} &\lesssim_s \varepsilon(\|\widehat{\imath}\|_{s+1}^{\gamma,\mathcal{O}} + \|\mathfrak{I}\|_{s+1}^{\gamma,\mathcal{O}}\|\widehat{\imath}\|_{s_0+1}^{\gamma,\mathcal{O}}),\\
\|\mathtt{D}_i^2 X_{H_\varepsilon}(i)[\widehat{\imath},\widehat{\imath}]\|_s^{\gamma,\mathcal{O}} &\lesssim_s \varepsilon(\|\widehat{\imath}\|_{s+1}^{\gamma,\mathcal{O}}\|\widehat{\imath}\|_{s_0+1}^{\gamma,\mathcal{O}} + \|\mathfrak{I}\|_{s+1}^{\gamma,\mathcal{O}}(\|\widehat{\imath}\|_{s_0+1}^{\gamma,\mathcal{O}})^2).
\end{aligned}$$

In the sequel we will use that, by the diophantine condition (5.7), the operator $(\omega\cdot\partial_\varphi)^{-1}$ is defined for all functions u with zero φ-average, and satisfies

$$\|(\omega\cdot\partial_\varphi)^{-1}u\|_s \lesssim_s \gamma^{-1}\|u\|_{s+\tau}, \quad \|(\omega\cdot\partial_\varphi)^{-1}u\|_s^{\gamma,\mathcal{O}} \lesssim_s \gamma^{-1}\|u\|_{s+2\tau+1}^{\gamma,\mathcal{O}}.$$

6. Approximate Inverse

We want to solve the nonlinear functional equation (see (5.10))

$$\mathcal{F}(i,\zeta) = 0 \tag{6.1}$$

by applying a Nash–Moser scheme. It is well known that the main issue in implementing this algorithm concerns the approximate inversion of the linearized operator of $\mathcal{F}$ at any approximate solution (i_n,ζ_n), namely $\mathtt{D}\mathcal{F}(i_n,\zeta_n)$. Note that $\mathtt{D}\mathcal{F}(i_n,\zeta_n)$ is independent of ζ_n. One of the main problems is that the (θ, y, z)-components of $\mathtt{D}\mathcal{F}(i_n,\zeta_n)$ are coupled and then the linear system

$$\mathtt{D}\mathcal{F}(i_n,\zeta_n)[\widehat{\imath},\widehat{\zeta}] = \omega\cdot\partial_\varphi\widehat{\imath} - \mathtt{D}_i X_{H_\varepsilon}(i_n)[\widehat{\imath}] - (0,\widehat{\zeta},0) = g = (g^{(\theta)}, g^{(y)}, g^{(z)}) \tag{6.2}$$

is quite involved. In order to approximately solve (6.2) we follow the scheme developed by Berti–Bolle in [9] which describe a way to approximately triangularize (6.2). This method has been applied in [4,39]. Since the strategy is identical to [39] we only summarize it and underline the differences which mainly come from the symplectic structure. For a fully detailed expository presentation see [40].

We now study the solvability of Eq. (6.2) at an approximate solution, which we denote by (i_0,ζ_0), $i_0(\varphi) = (\theta_0(\varphi), y_0(\varphi), z_0(\varphi))$ in order to keep the notations of [4], [39] . Assume the following hypothesis, which we shall verify at any step of the Nash–Moser iteration,

- **Assumption** The map $\omega\mapsto i_0(\omega)$ is a Lipschitz function defined on some subset $\mathcal{O}_0\subseteq\mathcal{G}_0\subseteq\Omega_\varepsilon$ (recall (5.7),(5.2)) and, for some $\mathfrak{p}_0 := \mathfrak{p}_0(\nu) > 0$,

$$\|\mathfrak{I}_0\|_{s_0+\mathfrak{p}_0}^{\gamma,\mathcal{O}_0} \le \varepsilon^{9-2b}\gamma^{-1}, \quad \|Z\|_{s_0+\mathfrak{p}_0}^{\gamma,\mathcal{O}_0} \le \varepsilon^{9-2b}, \quad \gamma = \varepsilon^{2b}, \tag{6.3}$$

  where $\mathfrak{I}_0(\varphi) := i_0(\varphi) - (\varphi,0,0)$ and Z is the error function

$$Z(\varphi) := (Z_1,Z_2,Z_3)(\varphi) := \mathcal{F}(i_0,\zeta_0)(\varphi) = \omega\cdot\partial_\varphi i_0(\varphi) - X_{H_{\varepsilon,\zeta_0}}(i_0(\varphi)). \tag{6.4}$$

By estimating the Sobolev norm of the function Z we can measure how the embedding i_0 is close to being invariant for $X_{H_{\varepsilon,\zeta_0}}$. If $Z = 0$ then i_0 is a solution. In general we say that i_0 is "approximately invariant" up to order $O(Z)$. We observe that by Lemma 6.1 in [4] we have that if i_0 is a solution, then the parameter ζ_0 has to be naught, hence the embedded torus i_0 supports a quasi-periodic solution of the "original" system with Hamiltonian H_ε (see (5.1)).

By [9] we know that it is possible to construct an embedded torus $i_\delta(\varphi) = (\theta_0(\varphi), y_\delta(\varphi), z_0(\varphi))$, which differs from i_0 only for a small modification of the y-component, such that the 2-form $\mathcal{W}$ (recall (4.9)) vanishes on the torus $i_\delta(\mathbb{T}^\nu)$, namely i_δ is isotropic. In particular $i_\delta(\varphi)$ is approximately invariant up to order $O(Z)$ (see Lemma 7 in [9]) and, more precisely, there exists $\tilde{\mathfrak{p}} := \tilde{\mathfrak{p}}(\nu) > 0$ such that

$$\|i_\delta - i_0\|_s^{\gamma,\mathcal{O}_0} \lesssim_s \|\mathfrak{I}_0\|_{s+\tilde{\mathfrak{p}}}^{\gamma,\mathcal{O}_0}. \tag{6.5}$$

The strategy is to construct an approximate inverse for $\mathtt{D}\mathcal{F}(i_0, \zeta_0)$ by starting from an approximate inverse for the linear operator $\mathtt{D}\mathcal{F}(i_\delta, \zeta_0)$. The advantage of analyzing the linearized problem at i_δ is that it is possible to construct a *symplectic* change of variable which approximately triangularizes the linear system thanks to the isotropicity of i_δ. For the details we refer to [9] and [4], here we only give the relevant definitions and state the main result. We define the symplectic change of coordinates

$$\begin{pmatrix} \theta \\ y \\ z \end{pmatrix} := G_\delta \begin{pmatrix} \varphi \\ \eta \\ w \end{pmatrix} := \begin{pmatrix} \theta_0(\varphi) \\ y_\delta(\varphi) + [\partial_\varphi \theta_0(\varphi)]^{-T} \eta + [(\partial_\theta \tilde{z}_0)(\theta_0(\varphi))]^T \, J^{-1} w \\ z_0(\varphi) + w \end{pmatrix} \tag{6.6}$$

where $\tilde{z}_0 := z_0(\theta_0^{-1}(\theta))$. We denote the transformed Hamiltonian by $K := K(\varphi, \eta, w, \zeta_0)$. We then define

$$\mathcal{L}_\omega := \omega \cdot \partial_\varphi - J K_{02}(\varphi), \tag{6.7}$$

where K_{02} is the linear operator representing the terms quadratic in w of K, i.e.

$$\frac{1}{2}(K_{02}(\varphi)[w], w) := \Pi^{d_w=2} K = \Pi^{d_w=2} H_\varepsilon \circ G_\delta. \tag{6.8}$$

$\mathcal{L}_\omega$ corresponds to the w-component of the linearized operator after the change of variable G_δ.

In [9] (see also [4,39]) the following result is proved.

Theorem 6.1. *Assume* (6.3) *and the following*
Inversion Assumption *There exist* $\mathfrak{p}_1 := \mathfrak{p}_1(\nu) > 0$ *and a set* $\Omega_\infty \subset \mathcal{G}_0 \subseteq \Omega_\varepsilon$ *such that for all* $\omega \in \Omega_\infty$ *and every function* $g \in H^{s+2\tau+1} \cap H_S^\perp$, *there exists a solution* $h := \mathcal{L}_\omega^{-1} g$ *of the linear equation* $\mathcal{L}_\omega h = g$ *which satisfies*

$$\|\mathcal{L}_\omega^{-1} g\|_s^{\gamma,\Omega_\infty} \lesssim_s \gamma^{-1}(\|g\|_{s+2\tau+1}^{\gamma,\Omega_\infty} + \varepsilon\gamma^{-5/2}\|\mathfrak{I}_\delta\|_{s+\mathfrak{p}_1}^{\gamma,\mathcal{O}_0}\|g\|_{s_0}^{\gamma,\Omega_\infty}). \tag{6.9}$$

Then there exists $\mu := \mu(\nu)$ *such that, for all* $\omega \in \Omega_\infty$ *there exists a linear operator* $\mathbf{T}_0$ *such that:*

1. *For all* $g := (g^{(\theta)}, g^{(y)}, g^{(z)})$, *one has*

$$\|\mathbf{T}_0 g\|_s^{\gamma,\Omega_\infty} \lesssim_s \gamma^{-1}(\|g\|_{s+\mu}^{\gamma,\Omega_\infty} + \varepsilon\gamma^{-5/2}\|\mathfrak{I}_0\|_{s+\mu}^{\gamma,\mathcal{O}_0}\|g\|_{s_0+\mu}^{\gamma,\Omega_\infty}). \tag{6.10}$$

2. $\mathbf{T}_0$ *is an approximate inverse of* $\mathtt{D}\mathcal{F}(i_0)$*, namely*

$$\|(\mathtt{D}\mathcal{F}(i_0)\circ\mathbf{T}_0-\mathrm{I})g\|_s^{\gamma,\Omega_\infty}\lesssim_s \varepsilon^{2b-1}\gamma^{-2}\Big(\|\mathcal{F}(i_0,\zeta_0)\|_{s_0+\mu}^{\gamma,\mathcal{O}_0}\|g\|_{s+\mu}^{\gamma,\Omega_\infty}+\{\|\mathcal{F}(i_0,\zeta_0)\|_{s+\mu}^{\gamma,\mathcal{O}_0}+\varepsilon\gamma^{-5/2}\|\mathcal{F}(i_0,\zeta_0)\|_{s_0+\mu}^{\gamma,\mathcal{O}_0}\|\mathfrak{I}_0\|_{s+\mu}^{\gamma,\mathcal{O}_0}\}\|g\|_{s_0+\mu}^{\gamma,\Omega_\infty}\Big). \tag{6.11}$$

6.1. The linearized operator in the normal directions. Recalling the assumption (6.3), in the sequel we assume that $\mathfrak{I}_\delta := \mathfrak{I}_\delta(\varphi;\omega)=i_\delta(\varphi;\omega)-(\varphi,0,0)$ satisfies, for some $\mathfrak{p}_1>0$,

$$\|\mathfrak{I}_\delta\|_{s_0+\mathfrak{p}_1}^{\gamma,\mathcal{O}_0}\lesssim\varepsilon^{9-2b}\gamma^{-1}. \tag{6.12}$$

We note moreover that G_δ in (6.6) is the identity plus a translation plus a finite rank linear operator; moreover, assuming (6.12), one has that G_δ is $O(\varepsilon^{9-2b}\gamma^{-1})$-close to the identity in low norm. Returning to the initial variables we set (see (4.7),(6.6))

$$T_\delta := A_\varepsilon(G_\delta(\varphi,0,0))=\varepsilon v_\delta+\varepsilon^b z_0,\quad v_\delta=\sum_{j\in S}\sqrt{\xi_j+\varepsilon^{2b-2}|\lambda(j)|y_{\delta j}(\varphi)}e^{\mathrm{i}(jx+\theta_{0j}(\varphi))} \tag{6.13}$$

and we have, for some $\sigma:=\sigma(\nu)>0$,

$$\|\Phi_B(T_\delta)\|_s^{\gamma,\mathcal{O}_0}\lesssim_s\varepsilon\,(1+\|\mathfrak{I}_\delta\|_{s+\sigma}^{\gamma,\mathcal{O}_0}),\qquad \|\mathtt{D}_i\Phi_B(T_\delta)[\widehat{\imath}]\|_s\lesssim_s\varepsilon(\|i\|_{s+\sigma}+\|\mathfrak{I}_\delta\|_{s+\sigma}\|i\|_{s_0+\sigma}). \tag{6.14}$$

By following Sect. 7 in [4] (see Lemma 7.1), K_{02} in (6.8) has rather explicit estimates.

Proposition 6.2. *Assume* (6.12)*. Then there exists* $\sigma_0=\sigma_0(\nu)>0$ *such that the following holds. The Hamiltonian operator* $\mathcal{L}_\omega$ *in* (6.7) *has the form*

$$\mathcal{L}_\omega=\Pi_S^\perp\big(\omega\cdot\partial_\varphi-J\circ(1+a_0(\varphi,x))+\mathcal{Q}_0\big),\qquad a_0(\varphi,x):=-(\Phi_B(T_\delta)+\partial_u^2 f(\Phi_B(T_\delta))). \tag{6.15}$$

Recall that T_δ *is defined in* (6.13)*,* Φ_B *is the Birkhoff map given in Proposition* 3.2*,* f *is the Hamiltonian density in* (1.3)*. The operator* $\mathcal{Q}_0$ *is finite rank and has the form*

$$\mathcal{Q}_0(\varphi)w=\sum_{|j|\le C}\int_0^1(w,g_j(\tau,\varphi))_{L^2(\mathbb{T})}\,\chi_j(\tau,\varphi)\,d\tau. \tag{6.16}$$

In particular we divide $\mathcal{Q}_0=\sum_{i=1}^5\varepsilon^i\mathcal{R}_i+\mathcal{R}_{>5}$*, where the* $\mathcal{R}_i$*,* $\mathcal{R}_{>5}$ *are finite rank operators. Moreover we have*

$$\|a_0\|_s^{\gamma,\mathcal{O}_0}\lesssim_s\varepsilon(1+\|\mathfrak{I}_\delta\|_{s+\sigma_0}^{\gamma,\mathcal{O}_0}),\qquad \|\mathtt{D}_i a_0[\widehat{\imath}]\|_s\lesssim_s\varepsilon(\|\widehat{\imath}\|_{s+\sigma_0}+\|\mathfrak{I}_\delta\|_{s+\sigma_0}\|\widehat{\imath}\|_{s_0}). \tag{6.17}$$

The remainders $\mathcal{R}_i$ *do not depend on* $\mathfrak{I}_\delta$ *and satisfy*

$$\|g_j^{(i)}\|_s^{\gamma,\mathcal{O}_0}+\|\chi_j^{(i)}\|_s^{\gamma,\mathcal{O}_0}\lesssim_s 1, \tag{6.18}$$

while $\mathcal{R}_{>5}$ *satisfies*

$$\|g_j^{>5}\|_s^{\gamma,\mathcal{O}_0}\|\chi_j^{>5}\|_{s_0}^{\gamma,\mathcal{O}_0}+\|g_j^{>5}\|_{s_0}^{\gamma,\mathcal{O}_0}\|\chi_j^{>5}\|_s^{\gamma,\mathcal{O}_0}\lesssim_s\varepsilon^6+\varepsilon^2\|\mathfrak{I}_\delta\|_{s+\sigma}^{\gamma,\mathcal{O}_0}, \tag{6.19}$$

$$\|\mathtt{D}_i g_j^{>5}[\widehat{\imath}]\|_s\|\chi_j^{>5}\|_{s_0}+\|\mathtt{D}_i g_j^{>5}[\widehat{\imath}]\|_{s_0}\|\chi_j^{>5}\|_s+\|g_j^{>5}\|_{s_0}\|\mathtt{D}_i\chi_j^*\|_s+\|g_j^{>5}\|_s\|\mathtt{D}_i\chi_j^{>5}\|_{s_0}$$

$$\lesssim_s \varepsilon^2 \|\widehat{\imath}\|_{s+\sigma} + \varepsilon^b \|\mathfrak{I}_\delta\|_{s+\sigma} \|\widehat{\imath}\|_{s_0+\sigma}. \tag{6.20}$$

Finally, recalling the Definition 2.3, we have

$$\mathfrak{M}^{\gamma}_{\mathcal{Q}_0}(0,s) \lesssim_s \varepsilon^2(1 + \|\mathfrak{I}_\delta\|^{\gamma,\mathcal{O}_0}_{s+\sigma_0}), \tag{6.21}$$

$$\mathfrak{M}_{\mathtt{D}_i \mathcal{Q}_0[\widehat{\imath}]}(0,s) \lesssim_s \varepsilon^2 \|\widehat{\imath}\|_{s+\sigma_0} + \varepsilon^{2b-1} \|\mathfrak{I}_\delta\|_{s+\sigma_0} \|\widehat{\imath}\|_{s_0+\sigma_0}. \tag{6.22}$$

Proof. The expression (6.15) follows from the definition (6.8) by remarking that G_δ and the weak BNF transformation Φ_B is the identity plus a finite rank operator, while the action angle change of coordinates is a rescaling plus a finite rank operator (acting only on the v). Then, in applying the chain rule, we get

$$\begin{aligned} \mathtt{D}_w \nabla_w (H_\varepsilon \circ G_\delta) &\overset{(4.10)}{=} \varepsilon^{-2b} (\mathtt{D}_z \nabla_z (\mathcal{H} \circ A_\varepsilon)) \circ G_\delta + R_1 = (\mathtt{D}_z \nabla_z \mathcal{H}) \circ A_\varepsilon \circ G_\delta + R_1 \\ &\overset{(3.8)}{=} (\mathtt{D}_z \nabla_z (H \circ \Phi_B)) \circ A_\varepsilon \circ G_\delta + R_1 \\ &= (\mathtt{D}_z \nabla_z H) \circ \Phi_B \circ A_\varepsilon \circ G_\delta + R_1 + R_2 \end{aligned} \tag{6.23}$$

where the finite rank part contains all the terms where a derivative falls on the change of variables. Then (6.15) follows from the definition of H in (1.4). Regarding the estimates, (6.17) follows from (6.14); regarding the bounds (6.18), (6.19), we split the finite rank part $R_1 + R_2$ as follows. The operator R_1 contains all terms arising form derivatives of G_δ. By tame estimates on the map G_δ (see for instance Lemma 6.7 in [4]), it satisfies the bounds (6.19) and we put it in $\mathcal{R}_{>5}$. The finite rank term R_2 comes from the Birkhoff map. This is an analytic map so we consider the Taylor expansion

$$\Phi_B(u) = u + \sum_{i=2}^{5} \Psi_i(u) + \Psi_{\geq 6}(u), \tag{6.24}$$

where each $\Psi_i(u)$ is homogeneous of degree i in u, while $\Psi_{\geq 6} = O(u^6)$ and all map $H^1_0(\mathbb{T})$ in itself. We have to evaluate Φ_B and its derivatives (up to order two) at $u = T_\delta = \varepsilon v_\delta + \varepsilon^b z_0$. We denote by $\overline{v}$ the function

$$\overline{v}(\varphi, x) := \sum_{j \in S} \sqrt{\xi_j} e^{\mathrm{i}(jx + \mathtt{l}(j)\cdot\varphi)} = A_\varepsilon(\varphi, 0, 0) \tag{6.25}$$

where $\mathtt{l}(\overline{\jmath}_i)$ is the i-th vector of the canonical basis of $\mathbb{Z}^\nu$ and is such that $\mathtt{l}(-\overline{\jmath}_i) = -\mathtt{l}(\overline{\jmath}_i)$.
We observe that[2]

$$\|v_\delta - \overline{v}\|^{\gamma,\mathcal{O}_0}_s \lesssim \|\mathfrak{I}_\delta\|^{\gamma,\mathcal{O}_0}_s,$$

and hence we can expand

$$\Phi_B(T_\delta) = \varepsilon \overline{v} + \sum_{i=2}^{5} \varepsilon^i \Psi_i(\overline{v}) + \tilde{q} = \Phi_B^{\leq 5} + \tilde{q}, \tag{6.26}$$

where $\tilde{q}$ is a remainder which satisfies

$$\|\tilde{q}\|^{\gamma,\mathcal{O}_0}_s \lesssim_s \varepsilon^6 + \varepsilon \|\mathfrak{I}_\delta\|^{\gamma,\mathcal{O}_0}_s, \quad \|\mathtt{D}_i \tilde{q}[\widehat{\imath}]\|_s \lesssim_s \varepsilon(\|\widehat{\imath}\|_s + \|\mathfrak{I}_\delta\|_s \|\widehat{\imath}\|_{s_0}). \tag{6.27}$$

[2] The function $\varepsilon\overline{v}$ represents a torus supporting a quasi-periodic motion which is invariant for the system (5.1) with $P = 0$, namely it is the approximate solution from which we bifurcate.

Then in $\mathcal{R}_i$ we include all the terms homogeneous of degree i coming from derivatives of $\Phi_B - \mathrm{I}$, evaluated at $\tilde{q} = 0$; we put in $\mathcal{R}_{>5}$ all the rest. The (6.21), (6.22) follows by (6.18), (6.19) and (6.20). □

Remark 6.3. The motivation for separating the $\mathcal{R}_i$ and $\mathcal{R}_{>5}$ is the following. Consider the Hamiltonian H_ε as a function of ξ instead of ω. Then in all our expressions we can, and shall, evidence a *purely polinomial* term $\sum_{i=0}^5 \varepsilon^i f_i$ (where the f_i are ε independent) plus a *remainder*, which is not analytic in ε, of size $\varepsilon^6 + \varepsilon \|\mathfrak{I}_\delta\|_s^{\gamma,\mathcal{O}_0}$. By the assumption (6.3), this means that in *low norm* $s = s_0 + \mathfrak{p}_1$ all these remainders are negligible w.r.t. terms of order ε^5. This distinction is needed because, due to the resonant nature of the DP equation, we need to perform (see Sects. 7.1 and 7.2) five steps of the order reduction and of the linear BNF by hand, before entering in a perturbative regime.
In this framework $\mathcal{R}_{>5}$ is purely a remainder, while the $\mathcal{R}_i$ are homogeneous polynomial terms. One could apply the same division to the non finite rank terms, one would get

$$\begin{aligned}&\Pi_S^\perp\big(\omega(\xi)\cdot\partial_\varphi h - J\,[(1-\Phi_B(T_\delta) - \partial_u^2 f(\Phi_B(T_\delta)))\\ &\quad = \Pi_S^\perp\big(\overline{\omega}\cdot\partial_\varphi h + \varepsilon^2\mathbb{A}\xi\cdot\partial_\varphi h - J\,(1-\Phi_B^{\le 5}(\varepsilon\overline{v}))h + gh\end{aligned} \tag{6.28}$$

where g satisfies the same estimates as (6.19).

6.1.1. Hamiltonian of the linearized operator Following Remark 6.3, we evidence the terms homogeneous in the Hamiltonian of $\mathcal{L}_\omega$, let us call it H, whose Hamiltonian vector fields have degree ≤ 5, since they are NOT perturbative. As explained in (6.26) this entails expanding the map $\Phi_B(T_\delta)$ in powers of ε up to order five plus a *small remainder* $\tilde{q}$.

We consider the symplectic form in the extended phase space $(\varphi,\eta,z)\in\mathbb{R}^\nu\times\mathbb{R}^\nu\times H_S^\perp$

$$\Omega_e(\varphi,\eta,z) := d\varphi\wedge d\eta + \sum_{j\in S^c}\frac{1}{\mathrm{i}\lambda(j)}dz_j\wedge dz_{-j} \tag{6.29}$$

with the Poisson brackets (recalling $\{\cdot,\cdot\}$ defined in (2.2))

$$\{F,G\}_e := \partial_\varphi F\partial_\eta G - \partial_\eta F\partial_\varphi G + \{F,G\}. \tag{6.30}$$

The Hamiltonian of the operator (6.15) respect to the symplectic form (6.29) is (see (6.28))

$$\mathsf{H} := \mathsf{H}_0 + \sum_{i=1}^5 \varepsilon^i \mathsf{H}_i + H_{>5} + \sum_{i=2}^5 \varepsilon^i \mathsf{H}_{\mathcal{R}_i} + \mathsf{H}_{\mathcal{R}_{>5}} \tag{6.31}$$

with

$$\begin{aligned}&\mathsf{H}_0 = \overline{\omega}\cdot\eta + \frac{1}{2}\int_{\mathbb{T}} z^2\,dx,\quad \mathsf{H}_1 = -\frac{1}{2}\int_{\mathbb{T}}\overline{v}\,z^2\,dx,\quad \mathsf{H}_2 = \mathbb{A}\xi\cdot\eta - \frac{1}{2}\int_{\mathbb{T}}\Psi_2(\overline{v})\,z^2\,dx,\\ &\mathsf{H}_i = -\frac{1}{2}\int_{\mathbb{T}}\Psi_i(\overline{v})\,z^2\,dx,\quad 3\le i\le 5,\ \text{and}\ \|X_{\mathsf{H}_{>5}}\|_s^{\gamma,\mathcal{O}_0}, \|X_{\mathsf{H}_{\mathcal{R}_{>5}}}\|_s^{\gamma,\mathcal{O}_0}\\ &\lesssim_s \varepsilon^6 + \varepsilon\|\mathfrak{I}_\delta\|_{s+\sigma}^{\gamma,\mathcal{O}_0}\end{aligned} \tag{6.32}$$

for some $\sigma > 0$. The functions $\mathsf{H}_{\mathcal{R}_i}$, $\mathsf{H}_{\mathcal{R}_{>5}}$ are the quadratic forms associated to the corresponding linear operators, thus the estimates on the Hamiltonian vector fields can be deduced from (6.19), (6.18). It is easily seen that

$$\mathsf{H}_0 + \sum_{i=1}^{5} \varepsilon^i (\mathsf{H}_i + \mathsf{H}_{\mathcal{R}_i}) = \Pi^{2d_y+d_z=2}(\mathcal{H}^{(\leq 7)} \circ A_\varepsilon)|_{\substack{y=\eta \\ \theta=\varphi}}.$$

Of course one can be even more explicit and write everything in terms of the original Hamiltonian (1.4) and of the generating functions of the weak BNF, for example one has

$$\begin{aligned} &\mathsf{H}_0 = (H^{(2)} \circ A_1)|_{\substack{y=\eta \\ \theta=\varphi}}, \qquad \mathsf{H}_1 \overset{(3.22)}{=} (H^{(3,2)} \circ A_1)|_{\substack{y=\eta \\ \theta=\varphi}}, \\ &\mathsf{H}_2 + \mathsf{H}_{\mathcal{R}_2} = (\Pi^{d_y=1}(Z_2^{(4,0)} \circ A_1) + (\Pi^{d_z=2}\{F^{(3,\leq 1)}, H^{(3)}\}) \circ A_1)|_{\substack{y=\eta \\ \theta=\varphi}}. \end{aligned} \tag{6.33}$$

The terms H_i can be computed explicitly, however we only need to prove that they fit the following definitions.

Definition 6.4. We say that a matrix $\mathtt{B} := \left((\mathtt{B})_j^{j'}(l-l')\right)_{j,j'\in\mathbb{Z},l,l'\in\mathbb{Z}^\nu}$ is *almost diagonal* if there exists a constant $\mathtt{C} > 0$ such that, if $(\mathtt{B})_j^{j'}(l-l') \neq 0$, then $\langle j - j', l - l'\rangle \leq \mathtt{C}$ for all $j, j' \in S^c, l, l' \in \mathbb{Z}^\nu$.

Let $B(\varphi)\colon H^s(\mathbb{T}) \to H^s(\mathbb{T})$ be a Töpliz in time operator (recall (2.12)). We say that $B(\varphi)$ is almost diagonal if its associated matrix is almost diagonal.

Let $H := H(\varphi)$ be a quadratic Hamiltonian of the form $H = (A(\varphi)z, z)_{L^2}$, where $A(\varphi)$ is a Töpliz in time operator. We say that H and its vector field are almost diagonal if $A(\varphi)$ is almost diagonal.

Remark 6.5. It is easy to verify that if X and Y are almost diagonal operators then $X+Y$, $X \circ Y$ are almost diagonal.

Definition 6.6. Let $p \in \mathbb{N}$ and $m \in \mathbb{R}$. We say that a pseudo differential operator $\mathfrak{B} = \mathrm{Op}(b(\varphi, x, j))$ (recall Definition 2.8) is *homogenenous* of degree p in the function $\overline{v}$ in (6.25) if its symbol $b(\varphi, x, j) \in S^m$ has the form

$$b(\varphi, x, j) := \sum_{j_1,\dots,j_p\in S} C_{j_1,\dots,j_p}(j)\sqrt{\xi_{j_1}\cdots\xi_{j_p}}\, e^{\mathrm{i}(j_1+\dots+j_p)x} e^{\mathrm{i}(\mathtt{l}(j_1)+\dots+\mathtt{l}(j_p))\cdot\varphi}. \tag{6.34}$$

Definition 6.7. Let $p \in \mathbb{N}$. We say that a Hamiltonian is pseudo differential and p-homogeneous if it has the form

$$H_p(z) = \frac{1}{2}\int_{\mathbb{T}} f_p(\overline{v})z \cdot z dx + \frac{1}{2}\int_{\mathbb{T}} \mathfrak{B}_p z \cdot z + \frac{1}{2}\int_{\mathbb{T}} \mathcal{R}_p z \cdot z, \tag{6.35}$$

where f_p is a homogeneous real valued function of $\overline{v}$ (of degree p) of the form

$$f_p(\overline{v}) := \sum_{j_1,\dots,j_p\in S} (f_p)_{j_1,\dots,j_p}\sqrt{\xi_{j_1}\cdots\xi_{j_p}}\, e^{\mathrm{i}(j_1+\dots+j_p)x} e^{\mathrm{i}(\mathtt{l}(j_1)+\dots+\mathtt{l}(j_p))\cdot\varphi}, \tag{6.36}$$

$\mathfrak{B}_p \in OPS^{-2}$ is a p-homogeneous pseudo differential operator according to Definition 6.6 which is self-adjoint w.r.t. to $(\cdot,\cdot)_{L^2}$; finally $\mathcal{R}_p$ is a *finite dimensional* operator of the form (6.16) with g_j, χ_j p-homogeneous functions of $\overline{v}$.

Lemma 6.8. *The Hamiltonians* $\mathsf{H}_i + \mathsf{H}_{\mathcal{R}_i}$ *in* (6.31) *are almost diagonal according to Definition* 6.4 *and pseudo differential homogeneous Hamiltonians according to Definition* 6.7.

Proof. It follows by Proposition 6.2 and Remark 6.3. □

Lemma 6.9. *Let* $p, q \in \mathbb{N}$ *and consider* H_p, G_q *two pseudo differential and homogenous Hamiltonians of degree respectively* p *and* q. *Then there is a pseudo differential and* $(p+q)$*-homogeneous Hamiltonian* $\widetilde{H}$ *such that* $X_{\widetilde{H}} = X_{\{H_p, G_q\}_e}$, *where* $\{\cdot,\cdot\}_e$ *are defined in* (6.30) *and* X_H *denotes the Hamiltonian vector field generated by* H.

Proof. By assumption H_p and G_q have the form (6.35) for some f_p, f_q real valued and some self-adjoint pseudo differential homogenenous operators $\mathfrak{B}_p$ and $\mathfrak{B}_q$. Then we have (recalling (2.2), (1.4) and (3.5))

$$\{H_p, G_q\}_e = \int_{\mathbb{T}} A_1 z \cdot z dx + \int_{\mathbb{T}} A_2 z \cdot z dx, \qquad A_1 := f_p \circ \partial_x \circ f_q,$$
$$A_2 := f_p \circ J \circ \mathfrak{B}_q + \mathfrak{B}_p \circ J \circ f_q + \mathfrak{B}_p \circ J \circ \mathfrak{B}_q + 3 f_p \circ \Lambda \partial_x \circ f_q.$$

One has that the Hamiltonian

$$\widetilde{H} := \frac{1}{2}\int_{\mathbb{T}} (A_1 + A_1^*) z \cdot z dx + \frac{1}{2}\int_{\mathbb{T}} (A_2 + A_2^*) z \cdot z dx$$

is equivalent to $\{H_p, G_q\}_e$ in the sense that they generate the same vector field. Here A_i^*, $i = 1, 2$, denotes the adjoint of A_i w.r.t. the L^2 scalar product. Notice that

$$A_1 + A_1^* = f_p f_q \partial_x + f_p (f_q)_x - f_q \circ \partial_x \circ f_p = f_p f_q \partial_x + f_p (f_q)_x - f_q f_p \partial_x$$
$$- f_q (f_p)_x = f_p (f_q)_x - f_q (f_p)_x,$$

which is an homogeneous function of $\overline{v}$ of degree $p+q$. Using the results on compositions of pseudo differential operators in Sect. 2 of [33], the fact that J is skew-self-adjoint, $\mathfrak{B}_i$, $i = p, q$, are self-adjoint, and f_q, f_p are real valued, we deduce that the operator A_2 is a skew-self-adjoint operator in OPS^{-1}. Hence, using the formula (2.13) in [33] for the adjoint, we have that $A_2 + A_2^*$ is pseudo differential homogeneous operator (according to Definition 6.6) in OPS^{-2}. □

7. Reduction and Inversion of the Linearized Operator

The aim of the section is to prove the claim in (6.9). As explained in the introduction, first one should reduce the unbounded parts of $\mathcal{L}_\omega$ and then use classical KAM reducibility results to diagonalize. The difficulties arise from the fact that a few steps of this procedure must be done by hand, since they do not fit the typical smallness conditions, see [33].

The key result of this section is the following.

Theorem 7.1. *Consider* $\mathcal{L}_\omega = \mathcal{L}_\omega(\mathfrak{I}_\delta)$ *in* (6.15) *and fix*

$$\tau = 2\nu + 6, \quad \mathtt{b}_0 := 6\tau + 6, \quad \mathtt{b} = \mathtt{b}_0 + s_0. \tag{7.1}$$

There exist $\mathcal{S} > s_0$ and $\mu_1 = \mu_1(\nu) > 0$ such that, if condition (6.12) is satisfied with $\mathfrak{p}_1 = \mu_1$, then the following holds. There exists a constant $m(\omega)$ defined for $\omega \in \Omega_\varepsilon$ with

$$|m-1-\varepsilon^2 c(\omega)|^{\gamma,\Omega_\varepsilon} \lesssim \varepsilon^4, \quad |m|^{lip} \lesssim 1, \quad c(\omega) := \vec{v}\cdot\xi, \quad \vec{v}_k = \frac{2}{3}(1+\bar{J}_k^2), \quad k = 1, \dots, \nu, \tag{7.2}$$

[3] *such that for all ω in the set $\mathcal{O}_\infty^{2\gamma}$, where (recall that $\mathcal{O}_0 \subseteq \mathcal{G}_0$, see (5.7))*

$$\mathcal{O}_\infty^{2\gamma} = \mathcal{O}_\infty^{2\gamma}(i) := \{\omega \in \mathcal{O}_0 : \ |\omega\cdot\ell - m(\omega) j| > \frac{2\gamma}{\langle \ell\rangle^\tau}, \forall \ell \in \mathbb{Z}^\nu, \ \forall j \in S^c\}, \tag{7.3}$$

there exists a real, bounded linear operator $\Upsilon = \Upsilon(\omega) : H^s_{S^\perp} \to H^s_{S^\perp}$, for all $s_0 \le s \le \mathcal{S}$, such that

$$\mathcal{L} := \Upsilon \mathcal{L}_\omega \Upsilon^{-1} = \Pi_S^\perp\big(\omega\cdot\partial_\varphi - mJ - \varepsilon^2 \mathfrak{D}(\omega) + \mathcal{P}_0\big) \tag{7.4}$$

where $\mathfrak{D}(\omega)$ is the diagonal operator of order -1 defined as $\mathfrak{D} := \mathfrak{D}(\omega) = \mathrm{diag}(\mathrm{i}\kappa_j)_{j\in S^c}$, with

$$\kappa_j := \kappa_j(\omega) := \lambda(j)\big(\mathfrak{l}_j(\omega) - c(\omega)\big) \in \mathbb{R}, \quad |\kappa_j|^{sup} \lesssim |j|^{-1}, \tag{7.5}$$

where $\mathfrak{l}_j$ is defined in (5.6). The constant m depends on i and for $\omega \in \mathcal{O}_\infty^{2\gamma}(i_1) \cap \mathcal{O}_\infty^{2\gamma}(i_2)$ one has

$$|\Delta_{12} m| \lesssim \varepsilon \|i_1 - i_2\|_{s_0+\mu_1}, \tag{7.6}$$

where $\Delta_{12} m := m(i_1) - m(i_2)$. The remainder $\mathcal{P}_0$ in (7.4) is defined and Lipschitz in ω belonging to the set $\mathcal{O}_\infty^{2\gamma}$ and is Lip--1-modulo tame (see Definition 2.7) with

$$\mathfrak{M}^{\sharp,\gamma^{3/2}}_{\mathcal{P}_0}(-1, s, \mathfrak{b}_0) \lesssim_s \varepsilon^{4-3a} + \varepsilon\gamma^{-1}\|\mathfrak{I}_\delta\|^{\gamma,\mathcal{O}_0}_{s+\mu_1}, \tag{7.7}$$

$$\|\langle D_x\rangle^{1/2}\underline{\Delta_{12}\mathcal{P}_0}\langle D_x\rangle^{1/2}\|_{\mathcal{L}(H^{s_0})}, \ \|\langle D_x\rangle^{1/2}\underline{\Delta_{12}\langle\partial_\varphi\rangle^{\mathfrak{b}_0}\mathcal{P}_0}\langle D_x\rangle^{1/2}\|_{\mathcal{L}(H^{s_0})} \lesssim \varepsilon\gamma^{-1}\|i_1 - i_2\|_{s_0+\mu_1}, \tag{7.8}$$

for all $\omega \in \mathcal{O}_\infty^{2\gamma}(i_1) \cap \mathcal{O}_\infty^{2\gamma}(i_2)$. Moreover if $u = u(\omega)$ depends on the parameter $\omega \in \mathcal{O}_\infty^{2\gamma}$ in a Lipschitz way then

$$\|\Upsilon^{\pm1} u\|^{\gamma,\mathcal{O}_\infty^{2\gamma}}_s \lesssim_s \|u\|^{\gamma,\mathcal{O}_\infty^{2\gamma}}_s + \varepsilon\gamma^{-1}\|\mathfrak{I}_\delta\|^{\gamma,\mathcal{O}_0}_{s+\mu_1}\|u\|^{\gamma,\mathcal{O}_\infty^{2\gamma}}_{s_0}, \quad s_0 \le s \le \mathcal{S}. \tag{7.9}$$

The result above has two relevant consequences. Firstly it shows that the operator $\mathcal{L}_\omega$ in (6.15) can be conjugated to an operator (see (7.4)) which is "diagonal", at the highest order of derivatives, plus a remainder which is -1-smoothing. In addition to this, thanks to a *linear BNF* procedure (performed in Sect. 7.2), the non-diagonal term $\mathcal{P}_0$ in (7.4) has a size much smaller than ε (see estimates (7.7), (7.8)). In particular it is "perturbative" w.r.t. the constant γ in (5.3). This allows us to apply the reducibility scheme of [33] in order to complete the diagonalization of the operator $\mathcal{L}$ (see Theorem 7.13). Then the inversion assumption (6.9) follows directly from Proposition 7.14).
Strategy of the Proof of Theorem 7.1.

[3] Notice that $\xi = \xi(\omega)$, recall (4.11).

- **Reduction at the highest order** The first step is to exploit the pseudo differential structure of the operator $\mathcal{L}_\omega$ in order to conjugate it to an operator which has constant coefficients up to a smoothing remainder of order -1. To this purpose we use changes of variables generated as the time-one flow map $\Phi^\tau|_{\tau=1}$ of Hamiltonians of the form

$$S(\tau,\varphi,z)=\int b(\tau,\varphi,x)z^2dx,\qquad b(\tau,\varphi,x):=\frac{\beta(\varphi,x)}{1+\tau\beta_x(\varphi,x)},\tag{7.10}$$

$$\partial_\tau\Phi^\tau u=\Pi_S^\perp[(J\circ b)\Pi_S^\perp[\Phi^\tau u]],\qquad \Phi^0u=u,\tag{7.11}$$

where β is some smooth function. In Proposition C.2 we show that Φ^τ is well defined as symplectic map on $H^s_{S^\perp}$ (see Lemma C.1) and study the structure of $\Phi^\tau\mathcal{L}_\omega(\Phi^\tau)^{-1}$. Proposition C.2 gives an explicit formula for the new coefficient at the highest order (see (C.17)). Then Corollary 3.6 of [31] (see also Proposition 3.6 in [33]) provides the solution for the Eq. (C.17)=const provided that some smallness condition is satisfied. This smallness condition has the form

$$C(s_1)\gamma^{-1}\|a_0\|^{\gamma,\mathcal{O}}_{s_1}\ll 1\tag{7.12}$$

for some $s_0+\mathfrak{p}_1>s_1>s_0$ and some constant $C(s_1)>0$. As shown in [33], due to the Hamiltonian structure, this reduces $\mathcal{L}_\omega$ to constant coefficients up to a correction of order -1.

Unfortunately, since here $\gamma=\varepsilon^{2+a}$, $a>0$, by (6.17), the coefficient $a_0(\varphi,x)$ in $\mathcal{L}_\omega$ does not satisfy (7.12). This is why we have to perform some preliminary steps in order to enter in the perturbative regime where we apply the scheme described in the proof of Corollary 3.6 in [31].

We first "regularize" the purely polynomial terms H_i (see (6.32)) by hand, by exploiting their homogeneity according to Definition 6.7. After that we are left with only unbounded terms which satisfy the smallness conditions of [33]. We "regularize" them by applying the results of [33] adapted to our slightly more general setting, see Proposition C.2.

Remark 7.2. In order to determine the correct change of variables in the regularization of H_i, it will be convenient to use the *Formal Lie expansions*. We recall that $H\circ(\Phi^\tau)^{-1}$ satisfies, for $\tau\in[0,1]$,

$$\partial_\tau(H\circ(\Phi^\tau)^{-1})=\{S(\tau),H\circ(\Phi^\tau)^{-1}\}.\tag{7.13}$$

By setting $S:=S(0)$, the Lie expansion of the conjugated Hamiltonian $H\circ\Phi^{-1}$ is the following:

$$H\circ(\Phi^\tau)^{-1}=H+\tau\{S,H\}_e+\frac{\tau^2}{2}\Big(\{S,\{S,H\}_e\}_e+\{(\partial_\tau S)(0),H\}_e\Big)+\cdots,\tag{7.14}$$

where the Poisson brackets $\{\cdot,\cdot\}_e$ are in (6.30). Recall that Φ is a C^k map from H^s to H^{s-k}. Therefore the Taylor expansion of the conjugated Hamiltonian coincides with the Lie series of the generator up to any order τ^k.

- **Linear BNF** The second step is to diagonalize the bounded terms. Here we diagonalize "by hand" the terms up to order ε^3, by exploiting the fact that they are almost diagonal according to Definition 6.4 and applying a linear BNF. Once this is done, the full diagonalization follows by a standard KAM reducibility theorem (see Theorem 7.13).

7.1. Reduction at the highest order. In the following we shall assume that the (6.12) holds with some $\mathfrak{p}_1 \gg 1$. The loss of regularity $\mathfrak{p}_1$ will be determined explicitly at the end of the section. In order to perform the *non-perturbative* steps, we construct changes of coordinates $\mathcal{B}_i$, $i = 1, 2, 3, 4, 5$, as the time-one flow maps generated by Hamiltonians as in (7.10). Then we set $\mathcal{L}_0 := \mathcal{L}_\omega$ and define iteratively $\mathcal{L}_i := \mathcal{B}_i \mathcal{L}_{i-1} \mathcal{B}_i^{-1}$. Note that $\mathcal{L}_0$ is pseudo differential plus a finite rank operator. Even though the $\mathcal{B}_i$ preserve the pseudo differential structure, in order to have a good quantitative control on the symbols we shall fix appropriate values

$$p \geq s_0, \quad \rho \geq s_0 + 6\tau + 9, \tag{7.15}$$

and write

$$\mathcal{L}_i = \mathcal{B}_i \mathcal{L}_{i-1} \mathcal{B}_i^{-1} = \Pi_S^\perp \Big(\omega \cdot \partial_\varphi - J \circ \big(1 + \varepsilon^2 c_i(\omega) + a_i(\varphi, x)\big) + \mathrm{Op}(\mathfrak{q}_i) + \widehat{\mathcal{Q}}_i \Big) \tag{7.16}$$

where $c_i(\omega)$ is a constant, $a_i, \mathfrak{q}_i$ are symbols, $\mathrm{Op}(\mathfrak{q}_i)$ is of order -1 and $\widehat{\mathcal{Q}}_i \in \mathfrak{L}_{\rho,p}$. This is a class of operators of order $-\rho$ which we introduced in [33] (we recall it in Definition C.5). Note that by Lemma C.7 $\mathcal{Q}_0 =: \widehat{\mathcal{Q}}_0$ belongs to $\mathcal{L}_{\rho,p}$ for all ρ, p, with bounds on $\mathbb{M}^\gamma_{\widehat{\mathcal{Q}}_0}(s, \mathfrak{b})$ given in the same lemma. Then one proves iteratively that

$$\begin{aligned} \|a_i\|_s^{\gamma, \mathcal{O}_0} &\lesssim_s \varepsilon^{i+1} + \varepsilon \|\mathfrak{I}_\delta\|^{\gamma, \mathcal{O}_0}_{s+\sigma_0+\sigma_{i+3}}, \ s \geq s_0, \\ \|\Delta_{12} a_i\|_p &\lesssim_p \varepsilon(1 + \|\mathfrak{I}_\delta\|_{p+\sigma_0+\sigma_{i+3}}) \|i_1 - i_2\|_{p+\sigma_0+\sigma_{i+3}}. \end{aligned} \tag{7.17}$$

$$\begin{aligned} |\mathfrak{q}_i|^{\gamma, \mathcal{O}_0}_{-1,s,\alpha} &\lesssim_{s,\alpha,\rho} \varepsilon(1 + \|\mathfrak{I}_\delta\|^{\gamma, \mathcal{O}_0}_{s+\sigma_0+\sigma_{i+3}}), \quad s \geq s_0, \\ |\Delta_{12} \mathfrak{q}_i|_{-1,p,\alpha} &\lesssim_{p,\alpha,\rho} \varepsilon(1 + \|\mathfrak{I}_\delta\|_{p+\sigma_0+\sigma_{i+3}}) \|i_1 - i_2\|_{p+\sigma_0+\sigma_{i+3}}. \end{aligned} \tag{7.18}$$

Note that the size of a_i (in the low norm) is decreasing in i. Regarding the remainders, the numbers $\mathbb{M}^\gamma_{\widehat{\mathcal{Q}}_i}(s, \mathfrak{b})$ control the norm of the corresponding operator, see Definition C.5. We have

$$\begin{aligned} \mathbb{M}^\gamma_{\widehat{\mathcal{Q}}_i}(s, \mathfrak{b}) &\lesssim_{s,\rho} \varepsilon(1 + \|\mathfrak{I}_\delta\|^{\gamma, \mathcal{O}_0}_{s+\sigma_0+\sigma_{i+3}}), \quad s_0 \leq s \leq \mathcal{S}, \qquad 0 \leq \mathfrak{b} \leq \rho - 2, \\ \mathbb{M}_{\Delta_{12}\widehat{\mathcal{Q}}_i}(p, \mathfrak{b}) &\lesssim_{p,\rho} \varepsilon(1 + \|\mathfrak{I}_\delta\|_{p+\sigma_0+\sigma_{i+3}}) \|i_1 - i_2\|_{p+\sigma_0+\sigma_{i+3}}, \qquad 0 \leq \mathfrak{b} \leq \rho - 3, \end{aligned} \tag{7.19}$$

with σ_0 defined in Proposition 6.2 and $\sigma_{i+3} > 0$ $i = 1, \ldots, 5$, depending only on ν (essentially σ_{i+3} are the losses coming from the application of Proposition C.2). Note that we can obtain (7.16) for any ρ, p satisfying (7.15); however, if we want (7.18) to hold for some given p, we have to assume a smallness condition (6.12) with $p + \sigma_0 + \sigma_{i+3} < \mathfrak{p}_1$.

Step (ε). Consider the Hamiltonian

$$S(\tau) := \frac{1}{2} \int_{\mathbb{T}} b_1(\tau, \varphi, x)\, z^2\, dx = \varepsilon S_1 + \varepsilon^2 \tau S_2 + \varepsilon^3 \tau^2 S_3 + S_4(\tau), \quad b_1 := \frac{\varepsilon \beta_1}{1 + \tau \varepsilon (\beta_1)_x}, \tag{7.20}$$

$$S_1 := \frac{1}{2} \int_{\mathbb{T}} \beta_1\, z^2\, dx, \quad S_2 := -\frac{1}{4} \int_{\mathbb{T}} \partial_x(\beta_1^2)\, z^2\, dx, \quad S_3 := \frac{1}{2} \int_{\mathbb{T}} \beta_1 (\beta_1)_x^2 z^2\, dx, \tag{7.21}$$

with $S_4(\tau) \sim O(\tau^3 \varepsilon^4)$ and for some function β_1 of the form (6.36) with $p = 1$ and some coefficients $(\beta_1)_j$, $j \in S$, to be determined. The Hamiltonian system associated

to $S(\tau)$ is of the form (7.11) with $b \rightsquigarrow b_1$. We call $\mathcal{B}_1$ the flow at time-one generated by $S(\tau)$, then the Hamiltonian of the conjugated linearized operator $\mathcal{B}_1\mathcal{L}_\omega\mathcal{B}_1^{-1}$ is (recall (6.31), (6.32) and Remark 7.2)

$$\mathsf{H}\circ\mathcal{B}_1^{-1} = \mathsf{H}_0 + \varepsilon\mathsf{H}_1^{(1)} + \varepsilon^2\mathsf{H}_2^{(1)} + \varepsilon^3\mathsf{H}_3^{(1)} + o(\varepsilon^3), \qquad \mathsf{H}_1^{(1)} := \{S_1, \mathsf{H}_0\}_e + \mathsf{H}_1, \tag{7.22}$$

$$\mathsf{H}_2^{(1)} := \frac{1}{2}\{S_1, \{S_1, \mathsf{H}_0\}_e\}_e + \{S_1, \mathsf{H}_1\}_e + \frac{1}{2}\{S_2, \mathsf{H}_0\}_e + \mathsf{H}_2 + \mathsf{H}_{\mathcal{R}_2}, \tag{7.23}$$

where, by Lemma 6.9, $\mathsf{H}_3^{(1)}$ is some pseudo differential 3-homogeneous Hamiltonian of the form (6.35). Notice that also $H_1^{(1)}$ and $\mathsf{H}_2^{(1)}$ (for $\eta = 0$) are pseudo differential and 1-homogeneous, resp. 2-homogenenous, Hamiltonians according to Definition 6.7. We want to solve the following equation

$$\mathsf{H}_1^{(1)} = \mathsf{H}_1 + \{S_1, \mathsf{H}_0\}_e = \mathsf{H}_1 + \{S_1, \mathsf{H}_0\} + \overline{\omega}\cdot\partial_\varphi S_1 = \int \mathfrak{B}_1(z)\, z\, dx, \tag{7.24}$$

where $\mathfrak{B}_1$ is some pseudo differential operator of order -2. Recalling (6.32), and expanding $\{S_1, \mathsf{H}_0\}$ as in the proof of Lemma 6.9, we note that the Eq. (7.24) is equivalent to the following one

$$\overline{\omega}\cdot\partial_\varphi\beta_1 - (\beta_1)_x - \overline{v} = 0. \tag{7.25}$$

Hence we choose $\beta_1 = \frac{1}{3}(\Lambda\partial_x)^{-1}\overline{v}$ and we note that

$$\|\beta_1\|_s^{\gamma,\mathcal{O}_0} \lesssim_s 1, \quad \forall s \geq s_0. \tag{7.26}$$

With the choice in (7.25) we have

$$\mathfrak{B}_1 := [3\Lambda\partial_x, \beta_1]. \tag{7.27}$$

In this way the Hamiltonians in (7.22), (7.23) become

$$\mathsf{H}_1^{(1)} := \int \mathfrak{B}_1(z)\, z\, dx, \qquad \mathsf{H}_2^{(1)} := \mathsf{H}_2 + \frac{1}{2}\{S_2, \mathsf{H}_0\}_e + \mathsf{H}_{\mathcal{R}_2} + \frac{1}{2}\{S_1, \mathsf{H}_1\} + \frac{1}{2}\{S_1, \int_{\mathbb{T}} \mathfrak{B}_1(z)\, z\, dx\}. \tag{7.28}$$

By (7.26) the smallness assumption of Lemma C.1 is satisfied. By (6.17), (6.21), (6.22) and using the assumption (6.12) with $\mathfrak{p}_1$ sufficiently large the condition (C.15) holds. In this case $\mathfrak{q} \rightsquigarrow 0$ (see (C.13)).

Then Proposition C.2 applies and the new linearized operator $\mathcal{L}_1 := \mathcal{B}_1\mathcal{L}_\omega\mathcal{B}_1^{-1}$ has the form in (7.16) with $i = 1$ and $c_1 = 0$. By (7.26), (6.21), (C.19), (C.17), (6.15), we have that $\widehat{\mathcal{Q}}_1 \in \mathfrak{L}_{\rho,p}(\mathcal{O})$ (see Definition C.5) (with ρ, p satisfying (7.15)) and (7.18), (7.19) hold for $i = 1$.

The only estimates that are not given by Lemma C.2 are (7.17). The coefficient a_1 is given by (C.17) with $m \rightsquigarrow 1$, $a \rightsquigarrow a_0$, $a_+ \rightsquigarrow a_1$ and $\tilde{\beta}$ such that $x \mapsto x + \tilde{\beta}$ is the inverse of $x \mapsto x + \beta_1$. By the choice of β_1 in (7.25) we have eliminated the ε-terms from a_1. Hence by (7.26) and (6.17) we get (7.17) for $i = 1$.

Step (ε^2). Now we deal with the terms of order ε^2 of the Hamiltonian (6.31). We consider the auxiliary Hamiltonian

$$\tilde{S}(\tau) = \frac{1}{2}\int_{\mathbb{T}} b_2(x,\varphi)\, z^2\, dx = \varepsilon^2\tilde{S}_2 + \tilde{S}_4(\tau), \quad b_2 := \frac{\varepsilon^2\beta_2}{1 + \tau\varepsilon^2(\beta_2)_x},$$

$$\tilde{S}_2 := \frac{1}{2}\int_{\mathbb{T}} \beta_2 z^2 \, dx, \tag{7.29}$$

where $\tilde{S}_4(\tau) := \tilde{S}(\tau) - \varepsilon^2 \tilde{S}_2 \sim O(\tau\varepsilon^4)$ and β_2 is some function of the form (6.34), with $p = 2$, to be determined. Notice that $(\partial_\tau \tilde{S})(0) \sim O(\varepsilon^4)$. The Hamiltonian system associated to $\tilde{S}(\tau)$ is of the form (7.11) with $b \rightsquigarrow b_2$. If $\mathcal{B}_2$ is the flow at time-one generated by $\tilde{S}(\tau)$, then the Hamiltonian of the conjugated linearized operator $\mathcal{L}_2 := \mathcal{B}_2\mathcal{L}_1\mathcal{B}_2^{-1}$ is (recall (7.22), (7.28), (7.14))

$$\begin{aligned} &\mathsf{H} \circ \mathcal{B}_1^{-1} \circ \mathcal{B}_2^{-1} = \mathsf{H}_0 + \varepsilon \mathsf{H}_1^{(1)} + \varepsilon^2 \mathsf{H}_2^{(2)} + \varepsilon^3 \mathsf{H}_3^{(2)} + o(\varepsilon^3) \\ &\mathsf{H}_2^{(2)} := \mathsf{H}_2^{(1)} + \{\tilde{S}_2, \mathsf{H}_0\}_e, \qquad \mathsf{H}_3^{(2)} := \mathsf{H}_3^{(1)} + \{\tilde{S}_2, \mathsf{H}_1^{(1)}\}_e. \end{aligned} \tag{7.30}$$

We want to solve the equation

$$\mathsf{H}_2^{(2)} = \mathsf{H}_2^{(1)} + \overline{\omega} \cdot \partial_\varphi \tilde{S}_2 + \{\tilde{S}_2, \mathsf{H}_0\} = c + \int_{\mathbb{T}} \mathfrak{B}_2(z)\, z \, dx + \mathsf{H}_{\mathcal{R}_2}, \tag{7.31}$$

where $\mathfrak{B}_2$ is some pseudo differential and 2-homogeneous operator of order -2 (see Definition 6.6), c is some constant to be determined and $\mathsf{H}_{\mathcal{R}_2}$ (possibly different from the one in (6.31)) is a Hamiltonian with the form (6.35). By Lemma 6.9 we have that $\mathsf{H}_2^{(1)} + \{\tilde{S}_2, \mathsf{H}_0\}$ can be written in the form (6.35) with, in particular (see (6.24) for the definition of Ψ_2)

$$f_2(\overline{v}) := -\Psi_2(\overline{v}) + \frac{1}{4}\partial_{xx}(\beta_1^2) - \frac{1}{2}\beta_1 \overline{v}_x + \frac{1}{2}\overline{v}(\beta_1)_x, \tag{7.32}$$

and some $\mathfrak{B}_2 \in OPS^{-2}$, as in Definition 6.6, up to a finite rank remainder. Hence the Eq. (7.31) is equivalent to

$$\overline{\omega} \cdot \partial_\varphi \beta_2 - (\beta_2)_x + f_2(\overline{v}) = c. \tag{7.33}$$

Since f_2 in (7.32) has the form (6.36) with $p = 2$, we look for a function β_2 of the same form in (6.36) with some coefficients $(\beta_2)_{j_1,j_2} \in \mathbb{C}$. Hence Eq. (7.33) reads

$$\begin{aligned} &\big[\lambda(j_1)+\lambda(j_2) - (j_1 + j_2)\big](\beta_2)_{j_1,j_2} + (f_2)_{j_1,j_2} = 0, \quad \text{for} \quad \lambda(j_1) + \lambda(j_2) - (j_1 + j_2) \neq 0, \\ &(f_2)_{j_1,j_2} = c, \quad \text{for} \quad \lambda(j_1)+\lambda(j_2) - (j_1 + j_2) = 0. \end{aligned} \tag{7.34}$$

We have that, for $j_1, j_2 \in S$, $\lambda(j_1)+\lambda(j_2)-(j_1+j_2) = 0$ if and only if $j_1 + j_2 = 0$, since $j_1 j_2 \neq -1$. The terms with $j_1 = -j_2$ corresponds to the average in x of the function $f_2(\overline{v})$. Hence we set

$$c := \frac{1}{2\pi}\int_{\mathbb{T}} f_2(\overline{v}) dx. \tag{7.35}$$

and we evaluate explicitly it. The functions $\Psi_2(\overline{v})$ and $\partial_{xx}(\beta_1^2)$ do not contribute since they have zero average in space.

Recalling that $\beta_1 = \frac{1}{3}(\Lambda\partial_x)^{-1}\overline{v}$ we have

$$\begin{aligned} \int_{\mathbb{T}} f_2(\overline{v}) dx &= \frac{1}{6}\int_{\mathbb{T}} \Big((\Lambda^{-1}\partial_x^{-1}\partial_x\overline{v}) \cdot \overline{v} - (\Lambda^{-1}\partial_x^{-1}\overline{v}) \cdot \overline{v}_x\Big) dx = \frac{1}{3}\int_{\mathbb{T}} (\Lambda^{-1}\overline{v}) \cdot \overline{v} dx \\ &\overset{(3.5)}{=} \frac{1}{3}\int_{\mathbb{T}} (\overline{v}^2 + \overline{v}_x^2). \end{aligned}$$

Then the constant $c = c(\omega)$ (recall the (4.11)) in (7.35) is given by

$$c(\omega) = \frac{1}{3}\sum_{j\in S}(1+j^2)\,\xi_j = \frac{2}{3}\sum_{j\in S^+}(1+j^2)\,\xi_j. \tag{7.36}$$

By noting that

$$\|\varepsilon^2\beta_2\|_s^{\gamma,\mathcal{O}} \lesssim_s \varepsilon^2 \quad \forall s \geq s_0, \tag{7.37}$$

by (7.18)–(7.17) with $i = 1$ and using the assumption (6.12) with $\mathfrak{p}_1$ sufficiently large the smallness assumption of Lemma C.1 and the condition (C.15) are satisfied. In this case $\mathtt{q} \rightsquigarrow \mathtt{q}_1$, hence by (7.18), (7.19) the bounds (C.13), (C.14) hold with $\mathtt{k}_1 \rightsquigarrow \varepsilon$, $\mathtt{k}_2 \rightsquigarrow \varepsilon$, $\mathtt{k}_3 \rightsquigarrow \varepsilon$, $f \rightsquigarrow \mathfrak{I}_\delta$. Then Proposition C.2 applies and $\mathcal{L}_2 := \Phi_2\mathcal{L}_1\Phi_2^{-1}$ with $i = 2$ and $c_2(\omega) := c(\omega)$ given in (7.36). By (7.37), (6.21), (C.19), (C.17) we have that $\widehat{\mathcal{Q}}_2 \in \mathfrak{L}_{\rho,p}(\mathcal{O})$ (with ρ, p as in (7.15)) and (7.18), (7.19) hold for $i = 2$. By (7.37), (7.18)–(7.17) for $i = 1$, (C.20), we have that (7.18)–(7.17) holds for $i = 2$.

Steps (ε^3)-(ε^4)-(ε^5). Consider $i = 3, 4, 5$. We proceed exactly as in the previous steps. We consider a change of coordinates $\mathcal{B}_i$ as the time-one flow map of

$$u_\tau = \Pi_S^\perp[\big(J \circ b_i(\tau)\big)\,u], \quad b_i := \frac{\varepsilon^i\beta_i}{1+\varepsilon^i\tau(\beta_i)_x} \tag{7.38}$$

for some smooth function β_i of the form (6.36) (with $p = i$) to be determined. Using Lemma 6.9 for the Hamiltonians of order ε^i, $i = 3, 4, 5$, we can choose β_i in order to solve an equation like the following

$$\overline{\omega}\cdot\partial_\varphi\beta_i - (\beta_i)_x = f_i(\overline{v}), \tag{7.39}$$

where f_i is a homogeneous function as in (6.36) (with $p = i$). The condition (1.13) implies that the Eq. (7.39) for $i = 4$ is solved up to remainders of the form

$$d(\omega) := d(\xi(\omega)) = \sum_{j_1,j_2\in S}\mathtt{d}(j_1, j_2)\xi_{j_1}\xi_{j_2}. \tag{7.40}$$

By (A.3) there are no small divisors for (7.39) if $i = 3$ or $i = 5$. By (7.18), (7.19) and by noting that

$$\|\varepsilon^i\beta_i\|_s^{\gamma,\mathcal{O}} \lesssim_s \varepsilon^i \quad i = 3, 4, 5, \ \forall s \geq s_0, \tag{7.41}$$

the smallness assumption of Lemma C.1 and the condition (C.15) are satisfied for the system (7.38). Arguing as in the previous steps we obtain that $\mathcal{L}_5 := \mathcal{B}_5\mathcal{B}_4\mathcal{L}_3\mathcal{B}_4^{-1}\mathcal{B}_5^{-1}$ has the form (7.16) with $i = 5$ and $c_5 = c_2 + \varepsilon^2 d(\xi)$. Moreover the bounds (7.18)–(7.17) hold for $i = 5$.

Remark 7.3. Since the symplectic maps $\mathcal{B}_i, i = 1, \ldots, 5$ are smooth in $\varepsilon\,\overline{v}$ (see (6.25) and Remark 7.2) and the Hamiltonian H has the Taylor expansion (6.31), then the operators $\mathrm{Op}(\mathtt{q}_5)$, $\widehat{\mathcal{Q}}_5$ in (7.16) may be expanded, in degree of homogeneity of $\varepsilon\overline{v}$, in the following way (see Remark C.3)

$$\mathrm{Op}(\mathtt{q}_5) = \sum_{i=1}^{3}\varepsilon^i\,\mathtt{q}_5^{(i)} + \mathtt{q}_5^{(\geq 4)}, \quad \widehat{\mathcal{Q}}_5 := \sum_{i=1}^{3}\varepsilon^i\,\widehat{\mathcal{Q}}_5^{(i)} + \widehat{\mathcal{Q}}_5^{(\geq 4)}$$

with

$$|\mathfrak{q}_5^{(\geq 4)}|^{\gamma,\mathcal{O}_0}_{-1,s,\alpha} \lesssim_{s,\alpha,\rho} \varepsilon^4 + \varepsilon\|\mathfrak{I}_\delta\|^{\gamma,\mathcal{O}_0}_{s+\sigma}, \quad \mathbb{M}^{\gamma}_{\widehat{\mathcal{Q}}_5^{(\geq 4)}}(s) \lesssim_{s,\rho} \varepsilon^4 + \varepsilon\|\mathfrak{I}_\delta\|^{\gamma,\mathcal{O}_0}_{s+\sigma} \tag{7.42}$$

for some $\sigma > 0$. Following Remark C.3 the $\widehat{\mathcal{Q}}_5^{(i)}$, $\widehat{\mathcal{Q}}_5^{(\geq 4)}$ are in $\mathfrak{L}_{\rho',p'}$, as is habitual we rename them ρ, p.

By (6.31) and the fact that the generators β_i in (7.10) are $\mathfrak{I}_\delta$-independent, it is clear that $\mathfrak{q}_5^{(\geq 4)}$ and $\widehat{\mathcal{Q}}_5^{(\geq 4)}$ contain terms of size ε^4, which are functions just of $\overline{v}$, and terms dependent also on $\mathfrak{I}_\delta$ of "size" $O(\varepsilon\|\mathfrak{I}_\delta\|_{s+\sigma})$, see the estimates (7.17), (7.18), (7.19). By the uniqueness of the Taylor expansion we have that $\sum_{i=1}^3 \varepsilon^i \left(\mathfrak{q}_5^{(i)} + \widehat{\mathcal{Q}}_5^{(i)}\right)$ coincide with the vector field $-\sum_{i=1}^3 \varepsilon^i J\nabla\mathcal{K}_i$ where, recalling (7.22), (7.30), (7.23), (7.36),

$$\mathcal{K}_1 := \mathsf{H}_1^{(1)}, \qquad \mathtt{Z}_0 + \mathcal{K}_2 := \mathsf{H}_2^{(2)}, \qquad \mathtt{Z}_0 := \mathbb{A}\xi\cdot\eta + \frac{c(\omega)}{2}\int_{\mathbb{T}} z^2\,dx, \tag{7.43}$$

and $\mathcal{K}_3$ is some pseudo differential 3-homogeneous Hamiltonian as in (6.35) with the corresponding function $f_3(\overline{v}) = 0$.

Now we apply Proposition 3.6 in [33] (or Corollary 3.6 in [31]) in order to make constant the coefficient a_5 of the linearized operator $\mathcal{L}_5$, namely we find β such that

$$\omega\cdot\partial_\varphi\beta - (1 + \varepsilon^2 c(\omega) + \varepsilon^4 d(\omega) + a_5(\varphi, x))(1+\beta_x) = \text{constant}. \tag{7.44}$$

Note that, by (7.17) with $i = 5$ and (6.12), the smallness condition (7.12) is satisfied by the function a_5. We have the following.

Proposition 7.4. *There exists $\beta^{(\infty)}(\varphi, x)$ such that $(\varphi, x) \mapsto (\varphi, x + \beta^{(\infty)}(\varphi, x))$ is a diffeomorphism of the torus $\mathbb{T}^{\nu+1}$ with the following estimates (recall (7.3)),*

$$\begin{aligned} &\|\beta^{(\infty)}\|^{\gamma,\mathcal{O}^{2\gamma}_\infty}_s \lesssim_s \gamma^{-1}\|a_5\|^{\gamma,\mathcal{O}_0}_{s+2\tau+4}, \quad \forall s \geq s_0,\\ &\|\Delta_{12}\beta^{(\infty)}\|_p \lesssim_p \varepsilon\gamma^{-1}(1 + \|\mathfrak{I}_\delta\|_{p+\widehat{\sigma}})\|i_1 - i_2\|_{p+\widehat{\sigma}}, \end{aligned} \tag{7.45}$$

for some $\widehat{\sigma} > 0$, and the following holds. If β is the function such that $(\varphi, x) \mapsto (\varphi, x + \beta(\varphi, x))$ is the inverse of the above diffeomorphism and $\mathcal{B}_6$ is the flow of the Hamiltonian PDE

$$u_\tau = \Pi_S^\perp[\big(J\circ b(\tau)\big)\,u)], \qquad b(\tau) := b(\tau,\varphi,x) = \frac{\beta}{1+\tau\beta_x}, \tag{7.46}$$

then the conjugated of the operator $\mathcal{L}_5$ in (7.16) with $i = 5$ is

$$\mathcal{L}_6 := \mathcal{B}_6\,\mathcal{L}_5\,\mathcal{B}_6^{-1} = \Pi_S^\perp\Big(\omega\cdot\partial_\varphi - mJ + \mathcal{Q}_6\Big), \tag{7.47}$$

where $\mathcal{Q}_6 = \mathrm{Op}(\mathfrak{q}_6) + \widehat{\mathcal{Q}}_6$ is of order -1, as in Proposition C.2, and m is a constant such that

$$\begin{aligned} &|m - 1 - \varepsilon^2 c(\omega) - \varepsilon^4 d(\omega)|^\gamma \lesssim \varepsilon^{10}\gamma^{-2}, \quad |m|^{lip} \lesssim 1,\\ &|\Delta_{12} m| \lesssim \varepsilon\|i_1 - i_2\|_{s_0+2}, \quad \forall\omega\in\mathcal{O}^{2\gamma}_\infty. \end{aligned} \tag{7.48}$$

Moreover, for any $s \geq s_0$,

$$|\mathfrak{q}_6|^{\gamma,\mathcal{O}_\infty^{2\gamma}}_{-1,s,\alpha} \lesssim_s \varepsilon(1+\|\mathfrak{I}_\delta\|^{\gamma,\mathcal{O}_0}_{s+\widehat{\sigma}}), \qquad |\Delta_{12}\mathfrak{q}_6|_{-1,p,\alpha} \lesssim_p \varepsilon\gamma^{-1}(1+\|\mathfrak{I}_\delta\|_{p+\widehat{\sigma}})\|i_1-i_2\|_{p+\widehat{\sigma}}, \tag{7.49}$$

and $\widehat{\mathcal{Q}}_6 \in \mathfrak{L}_{\rho,p}$, for $s_0 \leq s \leq \mathcal{S}$, satisfies

$$\mathbb{M}^{\gamma}_{\widehat{\mathcal{Q}}_6}(s,\mathfrak{b}) \lesssim_s \varepsilon(1+\|\mathfrak{I}_\delta\|^{\gamma,\mathcal{O}_0}_{s+\widehat{\sigma}}), \quad 0 \leq \mathfrak{b} \leq \rho-2, \tag{7.50}$$

$$\mathbb{M}_{\Delta_{12}\widehat{\mathcal{Q}}_6}(p,\mathfrak{b}) \lesssim_p \varepsilon\gamma^{-1}(1+\|\mathfrak{I}_\delta\|_{p+\widehat{\sigma}})\|i_1-i_2\|_{p+\widehat{\sigma}}, \quad 0 \leq \mathfrak{b} \leq \rho-3 \tag{7.51}$$

with $\widehat{\sigma} = \sigma_0+\sigma_9+\rho+s_1-s_0$ for some σ_9, possibly larger than σ_8 (recall (7.16) with $i=5$ and s_1 given in Proposition 3.6 in [33]).

Proof. The first order linear differential operator (recall (7.36), (7.40))

$$\omega\cdot\partial_\varphi - \big(1+\varepsilon^2 c(\omega)+\varepsilon^4 d(\xi)+a_5(\varphi,x)\big)\partial_x \tag{7.52}$$

defined on $H^s_{S^\perp}(\mathbb{T}^{\nu+1})$ is associated to the vector field on $\mathbb{T}^{\nu+1}$

$$X_0 := \omega\cdot\frac{\partial}{\partial\varphi} - \big(1+\varepsilon^2 c(\omega)+\varepsilon^4 d(\xi)+a_5(\varphi,x)\big)\frac{\partial}{\partial x}. \tag{7.53}$$

For $\mathfrak{p}_1$ in (6.12) large enough, i.e. if $\mathfrak{p}_1 \gg \sigma_0+s_1+\sigma_8$, and by (7.17) with $i=5$, we have

$$C(s_1)\,\gamma^{-1}\|a_5\|^{\gamma,\mathcal{O}_\infty^{2\gamma}}_{s_1} \leq C(s_1)\varepsilon^{4-3a} = \delta^* \ll 1,$$

provided that ε is small enough. This is the condition (7.12), hence Proposition 3.6 in [33] applies to the vector field (7.53). Thus there exist $\beta^{(\infty)}$ and m such that the bounds (7.45), (7.48) hold. In particular the second bound in (7.45) follows by Lemma 3.7 in [33]. Moreover the operator (7.52) conjugated by the transformation

$$\mathcal{T}_{\beta^{(\infty)}}: u(\varphi,x) \mapsto u(\varphi, x+\beta^{(\infty)}(\varphi,x)),$$

is associated to the vector field

$$(\mathcal{T}_{\beta^{(\infty)}})_* X_0 = \omega\cdot\frac{\partial}{\partial\varphi} + (\mathcal{T}_{\beta^{(\infty)}})^{-1}$$
$$\Big(\omega\cdot\partial_\varphi\beta^{(\infty)} - (1+\varepsilon^2 c(\omega)+\varepsilon^4 d(\xi)+a_5(\varphi,x)(1+\beta^{(\infty)}_x)\Big)\frac{\partial}{\partial x},$$

and by Proposition 3.6 in [33] we have that

$$\omega\cdot\partial_\varphi\beta^{(\infty)} - (1+\varepsilon^2 c(\omega)+\varepsilon^4 d(\omega)+a_5(\varphi,x))(1+\beta^{(\infty)}_x) = -m. \tag{7.54}$$

By Lemma 11.4 in [1] the function $\beta^{(\infty)}$ satisfies the bound (7.45). By (7.45) and $\mathfrak{p}_1$ large enough, for ε small enough, the function $\beta^{(\infty)}$ satisfies the smallness condition of Lemma C.1, indeed

$$\|\beta^{(\infty)}\|^{\gamma,\mathcal{O}_0}_{s_0+\sigma_1} \leq C(s_1)\gamma^{-1}\|a_5\|^{\gamma,\mathcal{O}_\infty^{2\gamma}}_{s_0+\sigma_1+2\tau+4} \overset{(7.17)}{\leq} C(s_1)\,\gamma^{-1}\big(\varepsilon^6+\varepsilon\|\mathfrak{I}_\delta\|^{\gamma,\mathcal{O}_0}_{s_0+\mathfrak{p}_1}\big)$$
$$\overset{(6.12)}{\leq} C(s_1)\varepsilon^{4-3a}. \tag{7.55}$$

Hence $\mathcal{B}_6$ is well defined. By (7.17), (7.18), (7.19), $i=5$, the bounds (C.13), (C.14) hold with $\mathtt{k}_1 \rightsquigarrow \varepsilon^6$, $\mathtt{k}_2 \rightsquigarrow \varepsilon$, $\mathtt{k}_3 \rightsquigarrow \varepsilon\gamma^{-1}$ and Proposition C.2 applies and the thesis follows. □

Let us define

$$\mathcal{B} := \mathcal{B}_6 \circ \mathcal{B}_5 \circ \mathcal{B}_4 \circ \mathcal{B}_3 \circ \mathcal{B}_2 \circ \mathcal{B}_1 . \tag{7.56}$$

Then the Hamiltonian of the operator $\mathcal{L}_6$ is (recall (6.31), (7.43) and (7.47))

$$\mathcal{K} := \mathsf{H} \circ \mathcal{B}^{-1} = \mathsf{H}_0 + \varepsilon \mathcal{K}_1 + \varepsilon^2 \Big(\mathtt{Z}_0 + \mathcal{K}_2 \Big) + \varepsilon^3 \mathcal{K}_3 + o(\varepsilon^3). \tag{7.57}$$

Notice also that $\{\mathtt{H}_0, \mathtt{Z}_0\}_e = 0$. The expansion (7.57) allows us, together with Remark 7.3, to give a more precise expression of the remainder $\mathcal{Q}_6$ in (7.47). This is the content of Lemma 7.6 in the next section.

7.2. Linear Birkhoff normal form. The aim of this section is to eliminate $\mathcal{K}_1$, $\mathcal{K}_3$ and normalize the Hamiltonian $\mathcal{K}_2$ from (7.57). Our first point is that the -1 smoothing remainder $o(\varepsilon^3)$ belongs to a special class of operators defined in Definition C.6 and denoted by $\mathfrak{C}_{-1}$. It turns out that this class is preserved under the changes of variables used in the linear Birkhoff normal form procedure (see Lemmata C.8, C.9).

Remark 7.5. In the following steps of linear Birkhoff normal form we shall use the relation

$$\sum_{i=1}^{\nu} \overline{\jmath}_i \ell_i + j' - j = 0, \quad \text{if } \ |\ell| \leq 3, \quad \forall j,\ j' \in S^c, \tag{7.58}$$

which holds by the conservation of momentum.

Lemma 7.6. *Recall* (7.47). *We have*

$$\mathcal{L}_6 = \Pi_S^{\perp} \big(\omega \cdot \partial_\varphi - mJ - \varepsilon X_{\mathcal{K}_1} - \varepsilon^2 X_{\mathcal{K}_2} - \varepsilon^3 X_{\mathcal{K}_3} + \mathfrak{R} \big) \tag{7.59}$$

where $X_{\mathcal{K}_i} := J \nabla \mathcal{K}_i$, $i = 1, 2, 3$, *are almost diagonal and in* $\mathfrak{C}_{-1}(\mathcal{O}_\infty^{2\gamma})$ *(recall Definitions 6.4 and C.6) satisfying*

$$\mathbb{B}^{\gamma}_{\varepsilon^k J \nabla \mathcal{K}_k}(s) \leq \varepsilon^k C(s), \quad k = 1, 2, 3. \tag{7.60}$$

The remainder $\mathfrak{R}$ *belongs to* $\mathfrak{C}_{-1}(\mathcal{O}_\infty^{2\gamma})$ *and satisfies*

$$\mathfrak{R} := \mathcal{Q}_6 + \varepsilon X_{\mathcal{K}_1} + \varepsilon^2 X_{\mathcal{K}_2} + \varepsilon^3 X_{\mathcal{K}_3} . \tag{7.61}$$

$$\mathbb{B}^{\gamma}_{\mathfrak{R}}(s) \lesssim_s \varepsilon^{4-3a} + \varepsilon \gamma^{-1} \|\mathfrak{I}_\delta\|^{\gamma, \mathcal{O}_0}_{s+\widehat{\sigma}}, \qquad \mathbb{B}_{\Delta_{12}\mathfrak{R}}(s_0) \lesssim \varepsilon \gamma^{-1} \|i_1 - i_2\|_{s_0+\widehat{\sigma}}, \tag{7.62}$$

for $\widehat{\sigma}$ *given in Proposition 7.4.*

Proof. By the discussion of Sect. 7.1 $\mathcal{K}_i$, $i = 1, 2, 3$, are of the form (6.35) with $f_i = 0$ for $i = 1, 2, 3$. Hence the vector field $X_{\mathcal{K}_i}$ are pseudo differential of order -1 up to a finite rank term. In addition, they are almost diagonal by (6.34) and the momentum condition (7.58). By Lemma C.8-(ii) $\varepsilon X_{\mathcal{K}_1}$, $\varepsilon^2 X_{\mathcal{K}_2}$, $\varepsilon^3 X_{\mathcal{K}_3}$ belong to $\mathfrak{C}_{-1}$ and, by (7.27), (7.26), (7.35), (7.37), (7.41), satisfy (7.60). By Proposition 7.4, the choice of ρ as in (7.15) and by Lemma C.8-(i) taking $p = s_0$ and $\mathfrak{p}_1$ large enough, $\mathcal{Q}_6 \in \mathfrak{C}_{-1}$. Thus $\mathfrak{R} \in \mathfrak{C}_{-1}$.

Note that only $\mathcal{Q}_6$ in (7.61) depend on the torus embedding i_δ, then the second bound in (7.62) follows by Lemma C.8-(i), (ii), (7.49) and (7.51). To prove the first bound in (7.62) we reason as follows.

By (7.45) and (7.17) with $i = 5$ we have that

$$\|\beta^{(\infty)}\|_s^{\gamma,\mathcal{O}_\infty^{2\gamma}} \lesssim_s \varepsilon^6\gamma^{-1} + \varepsilon\gamma^{-1}\|\mathfrak{I}_\delta\|_{s+2\tau+4+\sigma_0+\sigma_8}^{\gamma,\mathcal{O}_0}, \quad \|\beta^{(\infty)}\|_{s_0}^{\gamma,\mathcal{O}_\infty^{2\gamma}} \lesssim \varepsilon^{4-3a}. \tag{7.63}$$

Then the map $\mathcal{B}_6$ leaves invariant (using Remark 7.2) the terms of size $\varepsilon, \varepsilon^2, \varepsilon^3$ in $\mathcal{L}_5$, and hence, by Remark 7.3, those terms in $\mathcal{Q}_6$ are given by $-\varepsilon J\nabla\mathcal{K}_1$, $-\varepsilon^2 J\nabla\mathcal{K}_2$, $-\varepsilon^3 J\nabla\mathcal{K}_3$.

From the proof of the bounds (C.18), (C.20) in Proposition C.2 one can notice that the operators $\mathrm{Op}(\mathfrak{q}_6)$ and $\widehat{\mathcal{Q}}_6$ admit a "formal" expansion in $\beta^{(\infty)}$ (by expanding the flow in τ). Of course, by the discussion above, the biggest term in $\mathfrak{R}$ are the ones which are linear in $\beta^{(\infty)}$. Such term comes from the conjugation of $\mathcal{L}_5$ under the map $\mathcal{B}_6$, more precisely from the conjugation of

$$(J - \partial_x) \circ (1 + \varepsilon^2 c(\omega) + \varepsilon^4 d(\omega) + a_5(\varphi, x)).$$

We refer to the formula (3.11) in Proposition 3.1 of [33] to see the term bounded by the norm of $\beta^{(\infty)}$. Comparing the bounds (7.63) and (7.42) one can deduce the first bound in (7.62). □

In order to normalize the vector fields $\varepsilon^i J\nabla\mathcal{K}_i$ we will look for changes of coordinates Υ_i generated as one-time flow of quadratic Hamiltonians $H_{\mathtt{A}_i}$ described by almost diagonal matrices $\mathtt{A}_i$ (see (C.47), (C.48) ,(C.49), (C.50)). We remark that the Hamiltonian $\varepsilon^2\mathtt{Z}_0$ is left invariant by these changes of coordinates, since $\{\mathtt{H}_0, \mathtt{Z}_0\}_e = 0$. At any step of the procedure we shall verify that $J\mathtt{B}_i$ (see (C.49), (C.50)) are almost diagonal and belong to $\mathfrak{C}_{-1}$ in order to apply Lemma C.13, which guarantees well-posedness and tame estimates of Υ_i.

Step one (order ε) At this step we want to eliminate $\varepsilon X_{\mathcal{K}_1}$ from (7.59). We have

$$\mathcal{K}^{(1)} := \mathcal{K} \circ \Upsilon_1^{-1} = \mathsf{H}_0 + \varepsilon\mathcal{K}_1^{(1)} + \varepsilon^2\Big(\mathtt{Z}_0 + \mathcal{K}_2^{(1)}\Big) + \varepsilon^3\mathcal{K}_3^{(1)} + o(\varepsilon^3), \tag{7.64}$$

$$\begin{aligned}\mathcal{K}_1^{(1)} &:= \mathcal{K}_1 + \{H_{\mathtt{A}_1}, \mathsf{H}_0\}_e = \mathcal{K}_1 + \overline{\omega}\cdot\partial_\varphi H_{\mathtt{A}_1} + \{H_{\mathtt{A}_1}, \mathsf{H}_0\},\\ \mathcal{K}_2^{(1)} &:= \mathcal{K}_2 + \frac{1}{2}\{H_{\mathtt{A}_1}, \{H_{\mathtt{A}_1}, \mathsf{H}_0\}_e\}_e + \{H_{\mathtt{A}_1}, \mathcal{K}_1\}_e\\ &= \mathcal{K}_2 + \frac{1}{2}\{H_{\mathtt{A}_1}, \mathcal{K}_1^{(1)}\} + \frac{1}{2}\{H_{\mathtt{A}_1}, \mathcal{K}_1\}.\end{aligned} \tag{7.65}$$

We choose $\mathtt{A}_1$ such that

$$\mathcal{K}_1^{(1)} = \overline{\omega}\cdot\partial_\varphi H_{\mathtt{A}_1} + \{H_{\mathtt{A}_1}, \mathsf{H}_0\} + \mathcal{K}_1 = 0. \tag{7.66}$$

Recalling that $\mathcal{K}_1 := \mathsf{H}_1^{(1)}$, we have (see (7.27), (7.28))

$$\mathcal{K}_1(u) = \sum_{j,j'\in S^c} (\mathfrak{B}_1)_j^{j'}(\varphi)\, u_{j'}\, u_{-j}. \tag{7.67}$$

Then we choose $\mathtt{B}_1 = \mathfrak{B}_1$ in (C.49). By recalling the definition of $\mathfrak{B}_1$ in (7.27) it is easy to see that $J\mathtt{B}_1 \in \mathfrak{C}_{-1}$, since it is a pseudo differential operator of order -1. Moreover it is almost diagonal because $J, 3\Lambda\partial_x$ are diagonal operators and β_1 is a function supported on the finite set S.

Given $X \in \mathfrak{C}_{-1}$ to shorten the notation in the following lemma we write (recall (2.3))

$$\mathrm{ad}_X[\cdot] := [X, \cdot]. \tag{7.68}$$

Under this notation we have the following lemma.

Lemma 7.7. *The transformed operator is (recall (7.59), (7.64))*

$$\mathcal{L}_7 h := \Upsilon_1 \mathcal{L}_6 \Upsilon_1^{-1} = \Pi_S^{\perp}\Big(\omega\cdot\partial_\varphi - mJ - \varepsilon^2 X_{\mathcal{K}_2^{(1)}} - \varepsilon^3 X_{\mathcal{K}_3^{(1)}} + \mathcal{R}_7\Big) \tag{7.69}$$

where

$$X_{\mathcal{K}_2^{(1)}} := J\nabla\mathcal{K}_2^{(1)} = X_{\mathcal{K}_2} + \mathrm{ad}_{X_{\mathbb{A}_1}}[X_{\mathcal{K}_1}] + \frac{1}{2}\mathrm{ad}^2_{X_{\mathbb{A}_1}}[\omega\cdot\partial_\varphi - mJ], \tag{7.70}$$

$$X_{\mathcal{K}_3^{(1)}} := J\nabla\mathcal{K}_3^{(1)} = X_{\mathcal{K}_3} + \mathrm{ad}_{X_{\mathbb{A}_1}}[X_{\mathcal{K}_2}] + \frac{1}{2}\mathrm{ad}^2_{X_{\mathbb{A}_1}}[X_{\mathcal{K}_1}] + \frac{1}{6}\mathrm{ad}^3_{X_{\mathbb{A}_1}}[\omega\cdot\partial_\varphi - mJ], \tag{7.71}$$

the operator $\mathcal{R}_7 \in \mathfrak{C}_{-1}$ with

$$\mathbb{B}^{\gamma}_{\mathcal{R}_7}(s) \lesssim_s \varepsilon^{4-3a} + \varepsilon\gamma^{-1}\|\mathfrak{I}_\delta\|^{\gamma,\mathcal{O}_0}_{s+\tilde\sigma}, \quad \mathbb{B}_{\Delta_{12}\mathcal{R}_7}(s_0) \lesssim \varepsilon\gamma^{-1}(1+\|\mathfrak{I}_\delta\|_{s_0+\tilde\sigma})\|i_1 - i_2\|_{s_0+\tilde\sigma}, \tag{7.72}$$

for some $\tilde\sigma \geq \widehat{\sigma}$ (recall the loss of regularity in (7.49), (7.50), (7.51)).

Proof. By using (C.58) we have that formulæ (7.70), (7.71) hold and that

$$\begin{aligned}\mathcal{R}_7 :=\ & \mathfrak{R} + \varepsilon\,\mathrm{ad}_{X_{\mathbb{A}_1}}[\varepsilon^3 X_{\mathcal{K}_3} + \mathfrak{R}] + \frac{\varepsilon^2}{2}\mathrm{ad}^2_{X_{\mathbb{A}_1}}[-\varepsilon^2 X_{\mathcal{K}_2} - \varepsilon^3 X_{\mathcal{K}_3} + \mathfrak{R}] \\ & + \frac{\varepsilon^3}{6}\mathrm{ad}^3_{X_{\mathbb{A}_1}}[-\varepsilon X_{\mathcal{K}_1} - \varepsilon^2 X_{\mathcal{K}_2} - \varepsilon^3 X_{\mathcal{K}_3} + \mathfrak{R}] + \sum_{k\geq 4}\frac{\varepsilon^k}{k!}\mathrm{ad}^k_{X_{\mathbb{A}_1}}[\mathcal{L}_6].\end{aligned} \tag{7.73}$$

For $Y \in \mathfrak{C}_{-1}$ define $\mathcal{Z}_n := \sum_{k\geq n}\frac{\varepsilon^k}{k!}\mathrm{ad}^k_{X_{\mathbb{A}_1}}[Y]$ for any $n \geq 1$. By Lemma C.9, and using (C.54), we deduce that

$$\begin{aligned}\mathbb{B}^{\gamma}_{\mathcal{Z}_n}(s) &\lesssim_s C(s,n)\mathbb{B}^{\gamma}_{Y}(s_0) + C(s_0,n)\mathbb{B}^{\gamma}_{Y}(s) \\ \mathbb{B}_{\Delta_{12}\mathcal{Z}_n}(s_0) &\lesssim C(s_0,n)\Big(\mathbb{B}_Y(s_0) + \mathbb{B}_{\Delta_{12}Y}(s_0)\Big)\end{aligned} \tag{7.74}$$

for ε small enough. In (7.73) there are terms of the form $\mathrm{ad}^k_{X_{\mathbb{A}_1}}[Y]$, for some $k \geq 1$, with $Y = X_{\mathcal{K}_1}, X_{\mathcal{K}_2}, X_{\mathcal{K}_3}, \mathfrak{R}$ which belong to $\mathfrak{C}_{-1}$ by Lemmata C.8 and 7.6. We note that by (7.66)

$$\mathrm{ad}_{X_{\mathbb{A}_1}}[\omega\cdot\partial_\varphi - mJ] = -\varepsilon X_{\mathcal{K}_1} - (\omega-\overline{\omega})\cdot\partial_\varphi X_{\mathbb{A}_1} - (m-1)[X_{\mathbb{A}_1}, J] \in \mathfrak{C}_{-1} \tag{7.75}$$

since $\mathbb{A}_1$ is almost diagonal. Hence $(\omega-\overline{\omega})\cdot\partial_\varphi X_{\mathbb{A}_1}, [X_{\mathbb{A}_1}, J] \in \mathfrak{C}_{-1}$ (see the proof of Lemma C.13) and by Lemma C.8-(iii) the remainder $\mathcal{R}_7 \in \mathfrak{C}_{-1}$. By (7.73), (C.43), (C.44), (7.74), (7.75), (7.60), (7.62) and the fact that $|\omega - \overline{\omega}| \lesssim \varepsilon^2$ we get the bounds (7.72). □

Step two (order ε^2) At this step we want to normalize $\varepsilon^2 X_{\mathcal{K}_2^{(1)}}$ from (7.69). We have

$$\mathcal{K}^{(2)} := \mathcal{K}^{(1)}\circ\Upsilon_2^{-1} = \mathsf{H}_0 + \varepsilon^2\Big(\mathtt{Z}_0 + \mathcal{K}_2^{(2)}\Big) + \varepsilon^3\mathcal{K}_3^{(2)} + o(\varepsilon^3), \qquad \mathcal{K}_2^{(2)} := \{H_{\mathbb{A}_2}, \mathsf{H}_0\}_e + \mathcal{K}_2^{(1)}, \tag{7.76}$$

where $\mathcal{K}_2^{(1)}$ is given in (7.65) (see also (7.70)). We choose $\mathtt{A}_2$ in order to solve the following equation

$$\overline{\omega}\cdot\partial_\varphi H_{\mathtt{A}_2}+\{H_{\mathtt{A}_2},\mathsf{H}_0\}=\mathcal{K}_2^{(1)}-\Pi_{\mathrm{Ker}(\mathsf{H}_0)}\mathcal{K}_2^{(1)}. \tag{7.77}$$

Hence we choose $\mathtt{B}_2=\nabla\Pi_{\mathrm{Rg}(\mathsf{H}_0)}\mathcal{K}_2^{(1)}(u)$ in (C.49). Note that $X_{\mathcal{K}_2}$ is pseudo differential of order -1 and $J\mathtt{A}_1$, $X_{\mathcal{K}_1}$ belong to $\mathfrak{C}_{-1}$ and so also their Poisson brackets. Hence $J\mathtt{B}_2\in\mathfrak{C}_{-1}$. By Remark 6.5 we have that $J\mathtt{B}_2$ is also almost diagonal.

In order to perform the third step in the linear BNF we need to explicitly compute the corrections $O(\varepsilon^2)$ coming from $\Pi_{\mathrm{Ker}(\mathsf{H}_0)}\mathcal{K}_2^{(1)}$. The point is that a priori, it is not clear whether the resonant terms $\Pi_{\mathrm{Ker}(\mathsf{H}_0)}\mathcal{K}_2^{(1)}$ are supported only on trivial resonances. Our approach is then to show that the normal form we obtain must necessarily coincide with the formal one, which is relatively easy to compute.

Definition 7.8. Recalling the notations used in Sect. 3, we denote by $\Pi^{d_z\leq k}$, respectively $\Pi^{d_z=k}$, the projector of a homogenous Hamiltonian of degree n on the monomials with degree less or equal than k, respectively equal k, in the normal variable z, i.e.

$$\Pi^{d_z\leq k}H^{(n)}:=H^{(n,\leq k)},\quad \Pi^{d_z=k}H^{(n)}:=H^{(n,k)}.$$

We denote by Π_{triv} the projection onto trivial resonances (of the form (3.11)), i.e. monomials of the form

$$u_ju_{-j}u_iu_{-i}\dots u_ku_{-k}.$$

The following proposition allows to easily compute the resonant terms $\Pi_{\mathrm{Ker}(\mathsf{H}_0)}\mathcal{K}_2^{(1)}$ in (B.17).

Theorem 7.9 (Normal form identification). *Consider the symplectic change of coordinates A_ε in (4.7). Then*

$$\Pi^{d_z=2}\Pi_{\mathrm{Ker}(\mathtt{H}_0)}\Big(\mathtt{Z}_0+\mathcal{K}_2^{(1)}\Big)=\Big[\Pi_{\mathrm{triv}}\Pi^{d_z=2}\Big(\frac{1}{2}\{\mathfrak{F}^{(3)},H^{(3)}\}\Big)\Big]\circ A_{1|_{\substack{y=0\\ \theta=\varphi}}}, \tag{7.78}$$

where $A_1:=A_{\varepsilon_{|_{\varepsilon=1}}}$, $\mathtt{H}_0$ is in (6.32) and we set (recalling (2.4)) $\mathfrak{F}^{(3)}:=[\mathrm{ad}_{H^{(2)}}]^{-1}H^{(3)}$ with $H^{(3)}$ in (3.1).

Proof. The proof is postponed to the "Appendix B". □

As a consequence of the identification above, we have, by (7.78), (3.21), (3.1),

$$\Big(\frac{c(\omega)}{2}\sum_{j\in S^c}|u_j|^2+\Pi_{\mathrm{Ker}(\mathsf{H}_0)}\mathcal{K}_2^{(1)}\Big)=\frac{1}{2}\sum_{j\in S^c}\mathfrak{l}_j|u_j|^2 \tag{7.79}$$

where (recall (1.8))

$$\begin{aligned}\mathfrak{l}_j&:=\sum_{j_2\in S}\frac{\lambda(j_2+j)}{\lambda(j_2)+\lambda(j)-\lambda(j_2+j)}\xi_{j_2}\\&=\frac{2}{3}\sum_{j_2\in S^+}\frac{(1+j_2^2)(1+j^2)(2+j_2^2+j^2)}{(3+j_2^2-j_2j+j^2)(3+j_2^2+j_2j+j^2)}\xi_{j_2}.\end{aligned} \tag{7.80}$$

We define the diagonal operator (recall (7.36))

$$\mathfrak{D} := \mathfrak{D}(\xi) = \text{diag}\ (\mathrm{i}\kappa_j)_{j\in S^c}, \qquad \kappa_j = \lambda(j)\left(\mathfrak{l}_j - c(\omega)\right) \in \mathbb{R}. \tag{7.81}$$

Lemma 7.10. *We have*

$$\kappa_j = \lambda(j)\left(\mathfrak{l}_j - c(\omega)\right), \quad |j|\,|\kappa_j| \leq C \qquad \forall j \in S^c, \tag{7.82}$$

for an appropriate constant $C > 0$ depending on the set S.

Proof. Recalling the definitions (7.36) and (7.80) we have, for $j \in S^c$,

$$\mathfrak{l}_j - c(\omega) = -\frac{2}{3}\sum_{j_0\in S^+} \frac{(1+j_0^2)(7+5j_0^2+j_0^4+3j^2)}{(3+j_0^2-j_0j+j^2)(3+j_0^2+j_0j+j^2)}\xi_{j_0} = \frac{P(j)}{Q(j)}. \tag{7.83}$$

It is easy to prove that $|\lambda(j)| \leq 4|j|$, $3+j_0^2+j^2 \pm j_0 j \geq \frac{3}{4}j^2$ and $(1+j_0^2)(7+5j_0^2+j_0^4+3j^2) \leq 14 j_0^6 j^2$. Hence $|\kappa(j)| \lesssim \frac{|j|}{j^2}\sum_{j_0\in S^+} j_0^6$. □

Lemma 7.11. *The transformed operator is (recall (7.69))*

$$\mathcal{L}_8 := \Upsilon_2\mathcal{L}_7\Upsilon_2^{-1} = \Pi_S^{\perp}\Big(\omega\cdot\partial_\varphi - mJ - \varepsilon^2\mathfrak{D}(\xi) - \varepsilon^3 X_{\mathcal{K}_3^{(2)}} + \mathcal{R}_8\Big) \tag{7.84}$$

where $\mathcal{K}_3^{(2)} = \mathcal{K}_3^{(1)}$, $\mathfrak{D}(\xi)$ is the diagonal operator of order -1 defined in (7.81), $\mathcal{R}_8 \in \mathfrak{C}_{-1}$ satisfies

$$\mathbb{B}^{\gamma}_{\mathcal{R}_8}(s) \lesssim_s \varepsilon^{4-3a} + \varepsilon\gamma^{-1}\|\mathfrak{I}_\delta\|^{\gamma,\mathcal{O}_0}_{s+\tilde\sigma}, \quad \mathbb{B}_{\Delta_{12}\mathcal{R}_8}(s_0) \lesssim \varepsilon\gamma^{-1}(1+\|\mathfrak{I}_\delta\|_{s_0+\tilde\sigma})\|i_1-i_2\|_{s_0+\tilde\sigma}, \tag{7.85}$$

for some $\tilde\sigma$ possibly larger than the one in Lemma 7.7.

Proof. The proof follows by using the same arguments of the proof of Lemma 7.7. In particular, expanding the left hand side of (7.84) using (C.58) we get

$$\mathcal{R}_8 := \mathcal{R}_7 + \varepsilon^2\mathrm{ad}_{X_{\mathtt{A}_2}}[-\varepsilon^2\mathfrak{D}(\xi) - \varepsilon^3 X_{\mathcal{K}_3^{(2)}} + \mathcal{R}_8] + \sum_{k\geq 2}\frac{\varepsilon^{2k}}{k!}\mathrm{ad}_{X_{\mathtt{A}_2}}[\mathcal{L}_7]. \tag{7.86}$$

By (7.77) and Theorem 7.9 we have that

$$\mathrm{ad}_{X_{\mathtt{A}_2}}[\omega\cdot\partial_\varphi - mJ] + \mathcal{K}_2^{(1)} = -\mathfrak{D}(\xi) - (\omega-\overline{\omega})\cdot\partial_\varphi X_{\mathtt{A}_2} - (m-1)[X_{\mathtt{A}_2}, J].$$

By (7.10) $\mathfrak{D}(\xi) \in \mathfrak{C}_{-1}$ and by the fact that $\mathtt{A}_2$ is almost diagonal we have that $(\omega-\overline{\omega})\cdot\partial_\varphi X_{\mathtt{A}_2}, [X_{\mathtt{A}_2}, J] \in \mathfrak{C}_{-1}$. Then $\mathcal{R}_8 \in \mathfrak{C}_{-1}$. The bounds (7.85) are obtained by using the estimates (C.43), (C.44), (7.74), (C.54) and (7.72). □

Step three (order ε^3) At this step we eliminate $\varepsilon^3 X_{\mathcal{K}_3^{(2)}}$ from (7.84). Recalling that $\mathcal{K}_3^{(2)}$ is given in Lemma 7.11, we have

$$\begin{aligned}\mathcal{K}^{(3)} &:= \mathcal{K}^{(2)} \circ \Upsilon_3^{-1} = \mathsf{H}_0 + \varepsilon^2 \mathcal{K}_2^{(2)} + \varepsilon^3 \mathcal{K}_3^{(3)} + o(\varepsilon^3),\\ \mathcal{K}_3^{(3)} &:= \{H_{\mathtt{A}_3}, \mathsf{H}_0 + \varepsilon^2 \mathbb{A}\xi\cdot\eta + \frac{\varepsilon^2}{2}\sum_{j\in S^c} \mathfrak{l}_j(\xi) z_j z_{-j}\}_e + \mathcal{K}_3^{(2)}. \end{aligned} \tag{7.87}$$

Note that we consider in the normal form also the ε^2-terms. We want to solve the equation

$$\mathcal{D}_{\overline{\omega}+\varepsilon^2\mathbb{A}\xi} H_{\mathtt{A}_3} + \{H_{\mathtt{A}_3}, \mathsf{H}_0 + \frac{\varepsilon^2}{2}\sum_{j\in S^c} \mathfrak{l}_j(\xi) z_j z_{-j}\} + \mathcal{K}_3^{(2)} = 0. \tag{7.88}$$

Hence we choose the matrix $\mathtt{B}_3 := \nabla \mathcal{K}_3^{(1)}(u)$ (note that $\mathcal{K}_3^{(1)} = \mathcal{K}_3^{(2)}$). Recalling (7.71) it is easy to see that $J\mathtt{B}_3$ is sum of Lie brackets of elements of $\mathfrak{C}_{-1}$, hence by Lemma C.8 it belongs to $\mathfrak{C}_{-1}$. By the fact that $\mathtt{A}_1$ is almost diagonal and by Remark 6.5 we have that $J\mathtt{B}_3$ is almost diagonal.

Lemma 7.12. *The transformed operator is (recall (7.69))*

$$\mathcal{L}_9 := \Upsilon_3 \mathcal{L}_8 \Upsilon_3^{-1} = \Pi_S^{\perp}\Big(\omega\cdot\partial_\varphi - mJ - \varepsilon^2 \mathfrak{D}(\xi) + \mathcal{R}_9\Big) \tag{7.89}$$

where $\mathfrak{D}(\xi)$ is the diagonal operator of order -1 defined in (7.81), $\mathcal{R}_9 \in \mathfrak{C}_{-1}$ satisfies

$$\mathbb{B}^{\gamma}_{\mathcal{R}_9}(s) \lesssim_s \varepsilon^{4-3a} + \varepsilon\gamma^{-1}\|\mathfrak{I}_\delta\|^{\gamma,\mathcal{O}_0}_{s+\tilde\sigma}, \qquad \mathbb{B}_{\Delta_{12}\mathcal{R}_9}(s_0) \lesssim \varepsilon\gamma^{-1}\|i_1 - i_2\|_{s_0+\tilde\sigma}, \tag{7.90}$$

for some $\tilde\sigma$ possibly larger than the one in Lemma 7.11.

Proof. The proof follows the same arguments used for proving Lemma 7.7. By (C.58) we deduce

$$\mathcal{R}_9 := \mathcal{R}_8 + \varepsilon^3 \mathrm{ad}_{X_{\mathtt{A}_3}}[-\varepsilon^3 X_{\mathcal{K}_3^{(2)}} + \mathcal{R}_8] + \sum_{k\geq 2}\frac{1}{k!}\mathrm{ad}^k_{X_{\mathtt{A}_3}}[\mathcal{L}_8]. \tag{7.91}$$

We note that by (7.88) we have (recall (4.5) and (7.36))

$$\begin{aligned}\mathrm{ad}_{X_{\mathtt{A}_3}}[\omega\cdot\partial_\varphi - mJ - \varepsilon^2\mathfrak{D}(\xi)] &= \varepsilon^3 X_{\mathcal{K}_3^{(2)}} + (\omega - \overline{\omega} - \varepsilon^2\mathbb{A}\xi)\cdot\partial_\varphi \mathtt{A}_3\\ &\quad -(m - 1 - \varepsilon^2 c(\omega))[X_{\mathtt{A}_3}, J] \in \mathfrak{C}_{-1},\end{aligned}$$

since $\mathtt{A}_3$ is almost diagonal . Hence the bounds (7.90) follows by (C.54), (7.85) and by using Lemma C.9. □

Proof of Theorem 7.1. We choose $\mu_1 = \tilde\sigma$ given in Lemma 7.12. We consider p and $\mathfrak{p}_1$ so that

$$s_0 + \mathfrak{p}_1 - \widehat{\sigma} \geq p \geq s_0, \quad \sigma_9 + \sigma_0 + (s_1 - s_0) + \sigma_1 + \rho + 1 \leq \tilde\sigma, \qquad \tilde\sigma \leq \mathfrak{p}_1, \tag{7.92}$$

where $\tilde\sigma$ is the loss of regularity in Lemma 7.12, σ_0 has been introduced in Sect. 7, see estimates (6.17)–(6.22), $\sigma_1 > 0$ and s_1 are given respectively in Lemma C.1 and in Proposition 3.6 in [33].

We define the map (recall (C.48), (7.56))

$$\Upsilon := \Upsilon_3 \circ \Upsilon_2 \circ \Upsilon_1 \circ \mathcal{B}.$$

By Proposition 7.4 the map $\mathcal{B}$ is defined for $\omega \in \mathcal{O}_\infty^{2\gamma}$ (see (7.3)), and so also Υ. By (7.26), (7.37), (7.41), (C.55), (C.54), (7.45), (7.17) (with $i = 5$) and Lemma C.1 we have (7.9).

The result follows by setting $\mathcal{L} := \mathcal{L}_9$ (see (7.89)), $\mathcal{P}_0 := \mathcal{R}_9$, and m is the constant given by Proposition 7.4. Indeed (7.48) implies (7.2) and (7.6). Moreover, by Lemma 7.12, we have that $\mathcal{P}_0 \in \mathfrak{C}_{-1}$ and satisfies (7.90). Lemma C.10 implies that $\mathcal{P}_0$ is -1-modulo tame together with $\mathfrak{b}_0$ derivatives in the variable φ (recall the Definition C.32 and the fact that $\gamma^{3/2} < \gamma$). The bounds (7.7), (7.8) follow from (C.45)–(C.46) , the definition of $\mathbb{B}^\gamma_{\mathcal{P}_0}(s)$ (see (C.32),(C.34)) and (7.90). By Lemma 7.10 we deduce (7.5). □

7.3. KAM reducibilty and Inversion of the linearized operator. In this subsection we prove the claim (6.9) by diagonalizing the operator $\mathcal{L}$ in (7.4). We first write

$$\mathcal{L} := \omega \cdot \partial_\varphi - \mathbf{M}_0, \quad \mathbf{M}_0 := \mathcal{D}_0 + \mathcal{P}_0, \quad \mathcal{D}_0 := \mathrm{diag}(\mathrm{i}\, d_j^{(0)})_{j \in S^c},$$
$$d_j^{(0)} := d_j^{(0)}(\omega) = m(\omega)\lambda(j) + \varepsilon^2 \kappa_j(\omega). \tag{7.93}$$

Notice that (by the smallness condition (7.94)) Proposition 4.1 in [33] applies to the operator $\mathcal{L}$ in (7.4). Hence by following almost word by word the proof of Theorem 1.7 in [33] one has the following.

Theorem 7.13 (Reducibility). *Fix $\tilde{\gamma} \in [\gamma^{3/2}/4, 4\gamma^{3/2}]$. Assume that $\omega \mapsto i_\delta(\omega)$ is a Lipschitz function defined on $\mathcal{O}_0 \subseteq \Omega_\varepsilon$ (recall (6.3)), satisfying (6.12) with $\mathfrak{p}_1 \geq \mu_1$ where $\mu_1 := \mu_1(\nu)$ is given in Proposition 7.1. There exist $\delta_0 \in (0,1)$, $N_0 > 0$, $C_0 > 0$, such that, if (recall that by (4.8) $\gamma = \varepsilon^{2+a}$)*

$$N_0^{C_0} \varepsilon^{4-3a} \gamma^{-3/2} = N_0^{C_0} \varepsilon^{1-(9/2)a} \leq \delta_0, \tag{7.94}$$

then the following holds.

(i) ***(Eigenvalues)*** *For all $\omega \in \Omega_\varepsilon$ there exists a sequence*[4]

$$d_j^\infty(\omega) := d_j^\infty(\omega, i_\delta(\omega)) := m(\omega, i_\delta(\omega))\,\lambda(j) + \varepsilon^2 \kappa_j(\omega) + r_j^\infty(\omega, i_\delta(\omega)), \quad j \in S^c, \tag{7.95}$$

with m and κ_j in (7.48) and (7.81) respectively. Furthermore, for all $j \in S^c$

$$\sup_j \langle j \rangle |r_j^{(\infty)}| \gamma^{3/2} \lesssim \varepsilon^{4-3a}, \qquad r_j^\infty = -r_{-j}^\infty. \tag{7.96}$$

All the eigenvalues $\mathrm{i} d_j^\infty$ are purely imaginary.

[4] Whenever it is not strictly necessary, we shall drop the dependence on i_δ.

(ii) ***(Conjugacy)*** *For all ω in the set*

$$\begin{aligned}\Omega_\infty^{2\tilde\gamma} &:= \Omega_\infty^{2\tilde\gamma}(i_\delta)\\ &:= \left\{\omega \in \mathcal{O}_\infty^{2\gamma} : |\omega\cdot\ell + d_j^\infty(\omega) - d_k^\infty(\omega)| \geq \frac{2\tilde\gamma}{\langle\ell\rangle^\tau},\ \forall \ell\in\mathbb{Z}^\nu,\ \forall j,k\in S^c,\ j\neq k\right\}\end{aligned} \tag{7.97}$$

there is a real, bounded, invertible, linear operator $\Phi_\infty(\omega)\colon H^s_{S^\perp}(\mathbb{T}^{\nu+1}) \to H^s_{S^\perp}(\mathbb{T}^{\nu+1})$, with bounded inverse $\Phi_\infty^{-1}(\omega)$, that conjugates $\mathcal{L}$ in (7.4) to constant coefficients, namely

$$\begin{aligned}&\mathcal{L}_\infty(\omega) := \Phi_\infty(\omega)\circ\mathcal{L}\circ\Phi_\infty^{-1}(\omega) = \omega\cdot\partial_\varphi - \mathcal{D}_\infty(\omega),\\ &\mathcal{D}_\infty(\omega) := \mathrm{diag}_{j\in S^c}\{\mathrm{i}d_j^\infty(\omega)\}.\end{aligned} \tag{7.98}$$

The transformations Φ_∞, Φ_∞^{-1} are tame and they satisfy for $s_0 \leq s \leq \mathcal{S}$ (recall μ_1 in Theorem 7.1)

$$\begin{aligned}\|(\Phi_\infty^{\pm1} - \mathrm{I})h\|_s^{\gamma^{3/2},\Omega_\infty^{2\tilde\gamma}} \lesssim_s &\left(\varepsilon^{4-3a}\gamma^{-3/2} + \varepsilon\gamma^{-5/2}\|\mathfrak{I}_\delta\|_{s+\mu_1}^{\gamma,\mathcal{O}_0}\right)\|h\|_{s_0}^{\gamma^{3/2},\Omega_\infty^{2\tilde\gamma}}\\ &+ \varepsilon^{4-3a}\gamma^{-3/2}\|h\|_s^{\gamma^{3/2},\Omega_\infty^{\tilde\gamma}}.\end{aligned} \tag{7.99}$$

Moreover Φ_∞, Φ_∞^{-1} are symplectic, and $\mathcal{L}_\infty$ is a Hamiltonian operator.

(iii) ***(Dependence on $i_\delta(\omega)$)*** *Let $i_1(\omega)$ and $i_2(\omega)$ be two Lipschitz maps satisfying (6.12) with $\mathfrak{I}_\delta \rightsquigarrow i_k(\varphi) - (\varphi,0,0)$, $k=1,2$, and such that*

$$\|i_1 - i_2\|_{s_0+\mu_1} \lesssim \rho N^{-(\tau+1)} \tag{7.100}$$

for N sufficiently large and $0 < \rho < \gamma^{3/2}/4$. Fix $\gamma_1 \in [\gamma^{3/2}/2, 2\gamma^{3/2}]$ and $\gamma_2 := \gamma_1 - \rho$. Let $r_j^{(\infty)}(\omega, i_k(\omega))$ be the sequence in (7.95) with $\tilde\gamma \rightsquigarrow \gamma_k$ for $k=1,2$. Then for all $\omega\in\Omega_\infty^{\gamma_1}(i_1)$ we have, for some $\kappa > (3/2)\tau$,

$$\gamma^{-1}|\Delta_{12}m| + \sup_j \langle j\rangle|\Delta_{12}r_j^{(\infty)}| \leq \varepsilon\gamma^{-1}\|i_1 - i_2\|_{s_0+\mu_1} + \varepsilon^{4-3a}N^{-\kappa}. \tag{7.101}$$

Proof. The proof of items (i)–(ii) follow by Theorem 1.9 in [33]. The only point left to prove is item (iii). We apply Theorems 1.4, 1.5 in [33]. We have that

$$\gamma^{-1}|m^{(N)}(i_2) - m(i_1)| + \langle j\rangle|r_j^{(N)}(i_2) - r_j^{(\infty)}(i_1)| \lesssim \varepsilon\gamma^{-1}\|i_1 - i_2\|_{s_0+\mu_1} + \varepsilon^{4-3a}N^{-\kappa}, \tag{7.102}$$

where $\kappa > \tau$. Here $m^{(N)}(i_2)$, $r_j^{(N)}(i_2)$ are defined in (1.39) of [33] and they are an approximation of $m(i_2)$ and $r_j^{(\infty)}(i_2)$ satisfying

$$\gamma^{-1}|m^{(N)}(i_2) - m(i_2)| + \langle j\rangle|r_j^{(N)}(i_2) - r_j^{(\infty)}(i_2)| \lesssim \varepsilon^{4-3a}N^{-\kappa}. \tag{7.103}$$

The bounds (7.102), (7.103) imply the (7.101). □

7.4. Proof of the inversion assumption (6.9). We are in position to give estimates on the inverse of the operator $\mathcal{L}_\omega$ in (6.15). Let us now define (recall (5.3))

$$\mathcal{F}_\infty^{2\gamma}(i_\delta) := \{\omega \in \mathcal{O}_0 : |\omega \cdot \ell - d_j^\infty(\omega)| \geq \frac{2\gamma}{\langle \ell \rangle^\tau}, \quad \forall \ell \in \mathbb{Z}^\nu, \forall j \in S^c\}. \tag{7.104}$$

We deduce the inversion assumption (6.9) by the following result.

Proposition 7.14. *Assume the hypothesis of Theorem 7.13, (6.12) with* $\mathfrak{p}_1 \geq \mu_1 + 2\tau + 1$, *where* μ_1 *is given in Theorem 7.1. Then for all* $\omega \in \Omega_\infty := \Omega_\infty^{2\gamma^{3/2}}(i_\delta) \cap \mathcal{F}_\infty^{2\gamma}(i_\delta)$ *(see (7.97)), for any function* $g \in H^{s+2\tau+1}_{S^\perp}(\mathbb{T}^{\nu+1})$ *the equation* $\mathcal{L}_\omega h = g$ *has a solution* $h = \mathcal{L}_\omega^{-1} g \in H^s_{S^\perp}(\mathbb{T}^{\nu+1})$, *satisfying*

$$\|\mathcal{L}_\omega^{-1} g\|_s^{\gamma,\Omega_\infty} \lesssim_s \gamma^{-1}(\|g\|_{s+2\tau+1}^{\gamma,\Omega_\infty} + \varepsilon\gamma^{-5/2}\|\mathfrak{I}_\delta\|_{s+\mathfrak{p}_1}^{\gamma,\Omega_\infty}\|g\|_{s_0}^{\gamma,\Omega_\infty}). \tag{7.105}$$

Proof. We conjugated the operator $\mathcal{L}_\omega$ in (6.15) to a diagonal operator $\mathcal{L}_\infty = \chi \mathcal{L}_\omega \chi^{-1}$, see (7.98), with (recall (7.99) and Theorem 7.1) $\chi := \Phi_\infty \circ \Upsilon$. Moreover, by (7.9) and (7.99) we have the following estimates

$$\|\chi^{\pm 1} h\|_s^{\gamma,\Omega_\infty} \lesssim_s \|h\|_s^{\gamma,\Omega_\infty} + \varepsilon\gamma^{-5/2}\|\mathfrak{I}_\delta\|_{s+\mu_1}^{\gamma,\mathcal{O}_0}\|h\|_{s_0}^{\gamma,\Omega_\infty}.$$

We have

$$\mathcal{L}_\infty^{-1} g = \sum_{j \neq 0} \frac{g_{\ell j}}{\mathrm{i}\big(\omega \cdot \ell - d_j^\infty(\omega)\big)} e^{\mathrm{i}(\ell \cdot \varphi + jx)} \tag{7.106}$$

and then $\|\mathcal{L}_\infty^{-1} g\|_s^{\gamma,\Omega_\infty} \leq \gamma^{-1}\|g\|_{s+2\tau+1}^{\gamma,\Omega_\infty}$. Thus we get the estimate (7.105). □

8. The Nash–Moser Nonlinear Iteration

In this section we prove Theorem 5.4. It will be a consequence of the Nash–Moser theorem 8.1.

Consider the finite-dimensional subspaces

$$E_n := \{\mathfrak{I}(\varphi) = (\Theta, y, z)(\varphi) : \Theta = \Pi_n \Theta, \, y = \Pi_n y, \, z = \Pi_n z\}$$

where

$$N_n := N_0^{\chi^n}, \quad n = 0, 1, 2, \ldots, \quad \chi := 3/2, \quad N_0 > 0 \tag{8.1}$$

and Π_n are the projectors (which, with a small abuse of notation, we denote with the same symbol)

$$\begin{aligned}
&\Pi_n \Theta(\varphi) := \sum_{|\ell| < N_n} \Theta_\ell \, e^{\mathrm{i}\ell \cdot \varphi}, \ \Pi_n y(\varphi) := \sum_{|\ell| < N_n} y_\ell \, e^{\mathrm{i}\ell \cdot \varphi}, \\
&\quad \text{where } \Theta(\varphi) = \sum_{\ell \in \mathbb{Z}^\nu} \Theta_\ell \, e^{\mathrm{i}\ell \cdot \varphi}, \ y(\varphi) = \sum_{\ell \in \mathbb{Z}^\nu} y_\ell \, e^{\mathrm{i}\ell \cdot \varphi}, \\
&\Pi_n z(\varphi, x) := \sum_{|(\ell, j)| < N_n} z_{\ell j} \, e^{\mathrm{i}(\ell \cdot \varphi + jx)}, \text{ where } \ z(\varphi, x) = \sum_{\ell \in \mathbb{Z}^\nu, \, j \in S^c} z_{\ell j} \, e^{\mathrm{i}(\ell \cdot \varphi + jx)}.
\end{aligned} \tag{8.2}$$

We define $\Pi_n^\perp = \mathrm{I} - \Pi_n$. The classical smoothing properties hold, namely, for all $\alpha, s \geq 0$,

$$\begin{aligned} \|\Pi_n \mathfrak{I}\|_{s+\alpha}^{\gamma,\mathcal{O}} \leq N_n^\alpha \|\mathfrak{I}_\delta\|_s^{\gamma,\mathcal{O}}, \quad \forall\, \mathfrak{I}(\omega) \in H^s, \quad \|\Pi_n^\perp \mathfrak{I}\|_s^{\gamma,\mathcal{O}} \\ \leq N_n^{-\alpha} \|\mathfrak{I}\|_{s+\alpha}^{\gamma,\mathcal{O}}, \quad \forall\, \mathfrak{I}(\omega) \in H^{s+\alpha}. \end{aligned} \tag{8.3}$$

Recall (5.3), (4.8) for the definition of b we set $a := 2b - 2$. We define the following constants

$$\begin{aligned} &\alpha_0 := 3\mu + 3, \qquad \alpha := 3\alpha_0 + 1, \qquad \alpha_1 := (\alpha - 3\mu)/2, \\ &k := 3(\alpha_0 + \rho^{-1}) + 1, \qquad \beta_1 := 6\alpha_0 + 3\rho^{-1} + 3, \\ &\frac{1}{2}\left(\frac{1 - (9/2)a}{C_1(1+a)}\right) < \rho < \frac{1 - (9/2)a}{C_1(1+a)} \end{aligned} \tag{8.4}$$

where $\mu := \mu(\nu) > 0$ is the "loss of regularity" given by the Theorem 6.1 and C_1 is fixed below.

Theorem 8.1 (Nash–Moser). *Assume that $f \in C^\infty$ (see (1.3)). Let $\tau := 2\nu + 6$. Then there exist $C_1 > \max\{\alpha_0 + \alpha, C_0\}$ (where $C_0 := C_0(\nu)$ is the one in Theorem 7.13), $\delta_0 := \delta_0(\nu) > 0$ such that, if*

$$N_0^{C_1} \varepsilon^{b_*+1} \gamma^{-7/2} < \delta_0, \quad \gamma := \varepsilon^{2+a} = \varepsilon^{2b}, \quad N_0 := (\varepsilon\gamma^{-1})^\rho, \quad b_* = 9 - 2b, \tag{8.5}$$

then there exists $C_* = C_*(S) > 0$ such that for all $n \geq 0$ the following holds:

$(\mathcal{P}1)_n$ *there exists a function $(\mathfrak{I}_n, \zeta_n) \colon \mathcal{G}_n \subseteq \Omega_\varepsilon \to E_{n-1} \times \mathbb{R}^\nu, \omega \mapsto (\mathfrak{I}_n(\omega), \zeta_n(\omega))$, $(\mathfrak{I}_0, \zeta_0) := (0,0)$, $E_{-1} := \{0\}$, where the set $\mathcal{G}_0$ is defined in (5.7) and the sets $\mathcal{G}_n$ for $n \geq 1$ are defined inductively by:*

$$\begin{aligned} &\mathcal{G}_{n+1} := \bigcap_{i=0}^{2} \Lambda_{n+1}^{(i)}, \quad \text{with} \quad \Lambda_{n+1}^{(0)} := \left\{\omega \in \mathcal{G}_n : |\omega \cdot \ell + m(i_n) j| \geq \frac{2\gamma_n}{\langle \ell \rangle^\tau}, \ \forall j \in S^c, \ell \in \mathbb{Z}^\nu\right\}, \\ &\Lambda_{n+1}^{(1)} := \left\{\omega \in \mathcal{G}_n : |\omega \cdot \ell + d_j^\infty(i_n)| \geq \frac{2\gamma_n}{\langle \ell \rangle^\tau}, \ \forall j \in S^c, \ell \in \mathbb{Z}^\nu\right\}, \\ &\Lambda_{n+1}^{(2)} := \left\{\omega \in \mathcal{G}_n : |\omega \cdot \ell + d_j^\infty(i_n) - d_k^\infty(i_n)| \geq \frac{2\gamma_n^{3/2}}{\langle \ell \rangle^\tau}, \ \forall j,k \in S^c, j \neq k, \ell \in \mathbb{Z}^\nu\right\}, \end{aligned} \tag{8.6}$$

where $\gamma_n := \gamma(1 + 2^{-n})$, $\gamma_n^* := \gamma^{3/2}(1 + 2^{-n})$ and $d_j^\infty(\omega) := d_j^\infty(\omega, i_n(\omega))$ are defined in (7.95) (and $d_0^\infty(\omega) = 0$). Moreover $|\zeta_n|^{\gamma,\mathcal{G}_n} \lesssim \|\mathcal{F}(U_n)\|_{s_0}^{\gamma,\mathcal{G}_n}$ and

$$\|\mathfrak{I}_n\|_{s_0+\mu}^{\gamma,\mathcal{G}_n} \leq C_* \varepsilon^{b_*} \gamma^{-1}, \quad \|\mathcal{F}(U_n)\|_{s_0+\mu+3}^{\gamma,\mathcal{G}_n} \leq C_* \varepsilon^{b_*}, \tag{8.7}$$

where $U_n := (i_n, \zeta_n)$ with $i_n(\varphi) = (\varphi, 0, 0) + \mathfrak{I}_n(\varphi)$. The differences $\widehat{\mathfrak{I}}_n := \mathfrak{I}_n - \mathfrak{I}_{n-1}$ (where we set $\widehat{\mathfrak{I}}_0 := 0$) is defined on $\mathcal{G}_n$, and satisfy

$$\|\widehat{\mathfrak{I}}_1\|_{s_0+\mu}^{\gamma,\mathcal{G}_1} \leq C_* \varepsilon^{b_*} \gamma^{-1}, \quad \|\widehat{\mathfrak{I}}_n\|_{s_0+\mu}^{\gamma,\mathcal{G}_n} \leq C_* \varepsilon^{b_*} \gamma^{-1} N_{n-1}^{-\alpha}, \quad \forall n \geq 2. \tag{8.8}$$

$(\mathcal{P}2)_n$ $\|\mathcal{F}(U_n)\|_{s_0}^{\gamma,\mathcal{G}_n} \leq C_* \varepsilon^{b_*} N_{n-1}^{-\alpha}$ *where we set $N_{-1} := 1$.*

$(\mathcal{P}3)_n$ *(High Norms)* $\|\mathfrak{I}_n\|_{s_0+\beta_1}^{\gamma,\mathcal{G}_n} \leq C_* \varepsilon^{b_*} \gamma^{-1} N_{n-1}^k$ *and* $\|\mathcal{F}(U_n)\|_{s_0+\beta_1}^{\gamma,\mathcal{G}_n} \leq C_* \varepsilon^{b*} N_{n-1}^k$.

$(\mathcal{P}4)_n$ *(Measure) The measure of the "Cantor-like" sets* $\mathcal{G}_n$ *satisfies*

$$|\Omega_\varepsilon \setminus \mathcal{G}_0| \leq C_* \varepsilon^{2(\nu-1)}\gamma, \quad |\mathcal{G}_n \setminus \mathcal{G}_{n+1}| \leq C_* \varepsilon^{2(\nu-1)}\gamma N_{n-1}^{-1}. \tag{8.9}$$

Proof. To simplify the notations we omit the index γ, $\mathcal{G}_n$ on the norm $\|\cdot\|_s$.
Proof of $(\mathcal{P}_1)_0, (\mathcal{P}_2)_0, (\mathcal{P}_3)_0$. Recalling (5.10), we have, by the second estimate in (5.13),

$$\|\mathcal{F}(U_0)\|_s = \|\mathcal{F}((\varphi, 0, 0), 0)\|_s = \|X_P(\varphi, 0, 0)\|_s \lesssim_s \varepsilon^{9-2b}.$$

Hence the smallness conditions in $(\mathcal{P}_1)_0, (\mathcal{P}_2)_0, (\mathcal{P}_3)_0$ hold taking C_* large enough.
Assume that $(\mathcal{P}_1)_n, (\mathcal{P}_2)_n, (\mathcal{P}_3)_n$ *hold for some* $n \geq 0$, *and prove* $(\mathcal{P}_1)_{n+1}, (\mathcal{P}_2)_{n+1}, (\mathcal{P}_3)_{n+1}$. By (8.4) and (8.5)

$$N_0^{C_1}\varepsilon^{b_*+1}\gamma^{-7/2} = \varepsilon^{1-(9/2)a-\rho C_1(1+a)} < \delta_0$$

for ε small enough. If we take C_1 bigger than C_0 in Theorem 7.13 then (7.94) holds. In (6.12) we consider $\mathfrak{p}_0 = \mathfrak{p}_1 + \tilde{\mathfrak{p}}$, where $\mathfrak{p}_1 := \mu_1 + 2\tau + 1$ and $\tilde{\mathfrak{p}}$ appears in (6.5). Since $\mu \gg \mathfrak{p}_0$ (8.7) implies (6.3) and so (6.12), and Proposition 7.14 applies. Hence the operator $\mathcal{L}_\omega := \mathcal{L}_\omega(\omega, i_n(\omega))$ in (6.15) is defined on $\mathcal{O}_0 = \mathcal{G}_n$ and is invertible for all $\omega \in \mathcal{G}_{n+1}$ since $\mathcal{G}_{n+1} \subseteq \Omega_\infty^{2\gamma_n^*}(i_n) \cap \mathcal{F}_\infty^{2\gamma_n}(i_n)$ and the (7.105) holds. This means that the assumption (6.9) of Theorem 6.1 is verified with $\Omega_\infty = \mathcal{G}_{n+1}$. By Theorem 6.1 there exists an approximate inverse $\mathbf{T}_n(\omega) := \mathbf{T}_0(\omega, i_n(\omega))$ of the linearized operator $L_n(\omega) := \mathtt{D}\mathcal{F}(\omega, i_n(\omega), \zeta_n) \equiv \mathtt{D}\mathcal{F}(\omega, i_n(\omega), 0)$, satisfying (6.10). By (8.5), (8.7)

$$\|\mathbf{T}_n g\|_s \lesssim_s \gamma^{-1}(\|g\|_{s+\mu} + \varepsilon\gamma^{-5/2}\{\|\mathfrak{I}_n\|_{s+\mu} + \gamma^{-1}\|\mathfrak{I}_n\|_{s_0+\mu}\|\mathcal{F}(U_n)\|_{s+\mu}\}\|g\|_{s_0+\mu}), \tag{8.10}$$

$$\|\mathbf{T}_n g\|_{s_0} \lesssim \gamma^{-1}\|g\|_{s_0+\mu} \tag{8.11}$$

and, by (6.11), using also (8.5), (8.7), (8.3),

$$\begin{aligned}\|(L_n \circ \mathbf{T}_n - \mathrm{I})g\|_s \lesssim_s &\varepsilon^{2b-1}\gamma^{-2}(\|\mathcal{F}(U_n)\|_{s_0+\mu}\|g\|_{s+\mu} + \|\mathcal{F}(U_n)\|_{s+\mu}\|g\|_{s_0+\mu} \\ &+ \varepsilon\gamma^{-5/2}\|\mathfrak{I}_n\|_{s+\mu}\|\mathcal{F}(U_n)\|_{s_0+\mu}\|g\|_{s_0+\mu})\end{aligned} \tag{8.12}$$

$$\|(L_n \circ \mathbf{T}_n - \mathrm{I})g\|_{s_0} \lesssim \varepsilon^{2b-1}\gamma^{-2}(\|\Pi_n\mathcal{F}(U_n)\|_{s_0+\mu} + \|\Pi_n^\perp\mathcal{F}(U_n)\|_{s_0+\mu})\|g\|_{s_0+\mu} \tag{8.13}$$

$$\lesssim \varepsilon^{2b-1}\gamma^{-2}N_n^\mu\Big(\|\mathcal{F}(U_n)\|_{s_0} + N_n^{-\beta_1}\|\mathcal{F}(U_n)\|_{s_0+\beta_1}\Big)\|g\|_{s_0+\mu}. \tag{8.14}$$

Now, for all $\omega \in \mathcal{G}_{n+1}$, we can define, for $n \geq 0$,

$$U_{n+1} := U_n + H_{n+1}, \quad H_{n+1} := (\widehat{\mathfrak{I}}_{n+1}, \widehat{\zeta}_{n+1}) := -\tilde{\Pi}_n\mathbf{T}_n\Pi_n\mathcal{F}(U_n) \in E_n \times \mathbb{R}^\nu, \tag{8.15}$$

where $\tilde{\Pi}_n(\mathfrak{I}, \zeta) := (\Pi_n\mathfrak{I}, \zeta)$ with Π_n defined in (8.2). By construction we have

$$\begin{aligned}&\mathcal{F}(U_{n+1}) = \mathcal{F}(U_n) + L_nH_{n+1} + Q_n, \\ &Q_n := Q(U_n, H_{n+1}), \quad Q(U_n, H) := \mathcal{F}(U_n + H) - \mathcal{F}(U_n) - L_nH, \quad H \in E_n \times \mathbb{R}^\nu.\end{aligned}$$

Then, by the definition of H_{n+1} in (8.15), using $[L_n, \Pi_n]$ and writing $\tilde{\Pi}_n^\perp(\mathfrak{I}, \zeta) := (\Pi_n^\perp\mathfrak{I}, 0)$ we have

$$\mathcal{F}(U_{n+1}) = \mathcal{F}(U_n) - L_n\tilde{\Pi}_n\mathbf{T}_n\Pi_n\mathcal{F}(U_n) + Q_n = \mathcal{F}(U_n) - L_n\mathbf{T}_n\Pi_n\mathcal{F}(U_n)$$

$$
\begin{aligned}
&+ L_n\tilde{\Pi}_n^{\perp}\mathbf{T}_n\Pi_n\mathcal{F}(U_n) + Q_n\\
&= \mathcal{F}(U_n) - \Pi_n L_n\mathbf{T}_n\Pi_n\mathcal{F}(U_n) + (L_n\tilde{\Pi}_n^{\perp} - \Pi_n^{\perp}L_n)\mathbf{T}_n\Pi_n\mathcal{F}(U_n) + Q_n\\
&= \Pi_n^{\perp}\mathcal{F}(U_n) + R_n + Q_n + Q_n'
\end{aligned}
$$

where

$$
R_n := (L_n\tilde{\Pi}_n^{\perp} - \Pi_n^{\perp}L_n)\mathbf{T}_n\Pi_n\mathcal{F}(U_n), \quad Q_n' := -\Pi_n(L_n\mathbf{T}_n - \mathrm{I})\Pi_n\mathcal{F}(U_n). \tag{8.16}
$$

Lemma 8.2. *Define*

$$
w_n := \varepsilon\gamma^{-2}\|\mathcal{F}(U_n)\|_{s_0}, \qquad B_n := \varepsilon\gamma^{-1}\|\mathfrak{I}_n\|_{s_0+\beta_1} + \varepsilon\gamma^{-2}\|\mathcal{F}(U_n)\|_{s_0+\beta_1}. \tag{8.17}
$$

Then there exists $K := K(s_0, \beta_1) > 0$ *such that, for all* $n \geq 0$, *setting* $\alpha_0 := 3\mu + 3$

$$
w_{n+1} \leq K N_n^{\alpha_0+\rho^{-1}-\beta_1} B_n + K N_n^{\alpha_0} w_n^2, \qquad B_{n+1} \leq K N_n^{\alpha_0+\rho^{-1}} B_n. \tag{8.18}
$$

The proof of Lemma 8.2 follows almost word by word the proof of Lemma 9.2 in [4].
Proof of $(\mathcal{P}_3)_{n+1}$. By (8.18) and $(\mathcal{P}_3)_n$

$$
B_{n+1} \leq K N_n^{\alpha_0+\rho^{-1}} B_n \leq 2C_* K\varepsilon^{b_*+1}\gamma^{-2}N_n^{\alpha_0+\rho^{-1}}N_{n-1}^{k} \leq C_*\varepsilon^{b_*+1}\gamma^{-2}N_n^{k}, \tag{8.19}
$$

provided $2K N_n^{\alpha_0+\rho^{-1}-k}N_{n-1}^{k} \leq 1$, $\forall n \geq 0$. Choosing k as in (8.4) and N_0 large enough, i.e. for ε small enough. By (8.17) and the bound (8.19) $(\mathcal{P}_3)_{n+1}$ holds.
Proof of $(\mathcal{P}_2)_{n+1}$. Using (8.17), (8.18) and $(\mathcal{P}_2)_n$, $(\mathcal{P}_3)_n$, we get

$$
\begin{aligned}
&w_{n+1} \leq K N_n^{\alpha_0+\rho^{-1}-\beta_1} B_n + K N_n^{\alpha_0} w_n^2 \leq K N_n^{\alpha_0+\rho^{-1}-\beta_1} 2C_*\varepsilon^{b_*+1}\gamma^{-2}N_{n-1}^{k}\\
&+K N_n^{\alpha_0}(C_*\varepsilon^{b_*+1}\gamma^{-2}N_{n-1}^{-\alpha})^2
\end{aligned}
$$

and $w_{n+1} \leq C_*\varepsilon^{b_*+1}\gamma^{-2}N_n^{-\alpha}$ provided that

$$
4K N_n^{\alpha_0+\rho^{-1}-\beta_1+\alpha}N_{n-1}^{k} \leq 1, \quad 2K C_*\varepsilon^{b_*+1}\gamma^{-2}N_n^{\alpha_0+\alpha}N_{n-1}^{-2\alpha} \leq 1, \ \forall n \geq 0. \tag{8.20}
$$

The inequalities in (8.20) hold by (8.5), taking α as in (8.4), $C_1 > \alpha_0 + \alpha$ and δ_0 in (8.5) small enough. By (8.17), the inequality $w_{n+1} \leq C_*\varepsilon^{b_*+1}\gamma^{-2}N_n^{-\alpha}$ implies $(\mathcal{P}_2)_{n+1}$.
Proof of $(\mathcal{P}_1)_{n+1}$. The bound (8.8) for $\widehat{\mathfrak{I}}_1$ follows by (8.15), (8.10) (for $s = s_0 + \mu$) and $\|\mathcal{F}(U_0)\|_{s_0+2\mu} \lesssim_{s_0+2\mu} \varepsilon^{b_*}$. The bound (8.8) for $\widehat{\mathfrak{I}}_{n+1}$ follows by (8.2), $(\mathcal{P}_2)_n$ and (8.4). It remains to prove that (8.7) holds at the step $n+1$. We have

$$
\|\mathfrak{I}_{n+1}\|_{s_0+\mu} \leq \sum_{k=1}^{n+1}\|\widehat{\mathfrak{I}}_k\|_{s_0+\mu} \leq C_*\varepsilon^{b_*}\gamma^{-1}\sum_{k\geq 1}N_{k-1}^{-\alpha_1} \lesssim C_*\varepsilon^{b_*}\gamma^{-1} \tag{8.21}
$$

taking α_1 as in (8.4) and N_0 large enough, i.e. ε small enough. Moreover, using (8.2), $(\mathcal{P}_2)_{n+1}$, $(\mathcal{P}_3)_{n+1}$, (8.4) we get

$$
\begin{aligned}
\|\mathcal{F}(U_{n+1})\|_{s_0+\mu+1} &\leq N_n^{\mu+1}\|\mathcal{F}(U_{n+1})\|_{s_0} + N_n^{\mu+1-\beta_1}\|\mathcal{F}(U_{n+1})\|_{s_0+\beta_1}\\
&\leq C_*\varepsilon^{b_*}N_n^{\mu+1-\alpha} + C_*\varepsilon^{b_*}N_n^{\mu+1-\beta_1+k} \lesssim C_*\varepsilon^{b_*},
\end{aligned}
$$

which is the second inequality in (8.7) at the step $n+1$. The bound $|\zeta_{n+1}|^{\gamma} \lesssim \|\mathcal{F}(U_{n+1})\|_{s_0}^{\gamma}$ is a consequence of Lemma 6.1 in [39].
To conclude the proof of Theorem (8.1) it remains to show the bounds (8.9). This is done in the next section. □

8.1. Measure estimates. Let us define for $0 < \eta \leq \overline{j}_1^4 \sqrt{\varepsilon}$, $\sigma \geq 1$ and $n \in \mathbb{N}$

$$R_{\ell jk}(\eta, \sigma) := R_{\ell jk}(i_n, \eta, \sigma) := \{\omega \in \mathcal{G}_n : |\omega \cdot \ell + d_j^\infty - d_k^\infty| \leq 2\eta \langle \ell \rangle^{-\sigma}\}, \quad (8.22)$$

$$Q_{\ell j}(\eta, \sigma) := Q_{\ell j}(i_n, \eta, \sigma) := \{\omega \in \mathcal{G}_n : |\omega \cdot \ell + mj| \leq 2\eta \langle \ell \rangle^{-\sigma}\}, \quad (8.23)$$

$$P_{\ell j}(\eta, \sigma) := P_{\ell j}(i_n, \eta, \sigma) := \{\omega \in \mathcal{G}_n : |\omega \cdot \ell + d_j^\infty| \leq 2\eta \langle \ell \rangle^{-\sigma}\}. \quad (8.24)$$

Recalling (8.6) we can write, setting $\eta \rightsquigarrow \gamma_n$ for the sets $Q_{\ell j}(\eta, \sigma)$ and $P_{\ell j}(\eta, \sigma)$, $\eta \rightsquigarrow \gamma_n^*$ for the set $R_{\ell jk}(\eta, \sigma)$, and $\sigma \rightsquigarrow \tau$,

$$\mathcal{G}_n \backslash \mathcal{G}_{n+1} = \bigcup_{\ell \in \mathbb{Z}^\nu, j,k \in S^c} \Big(R_{\ell jk}(i_n, \gamma_n^*, \tau) \cup Q_{\ell j}(i_n, \gamma_n, \tau) \cup P_{\ell j}(i_n, \gamma_n, \tau) \Big). \quad (8.25)$$

Since, by (5.7) and $\gamma > \gamma^{3/2}$ (see (8.5)), $R_{\ell jk}(i_n) = \emptyset$ for $j = k$, in the sequel we assume that $j \neq k$. We start with a preliminary lemma, which gives a first relation between ℓ, j, k which must be satisfied in order to have non empty resonant sets.

Lemma 8.3. *Let $n \geq 0$. There is a constant $C > 0$ dependent of the tangential set and independent of $\ell, j, k, n, i_n, \omega$ such that the following holds:*

- *If $R_{\ell jk}(i_n, \eta, \sigma) \neq \emptyset$ then $|\ell| \geq C|\lambda(j) - \lambda(k)| \geq \frac{C}{2}|j - k|$;*
- *If $Q_{\ell j}(i_n, \eta, \sigma) \neq \emptyset$ then $|\ell| \geq C|j|$;*
- *If $P_{\ell j}(i_n, \eta, \sigma) \neq \emptyset$ then $|\ell| \geq C|j|$.*

Proof. If $R_{\ell jk}(i_n) \neq \emptyset$, then there exists ω such that

$$|d_j^\infty(\omega, i_n(\omega)) - d_k^\infty(\omega, i_n(\omega))| < 2\eta \langle \ell \rangle^{-\sigma} + |\omega \cdot \ell|.$$

Moreover, using (7.95), (7.96), (7.2), (7.5), we get $|d_j^\infty(\omega, i_n(\omega)) - d_k^\infty(\omega, i_n(\omega))| \geq \frac{1}{3}|\lambda(j) - \lambda(k)|$. Thus, for ε small enough

$$2|\overline{\omega}||\ell| \geq |\omega \cdot \ell| \geq \left(\frac{1}{3} - \frac{2\eta}{\langle \ell \rangle^\sigma |\lambda(j) - \lambda(k)|} \right) |\lambda(j) - \lambda(k)| \geq \frac{1}{4}|\lambda(j) - \lambda(k)|$$

and this proves the first claim on $R_{\ell jk}$. If $Q_{\ell j} \neq \emptyset$ then we have $|mj| < 2\eta \langle \ell \rangle^{-\sigma} + |\omega \cdot \ell|$. Hence, for ε small enough, we have

$$|j| \leq \frac{|\omega \cdot \ell|}{|m|} \leq \frac{1}{\tilde{C}}|\ell|, \qquad \tilde{C} := \frac{|m|}{4|\overline{\omega}|}.$$

Following the same arguments and by using that $|d_j| \geq C|j|$ for some constant $C > 0$ we get the last statement. □

8.1.1. Measure of a resonant set The aim of this subsection is to prove the following lemma.

Lemma 8.4. *There is $\mathtt{r}_0 > 0$ such that, for any $0 < \mathtt{r} \leq \mathtt{r}_0$, and any choice of $S^+ \in \mathcal{V}(\mathtt{r})$ we have that*

$$|R_{\ell jk}(\eta, \sigma)| \leq K \varepsilon^{2(\nu-1)} \eta \langle \ell \rangle^{-\sigma}, \quad (8.26)$$

for some $K = K(S)$. The same holds for $Q_{\ell j}(\eta, \sigma)$ and $P_{\ell j}(\eta, s)$.

The proof of the lemma above involves many arguments and we split it into several steps.

In several bounds we will evidence the dependence of the constants on the tangential set S in order to highlight that the smallness of the amplitudes ξ depends on the choice of the tangential sites.

Let us first consider the set $R_{\ell jk}$, which is the most difficult case. We study the sub-levels of the function $\omega \mapsto \phi_R(\omega)$ defined by (recall (4.5),(7.95))

$$\begin{aligned}\phi_R(\omega) &:= \mathrm{i}\omega\cdot\ell + d_j^\infty(\omega) - d_k^\infty(\omega) = \mathrm{i}\omega\cdot\ell + \mathrm{i}m(\omega)(\lambda(j)-\lambda(k))\\ &\quad + \mathrm{i}\varepsilon^2(\kappa_j-\kappa_k)(\omega) + (r_j^\infty - r_k^\infty)(\omega).\end{aligned} \tag{8.27}$$

We recall that (see (7.2), (7.48))

$$\begin{aligned}&m = 1+\varepsilon^2 c(\omega) + \mathfrak{r}_m(\omega),\quad c(\omega) = \vec v\cdot\xi(\omega),\quad \vec v := (2/3)(1+\overline{\jmath}_k^2)_{k=1}^\nu \in \mathbb{R}^\nu,\\ &\kappa_j(\omega) = \vec w_j\cdot\xi(\omega),\end{aligned} \tag{8.28}$$

where κ_j is defined in (7.5) (see also (7.40)) and

$$\mathfrak{r}_m := \varepsilon^4 d(\omega) + O(\varepsilon^{10}\gamma^{-2}),\quad |\mathfrak{r}_m|^\gamma \lesssim \overline{\jmath}_1\varepsilon^4,\quad |\nabla_\omega \mathfrak{r}_m| \lesssim \overline{\jmath}_1\varepsilon^2 + O(\varepsilon^{10}\gamma^{-3}). \tag{8.29}$$

We first study some properties of the function $\phi_R(\omega)$ in (8.27).

The small divisor $\phi_R(\omega)$ as an affine function of ω It will be useful to consider $\phi_R(\omega)$ in (8.27) as a small perturbation of an affine function in ω

$$\phi_R(\omega) = a_{jk} + b_{\ell jk}\cdot\omega + q_{jk}(\omega),\qquad \ell\in\mathbb{Z}^\nu,\ j,k\in S^c. \tag{8.30}$$

where

$$a_{jk} := \mathrm{i}\Big((\lambda(j)-\lambda(k))[1-\vec v\cdot\mathbb{A}^{-1}\overline{\omega}] + (\vec w_k - \vec w_j)\cdot\mathbb{A}^{-1}\overline{\omega}\Big), \tag{8.31}$$

$$b_{ljk} := \mathrm{i}\Big(\ell + (\lambda(j)-\lambda(k))\mathbb{A}^{-T}\vec v + \mathbb{A}^{-T}(\vec w_j - \vec w_k)\Big), \tag{8.32}$$

and the remainder $q_{jk}(\omega)$ satisfies

$$\begin{aligned}&|q_{jk}(\omega)|^{sup} \lesssim \overline{\jmath}_1\varepsilon^4|j-k| + \varepsilon^{4-3a},\\ &|q_{jk}(\omega)|^{lip} \le |\mathfrak{r}_m(\omega)|^{lip}|\lambda(j)-\lambda(k)| + |r_j^\infty - r_k^\infty|^{lip} \lesssim \overline{\jmath}_1\varepsilon^2|j-k| + \varepsilon^{1-4a}.\end{aligned} \tag{8.33}$$

Lemma 8.5. *Denoting $\vec p_j = \lambda(j)\vec v + \vec w_j$, we have the following bounds:*

$$\begin{aligned}&|\vec p_j| \lesssim \overline{\jmath}_1^4|j|,\qquad |\vec w_j| \lesssim \overline{\jmath}_1^6|j|^{-1},\qquad |\vec p_j - \vec p_k| \lesssim \overline{\jmath}_1^4|j-k|,\\ &|\vec w_j - \vec w_k| \lesssim \overline{\jmath}_1^8|j-k|(|j|^{-2}+|jk|^{-1}).\end{aligned}$$

Proof. The first bound follows by the fact that $3+\overline{\jmath}_i^2 + j^2 \pm j_2\overline{\jmath}_i \ge 3 + \frac{j^2+\overline{\jmath}_i^2}{2} \ge \frac{j^2}{2}$ and $(1+j^2)(1+\overline{\jmath}_i^2)(2+j^2+\overline{\jmath}_i^2) \lesssim \overline{\jmath}_1^4 j^4$. The others follow similarly. □

Fix $\alpha\in(0,1/2)$ and let

$$0<\beta<(2-\alpha)(\sigma+1)^{-1}. \tag{8.34}$$

We have the following estimates for sets $R_{\ell jk}$ with $|\ell|$ "large".

Lemma 8.6. *Let* $|\ell| > \varepsilon^{-\beta}$. *Then* $R_{\ell jk}(\eta, \sigma)$ *satisfies* (8.26).

Proof. In this proof we shall denote by $C(S)$ a running positive constant depending on the set S. Suppose that $|j-k| \le c_0|\ell|$ with $c_0 = c_0(S)$ small. By Lemmata 8.5 and 4.1 we have $|\mathbb{A}^{-T}(\vec{p}_j - \vec{p}_k)| \lesssim \overline{J}_1^4 |j-k| < |\ell|/2$ for c_0 sufficiently small. This means that $|b_{\ell jk}| \gtrsim |\ell|/2$. Now suppose that $|j-k| > c_0|\ell|$. Then

$$|a_{jk}| \geq |\lambda(j)-\lambda(k)|\Big(|1 - \mathbb{A}^{-1}\overline{\omega}\cdot\vec{v}| - \frac{\mathbb{A}^{-T}(\vec{w}_j - \vec{w}_k)}{|\lambda(j)-\lambda(k)|}|\Big)$$
$$\overset{(A.4),(A.7)}{\geq} |j-k|\Big(\frac{1}{2} - \frac{C(S)|j-k|}{c_0\,|\ell|\,|jk|}\Big) \geq |j-k|\Big(\frac{1}{2} - 2\,c_0^{-1}\,C(S)\varepsilon^{\beta}\Big) \geq |j-k|/4$$

for ε small enough. By (8.22), (8.33)

$$2|b_{\ell jk}||\overline{\omega}| \geq |b_{\ell jk}\cdot\omega| \geq |a_{jk}| - |\phi_{\ell jk}(\omega)| - |q_{jk}(\omega)|$$
$$\geq \Big(\frac{1}{4} - \frac{2\eta}{c_0\langle\ell\rangle^{\sigma+1}} - C(S)\varepsilon^4 - \frac{\varepsilon^{4-3a}}{|j-k|}\Big)|j-k| \geq \frac{1}{8}|j-k|,$$

for ε small enough and $\sigma \geq 1$. Again we have shown that $|b_{\ell jk}| > \delta|\ell|$ with $\delta := c_0/2|\overline{\omega}|$. Split $\omega = s\widehat{b} + b^{\perp}$ where $\widehat{b} := b/|b|$ and $b^{\perp}\cdot b = 0$. Let $\Psi_R(s) := \phi_R(s\widehat{b} + b^{\perp})$. For ε small enough, by (8.33), we get

$$|\Psi_R(s_1) - \Psi_R(p)| \geq (|b| - |q_{jk}|^{lip})|s_1 - p|$$
$$\geq \left(\delta - C(S)\varepsilon^2 - \frac{\varepsilon^{1-4a}}{|j-k|}\right)|j-k|\,|s_1-p| \geq \frac{\delta_1}{2}|j-k|\,|s_1-p|.$$

The lemma follows by Fubini's theorem. □

We now prove that if the main term (in size) of $\phi_R(\overline{\omega})$ is big enough and $|\ell|$ is bounded by some constant then the bad set $R_{\ell jk}(\eta,\sigma) = \emptyset$. We remark that

$$\phi_R(\overline{\omega}) - q_{jk}(\overline{\omega}) = a_{jk} + b_{\ell jk}\cdot\overline{\omega} = \overline{\omega}\cdot\ell + \lambda(j) - \lambda(k).$$

Lemma 8.7. *If* $|\ell| \le \varepsilon^{-\beta}$ *and* $|\overline{\omega}\cdot\ell + \lambda(j) - \lambda(k)| \geq \gamma_0\langle\ell\rangle^{-\sigma}$, *where* $\gamma_0 = \varepsilon^{\alpha}$, *then* $R_{\ell jk}(\eta,\sigma) = \emptyset$.

Proof. By definition

$$|\omega\cdot\ell + d_j^{\infty} - d_k^{\infty}| \geq \gamma_0\langle\ell\rangle^{-\sigma} - |b_{\ell jk}||\omega - \overline{\omega}| - 2|q_{jk}|^{sup}.$$

By Lemma 8.3 (recall (8.27)) we have $|j-k| \lesssim C|\ell|$ and so

$$|b_{\ell jk}||\omega-\overline{\omega}| + 2|q_{jk}|^{sup} \leq C(S)\varepsilon^2(|\ell| + |j-k|) \lesssim C(S)\varepsilon^2|\ell| \lesssim C(S)\varepsilon^{2-\beta}$$
$$\leq \varepsilon^{\alpha+\sigma\beta}/2 \leq \frac{\gamma_0}{2\langle\ell\rangle^{\sigma}} \tag{8.35}$$

for some $C(S) > 0$, for ε small enough, by (8.34). □

Lemma 8.8. *Let* $|\ell| \le \varepsilon^{-\beta}$ *and* $|\overline{\omega}\cdot\ell + \lambda(j) - \lambda(k)| \le \gamma_0\langle\ell\rangle^{-\sigma}$. *Then* $R_{\ell jk}(\eta,\sigma)$ *satisfies* (8.26).

Proof. Let us call $p := \overline{\omega}\cdot\ell+\lambda(j)-\lambda(k)$ and note that $|p| \lesssim \gamma_0 = \varepsilon^\alpha$. We also remark that $\ell \neq 0$ since for $j \neq k$ one has $|\lambda(j)-\lambda(k)| > 1/2$. We substitute p in the definition of $b_{\ell jk}$ (see (8.32))

$$|b_{\ell jk}| = \Big|\ell + (-\overline{\omega}\cdot\ell + p)\mathbb{A}^{-T}\vec{v} + \mathbb{A}^{-T}(\vec{w}_j - \vec{w}_k)\Big| \gtrsim |\ell - \mathbb{A}^{-T}\vec{v}\overline{\omega}^T\ell + \mathbb{A}^{-T}(\vec{w}_j - \vec{w}_k)| + \varepsilon^\alpha.$$

Then, using (A.5) in Lemma (A.1) we have $|b_{\ell jk}| \geq |\ell|\delta/2$ for ε small enough. The thesis follows reasoning as in Lemma (8.6). □

Proof of Lemma 8.4. For the sets $R_{\ell jk}$ the lemma follows by Lemmata 8.6, 8.7, 8.8.

The proof for the sets Q_{lj} and $P_{\ell j}$ follows using the same arguments used for $R_{\ell jk}$. Lemmata 8.6, 8.7 are identical, with the only difference that the non-resonance condition now reads respectively $|\overline{\omega}\cdot\ell + j| \geq \gamma_0\langle\ell\rangle^{-\sigma}, |\overline{\omega}\cdot\ell + \lambda(j)| \geq \gamma_0\langle\ell\rangle^{-\sigma}$ in the case of $Q_{\ell j}$ and $P_{\ell j}$. Regarding Lemma 8.8, it follows from (A.4) in the case of $Q_{\ell j}$ and from (A.6) in the case of $P_{\ell j}$. □

8.1.2. Summability

Lemma 8.9. *For $n \geq 1$, $|\ell| \leq N_{n-1}$, one has $R_{\ell jk}(i_n)$, $Q_{\ell j}(i_n)$, $P_{\ell j}(i_n) = \emptyset$.*

Proof. We first note that, by Lemma 8.3, if $|\lambda(j) - \lambda(k)| > C_1^{-1}|\ell|$ (for some $C_1 = C_1(S)$) then $R_{\ell jk}(i_n) = R_{\ell jk}(i_{n-1}) = \emptyset$, so that our claim is trivial. Otherwise, if

$$|\lambda(j) - \lambda(k)| \leq C_1^{-1}|\ell| \leq C_1^{-1}N_{n-1}.$$

By (7.101) (with $i_\delta^{(1)} \rightsquigarrow i_n$ and $i_\delta^{(2)} \rightsquigarrow i_{n-1}$, $N \rightsquigarrow N_{n-1}$) and (8.8) we have for all $j, k \in S^c$

$$|(d_j^\infty - d_k^\infty)(i_n) - (d_j^\infty - d_k^\infty)(i_{n-1})| \leq \varepsilon^{4-3a}N_{n-1}^{-\mathtt{a}} \qquad \forall\omega \in \mathcal{G}_n, \tag{8.36}$$

where $\mathtt{a} := \min\{\kappa, \alpha\}$ (recall α in (8.4) and κ in (7.13)). Now for all $j \neq k$, $|\ell| \leq N_{n-1}$, $\omega \in \mathcal{G}_n$ by (8.36)

$$\begin{aligned}|\omega\cdot\ell + d_j^\infty(i_n) - d_k^\infty(i_n)| &\geq |\omega\cdot\ell + d_j^\infty(i_{n-1}) - d_k^\infty(i_{n-1})| \\ &\quad - |(d_j^\infty - d_k^\infty)(i_n) - (d_j^\infty - d_k^\infty)(i_{n-1})| \\ &\geq 2\gamma_{n-1}^*\langle\ell\rangle^{-\tau} - \varepsilon^{4-3a}N_n^{-\mathtt{a}} \geq 2\gamma_n^*\langle\ell\rangle^{-\tau}\end{aligned} \tag{8.37}$$

since $\varepsilon^{4-3a}\gamma^{-3/2}N_n^{\tau-(2/3)\mathtt{a}}2^{n+1} \leq 1$. Since by definition $R_{\ell jk}(i_n) \subseteq \mathcal{G}_n$ then $R_{\ell jk}(i_n) = \emptyset$.

Now we prove that $Q_{\ell j}(i_{n-1}) \subseteq Q_{\ell j}(i_n)$. We have

$$\begin{aligned}|m(i_n) - m(i_{n-1})||j| &\overset{(7.48)}{\leq} C\varepsilon^3\|i_n - i_{n-1}\|_{s_0+2}|j| \overset{(8.8)}{\leq} C\varepsilon^{b_*+3}\gamma^{-1}N_{n-1}^{-\alpha}|j| \\ &\leq C\varepsilon^{b_*+3}\gamma^{-1}N_{n-1}^{-\alpha}|\ell|\end{aligned} \tag{8.38}$$

and then

$$\begin{aligned}|\omega\cdot\ell + m(i_n)j| &\geq |\omega\cdot\ell + m(i_{n-1})j| - |m(i_n) - m(i_{n-1})||j| \\ &\geq 2\gamma_{n-1}\langle\ell\rangle^{-\tau} - \varepsilon^{b_*+3}\gamma^{-1}N_{n-1}^{-\alpha+1} \geq 2\gamma_n\langle\ell\rangle^{-\tau}\end{aligned} \tag{8.39}$$

since $|\ell| \le N_{n-1}$.
As before, by (7.101), for all $j, k \in S^c$

$$|d_j^\infty(i_n) - d_j^\infty(i_{n-1})| \le \varepsilon^{4-3a} N_{n-1}^{-\mathrm{a}} \qquad \forall \omega \in \mathcal{G}_n. \tag{8.40}$$

For all $j \ne k$, $|\ell| \le N_{n-1}$, $\omega \in \mathcal{G}_n$ by (8.36)

$$\begin{aligned} |\omega \cdot \ell + d_j^\infty(i_n)| &\ge |\omega \cdot \ell + d_j^\infty(i_{n-1})| - |d_j^\infty(i_n) - d_j^\infty(i_{n-1})| \\ &\ge 2\gamma_{n-1}\langle \ell \rangle^{-\tau} - \varepsilon^{4-3a} N_{n-1}^{-\mathrm{a}} \ge 2\gamma_n \langle \ell \rangle^{-\tau} \end{aligned} \tag{8.41}$$

since $\varepsilon^{4-3a}\gamma^{-1} N_n^{\tau-(2/3)\mathrm{a}} 2^{n+1} \le 1$. □

We have proved that

$$\mathcal{G}_n \backslash \mathcal{G}_{n+1} \subseteq \bigcup_{\substack{j,k \in S^c \\ |\ell| > N_{n-1}}} \Big(R_{\ell jk}(i_n) \cup Q_{\ell j}(i_n) \cup P_{\ell j}(i_n) \Big), \quad \forall n \ge 1. \tag{8.42}$$

Lemma 8.10. *There exists* $\mathtt{C} > 0$ *such that if* $|j|, |k| \ge \mathtt{C}\langle \ell \rangle^{\nu+2} \gamma^{-(1/2)}$ *then (recall that* $\tau = 2\nu + 6 > \nu + 2$*)*

$$R_{\ell jk}(\gamma^{3/2}, \tau) \subseteq Q_{\ell, j-k}(\gamma, \nu + 2). \tag{8.43}$$

Proof. We have that

$$\begin{aligned} |\omega \cdot \ell + d_j^\infty - d_k^\infty| &\ge |\omega \cdot \ell + m(j-k)| - |m||\lambda(j) - j + k - \lambda(k)| - \varepsilon^2 |w_j - w_k| \\ &\quad - |r_j^\infty| - |r_k^\infty| \\ &\ge \frac{2\gamma}{\langle \ell \rangle^{\nu+2}} - 2|j-k| \frac{C}{|j||k|} - \frac{\tilde{C}\varepsilon^2}{\min\{|j|, |k|\}} \\ &\ge \frac{2\gamma}{\langle \ell \rangle^{\nu+2}} - \frac{C\gamma}{\mathtt{C}\langle \ell \rangle^{2(\nu+2)-1}} - \frac{\tilde{C}\varepsilon^2\sqrt{\gamma}}{\mathtt{C}\langle \ell \rangle^{\nu+2}} \\ &\ge \frac{\gamma}{\langle \ell \rangle^{\nu+2}} \Big(2 - \frac{C}{2\mathtt{C}\langle \ell \rangle^{\nu+1}} - \frac{\tilde{C}\varepsilon^2}{2\sqrt{\gamma}\mathtt{C}} \Big) \\ &\ge \frac{\gamma}{\langle \ell \rangle^{\nu+2}} \ge \frac{\gamma^{3/2}}{\langle \ell \rangle^{\tau}} \end{aligned} \tag{8.44}$$

for $\mathtt{C}$ big enough and since $\varepsilon^2(\sqrt{\gamma})^{-1} \ll 1$. □

We are in position to prove (8.9). We have, by (8.42),

$$\begin{aligned} \Big| \bigcup_{\ell \in \mathbb{Z}^\nu, j,k \in S^c} R_{\ell jk}(i_n) \Big| \le & \sum_{|\ell| > N_{n-1}, |j|, |k| \ge \mathtt{C}\langle \ell \rangle^{\nu+2}\gamma^{-(1/2)}} |R_{\ell jk}(i_n)| \\ & + \sum_{|\ell| > N_{n-1}, |j|, |k| \le 2\mathtt{C}\langle \ell \rangle^{\nu+2}\gamma^{-(1/2)}} |R_{\ell jk}(i_n)|. \end{aligned}$$

On one hand we have that, using Lemmata 8.4 and 8.10,

$$\sum_{|\ell| > N_{n-1}, |j|, |k| \ge \mathtt{C}\langle \ell \rangle^{\nu+2}\gamma^{-(1/2)}} |R_{\ell jk}(i_n)| \lesssim K \sum_{j-k=h, |h| \le C|\ell|} \varepsilon^{2(\nu-1)} \gamma \langle \ell \rangle^{-\nu-2}$$

$$\lesssim K\varepsilon^{2(\nu-1)}\gamma \sum_{|\ell|\geq N_{n-1}} \langle \ell\rangle^{-(\nu+1)}$$
$$\lesssim K\varepsilon^{2(\nu-1)}\gamma N_{n-1}^{-1}.$$

On the other hand

$$\sum_{\substack{|\ell|>N_{n-1},|j-k|\leq C|\ell|\\ |j|,|k|\leq 2C\langle\ell\rangle^{\tau_1}\gamma^{-(1/2)}}} |R_{\ell jk}(i_n)| \lesssim K\gamma^{(3/2)}\varepsilon^{2(\nu-1)} \sum_{|\ell|\geq N_{n-1}} \frac{|\ell|\langle\ell\rangle^{\nu+2}}{\sqrt{\gamma}\langle\ell\rangle^{\tau}}$$
$$\lesssim K\gamma\varepsilon^{2(\nu-1)} \sum_{|\ell|\geq N_{n-1}} \langle\ell\rangle^{-(\tau-\nu-3)}$$
$$\lesssim K\gamma\varepsilon^{2(\nu-1)} N_{n-1}^{-1}.$$

The discussion above implies estimates (8.9).

8.2. Conclusion of the Proof of Theorem 5.4. Theorem 8.1 implies that the sequence $(\mathfrak{I}_n, \zeta_n)$ is well defined for $\omega \in \mathcal{G}_\infty := \cap_{n\geq 0}\mathcal{G}_n$, $\mathfrak{I}_n$ is a Cauchy sequence in $\|\cdot\|_{s_0+\mu}^{\gamma,\mathcal{G}_\infty}$ (see (8.8)) and $|\zeta_n|^\gamma \to 0$. Therefore $\mathfrak{I}_n$ converges to a limit $\mathfrak{I}_\infty$ in norm $\|\cdot\|_{s_0+\mu}^{\gamma,\mathcal{G}_\infty}$ and, by $(\mathcal{P}2)_n$, for all $\omega \in \mathcal{G}_\infty$, $i_\infty(\varphi) := (\varphi, 0, 0) + \mathfrak{I}_\infty(\varphi)$ is a solution of

$$\mathcal{F}(i_\infty, 0) = 0 \qquad \text{with} \qquad \|\mathfrak{I}_\infty\|_{s_0+\mu}^{\gamma,\mathcal{G}_\infty} \lesssim \varepsilon^{9-2b}\gamma^{-1} \tag{8.45}$$

by (8.7), (8.5). Therefore $\varphi \mapsto i_\infty(\varphi)$ is an invariant torus for the Hamiltonian vector field X_{H_ε} (recall (5.1)). By (8.9),

$$|\Omega_\varepsilon \backslash \mathcal{G}_\infty| \leq |\Omega_\varepsilon \backslash \mathcal{G}_0| + \sum_{n\geq 0} |\mathcal{G}_n \backslash \mathcal{G}_{n+1}| \leq 2\, C_* \varepsilon^{2(\nu-1)}\gamma$$
$$+ C_*\varepsilon^{2(\nu-1)}\gamma \sum_{n\geq 1} N_{n-1}^{-1} \lesssim C_*\varepsilon^{2(\nu-1)}\gamma.$$

The set Ω_ε in (5.2) has measure $|\Omega_\varepsilon| = O(\varepsilon^{2\nu})$. Hence $|\Omega_\varepsilon \backslash \mathcal{G}_\infty|/|\Omega_\varepsilon| \to 0$ as $\varepsilon \to 0$ because $\gamma = o(\varepsilon^2)$, and therefore the measure of $\mathcal{C}_\varepsilon := \mathcal{G}_\infty$ satisfies (5.12).

It remains to show the linear stability of the embedding $i_\infty(\varphi)$. By the discussion of Sect. 6 (see also [4] for further details) and Sect. 7, since $i_\infty(\varphi)$ is isotropic and solves the Eq. (8.45), it is possible to find a change of coordinates G_∞ (of the form (6.6)), so that in the linearized system of the Hamiltonian $H_\varepsilon \circ G_\infty(\varphi, \eta, w)$ the equation for the actions is given by $\dot\eta = 0$. Moreover, by Sect. 7 the linear equation for the normal variables w is conjugated, by setting $w = \Upsilon \circ \Phi_\infty(z)$ to the diagonal system $\dot z_j - \mathrm{i}d_j^\infty(\omega) z_j = f_j(\omega t)$, $j \in S^c$, where $f(\omega t)$ is a forcing term.

Since $d_j^\infty \in \mathbb{R}$ a standard argument shows that the Sobolev norms of w do not increase in time. For further details see [2,4,40].

Acknowledgements. The authors wish to thank Andy Hone, Luca Biasco, Livia Corsi and Marcel Guardia for useful discussions and comments. The authors also thank the anonymous referees for their helpful remarks. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme under Grant Agreement No. 757802.

Appendix A. Non-degeneracy conditions

Proof of Lemma 4.1. Recalling (4.6) we introduce a matrix $\mathbb{K}$ so that $\mathbb{A} =: (2/9)\ \mathrm{diag}\left(\lambda(\overline{\jmath}_i)(1+\overline{\jmath}_i^2)\right)\mathbb{K}$. Now we show that the entries of $\mathbb{K}$ are bounded by some constant independent of the $\overline{\jmath}_i$. After some direct computations we have that

$$\mathbb{K}_{jj} = \frac{1+j^2}{j^2}, \quad j \in S^+,$$
$$\mathbb{K}_{jk} = 3\frac{(1+k^2)(2+k^2+j^2)}{(3+k^2+j^2+kj)(3+k^2+j^2-kj)}, \quad j,k \in S^+,\ j \neq k.$$

Obviously $|\mathbb{K}_{jj}| \leq 2$; regarding the off-diagonal terms, we note that $0 \leq 2kj \leq k^2+j^2$, hence $|\mathbb{K}_{jk}| \leq 12$ if $j \neq k$.

We consider the variables $x, p_2, \ldots, p_\nu$ defined as

$$\overline{\jmath}_1 =: 1/x, \quad \overline{\jmath}_i =: p_i/x, \quad 0 < p_i \leq 1, \tag{A.1}$$

so that $P(x, p_i) = \det(\mathbb{K})$ is a rational function. It is easily seen that $\mathbb{K}$ computed at $p_i = 1$ for all i, coincides with the matrix

$$(1+x^2)\left(\mathrm{I} + 2\,g(x)\,(U-\mathrm{I})\right), \qquad g(x) := (3x^2+1)^{-1},$$

where U is the matrix with components $U_{ij} = 1$ for any $i, j = 1, \ldots, \nu$. Its determinant is

$$\left(\frac{1+x^2}{1+3x^2}\right)^\nu (3x^2-1)^{\nu-1}(3x^2+2\nu-1). \tag{A.2}$$

We note that the absolute value of (A.2) is ≥ 1 at $x = 0$. We conclude that there exists $0 < \mathtt{r}_0 < 1$ such that

$$\text{if} \quad 0 \leq x < \mathtt{r}_0, \ |p_i - 1| \leq \mathtt{r}_0 \quad \text{then} \quad |P(x, p_i)| \geq 1/2.$$

This implies the thesis. □

Lemma A.1. *There exists $0 < \mathtt{r}_0 < 1$ such that, for any $S^+ \in \mathcal{V}(\mathtt{r})$ with $0 < \mathtt{r} \leq \mathtt{r}_0$ (see Definition 1.1), the following holds true:*

- $$\Big|\sum_{i=1}^{\nu} \frac{\overline{\jmath}_i}{1+\overline{\jmath}_i^2}\,\ell_i\Big| > \frac{\mathtt{r}}{2} \neq 0, \quad \forall \ell \in \mathbb{Z}^\nu, \quad |\ell| = 1, 2, 3, 5\,; \tag{A.3}$$
- $$\left|\det\left(\mathrm{I} - \mathbb{A}^{-T}\vec{v}(\overline{\omega})^T\right)\right| \geq 1\,; \tag{A.4}$$

- $$|\ell - \left(\mathrm{I} - \mathbb{A}^{-T}\vec{v}(\overline{\omega})^T\right)^{-1}\mathbb{A}^{-T}(\vec{w}_j - \vec{w}_k)| \geq \delta|\ell|, \quad \ell \in \mathbb{Z}^\nu,\ j,k \in S^c \tag{A.5}$$
- $$|\ell - \left(\mathrm{I} - \mathbb{A}^{-T}\vec{v}(\overline{\omega})^T\right)^{-1}\mathbb{A}^{-T}\vec{w}_j| \geq \delta|\ell|, \quad \ell \in \mathbb{Z}^\nu,\ j \in S^c \tag{A.6}$$

Here $\vec{v}$ and $\vec{w}_j$ are defined in (8.28) (see also (7.5)), $\overline{\omega}$ in (1.10) and δ is some appropriately small pure constant.

Proof of (A.3). The case $|\ell| = 1$ is trivial. For $|\ell| = 2$ we use the fact that the $\overline{\jmath}_i$ are all distinct. For $|\ell| = 3, 5$ we pass to the variables (A.1) and we get

$$\overline{\jmath}_1|\sum_{i=1}^{\nu}\frac{\overline{\jmath}_i}{1+\overline{\jmath}_i^2}\ell_i| = |\sum_{i=1}^{\nu}\frac{p_i}{x^2+p_i^2}\ell_i| =: L(x,p).$$

We notice that $L(0,1) = |\sum_i \ell_i| \geq 1$ (since $\sum_i \ell_i$ has the same parity of $|\ell|$) so, by continuity, there exists $0 < \mathtt{r}_0 < 1$ such that $L(x,p) > 1/2$ for all $0 \leq x < \mathtt{r}_0$, $|p_i - 1| \leq \mathtt{r}_0$. This implies the result. □

Proof of (A.4). We first note that (recall (8.28))

$$\det(\mathrm{I} - \mathbb{A}^{-T}\vec{v}\overline{\omega}^T) = 1 - \mathbb{A}^{-T}\vec{v}\cdot\overline{\omega}. \tag{A.7}$$

Consider the change of variables (A.1). One can note that the matrix $\mathbb{A}$ in (4.6) at $p_i = 1$, $i = 2, \dots, \nu$, is given by

$$\mathbb{A} := d(x)\big[\mathrm{I} + e(x)U\big], \qquad \mathbb{A}^{-1} := \frac{1}{d(x)}\big[\mathrm{I} - f(x)U\big],$$
$$d(x) := \frac{2(4x^2+1)(3x^4+2x^2+1)}{9x^3(3x^2+1)}, \qquad e(x) := \frac{2x^2}{3x^4+2x^2+1}, \quad f(x) := \frac{e(x)}{1+\nu e(x)}. \tag{A.8}$$

By (A.8) at $p_i = 1$ we have

$$\mathbb{A}^{-1}\vec{v}\cdot\overline{\omega} = \frac{2(4x^2+1)}{3x}(\vec{1}\cdot\mathbb{A}^{-1}\vec{1}) = 3\nu\,\frac{(4x^2+1)(3x^2+1)}{(4x^2+1)(3x^4+2(1+\nu)x^2+1)}. \tag{A.9}$$

We note that, for $x = 0$, one has $|\det(\mathrm{I} - \mathbb{A}^{-1}\vec{v}\overline{\omega}^T)| = 3\nu - 1 \geq 2$. Thus there is $0 < \mathtt{r}_0 < 1$ so that for all $0 \leq x < \mathtt{r}_0$, $|p_i - 1| \leq \mathtt{r}_0$ one has $|\det(\mathrm{I} - \mathbb{A}^{-1}\vec{v}\overline{\omega}^T)| \geq 1$. □

Proof of (A.5),(A.6). We systematically use the variables (A.1). We define $\Omega = \mathrm{diag}\,(\overline{\omega}_i)$, $V = \mathrm{diag}\,(\vec{v}_i)$ and write $\mathbb{A} = \Omega\mathbb{H}V$ where

$$\mathbb{H} = \frac{x^2+1}{3}[I + \frac{2}{(3x^2+1)}(U - I)] + O(|p-1|) = -\frac{1}{3}(I - 2U) + O(|p-1|, |x|).$$

Then $\vec{w}_j$ in (8.28) can be written as $\vec{w}_j = \Omega\, V\, b_j$ with

$$b_j \cdot \mathsf{e}_i = -\frac{\lambda(j)(7 + 5\overline{\jmath}_i^2 + \overline{\jmath}_i^4 + 3j^2)}{\lambda(\overline{\jmath}_i)((3+j^2)^2 + (6+j^2)\overline{\jmath}_i^2 + \overline{\jmath}_i^4)},$$

where e_i is the i-th vector of the canonical basis of $\mathbb{R}^\nu$. In the new coordinates (x, p) in (A.1), and setting $t = j(\overline{\jmath}_1)^{-1}$, this reads as

$$b_j \cdot \mathrm{e}_i =: b(t, x, p) \cdot \mathrm{e}_i = f(t, x) g(t, x, p_i) + t\, g(t, x, p_i),$$
$$f(t, x) := \frac{3\, t\, x^2}{x^2 + t^2}, \qquad g(t, x, p) := \frac{(x^2 + p_i^2)(7x^4 + 5x^2 p_i^2 + p_i^4 + 3x^2 t^2)}{p_i\,(4x^2 + p_i^2)((3x^2 + t^2)^2 + (6x^2 + t^2)p_i^2 + p_i^4)}.$$

We claim that

$$b(t, x, p) = -\frac{t}{t^4 + t^2 + 1}\vec{1} + O(|p - 1|, |x|) \tag{A.10}$$

where the term $O(|p - 1|, |x|)$ is uniform in t.
By direct computations we have

$$|g(t, x, p) - g(t, 0, 1)| \lesssim C\big(|x| + |p - 1|\big)(1 + t^2)^{-1}, \quad |g(t, x, p)| \le C(1 + t^2)^{-1},$$

with C independent of t, and for x, p in a neighborhood of $(0, 1)$. Moreover $\sup_t |f(t, x)| \le 3|x|^2/2$. Thus, for x, p sufficiently close to $(0, 1)$, we have that

$$\begin{aligned} |b(t, x, p) - b(t, 0, 1)| &\le |t||g(t, x, p) - g(t, 0, 1)| + |f(t, x)||g(t, x, p)| \\ &\lesssim O(|p - 1|, |x|), \end{aligned}$$

which implies the claim (A.10). Hence, setting $s = k(\overline{\jmath}_1)^{-1}$, we have

$$\begin{aligned} &\ell - \Omega^{-1}(\mathbb{H}^T - U)^{-1}\Omega(b(t, x, p) - b(s, x, p)) \\ &\quad = \ell - 3(1 + \nu)^{-1} h(t, s)\, \vec{1} + O(|p - 1|, |x|), \end{aligned}$$

where

$$h(t, s) := t\,(t^4 + t^2 + 1)^{-1} - s\,(s^4 + s^2 + 1)^{-1}.$$

We note that $|t\,(t^4 + t^2 + 1)^{-1}| \le 0.41$, hence each single component satisfies

$$|3(1 + \nu)^{-1} h(t, s) + O(|p - 1|, |x|) \cdot \mathrm{e}_i| \le 3(1 + \nu)^{-1}(1 - \delta) \le (1 - \delta)$$

provided that $|p - 1|, |x| \le \mathfrak{r}_0$ small enough. Hence, if $\nu \ge 2$, one has $|\ell_i - \big(\frac{3h(t,s)}{1+\nu}\,\vec{1} + O(|p - 1|, |x|)\big) \cdot \mathrm{e}_i| \ge \delta|\ell_i|$. This concludes the proof of (A.5). The proof of (A.6) is the same just setting $h(t) = t\,(t^4 + t^2 + 1)^{-1}$. □

Proof of Lemma 5.1. It is well known that, thanks to the choice of τ in (5.3), $|\Omega_\varepsilon \backslash \mathcal{G}_0^{(0)}| \le C\varepsilon^{2(\nu-1)}\gamma$ for some $C = C(S) > 0$. Thus we focus on the estimate for the measure of $\mathcal{G}_0^{(1)}$. For indexes $\ell \in \mathbb{Z}\backslash\{0\}$, $j, k \in S^c$ satisfying

$$|\ell| \le 3, \qquad \text{and} \qquad \sum_{i=1}^{\nu} \overline{\jmath}_i\, \ell_i + j = k, \tag{A.11}$$

we define the sets $T_{\ell jk} := \{\omega \in \Omega_\varepsilon : |\overline{\omega}\cdot\ell + \varepsilon^2 \mathbb{A}\xi\cdot\ell + \lambda(j) - \lambda(k) + \varepsilon^2(\lambda(j)\mathfrak{l}_j - \lambda(k)\mathfrak{l}_k)| \le C\gamma\}$. Recalling (5.5) we have that $\Omega_\varepsilon \backslash \mathcal{G}_0^{(1)} = \bigcup T_{\ell jk}$ where the union is restricted to ℓ, j, k satisfying (A.11).

Let us first study the ε-independent part of our small divisor. By (1.8) and (A.11) we have

$$\overline{\omega}\cdot\ell+\lambda(j)-\lambda(k)=3\sum_{i=1}^{\nu}\frac{\overline{\jmath}_i}{1+\overline{\jmath}_i^2}\ell_i+3\frac{j}{1+j^2}-3\frac{(\sum_{i=1}^{\nu}\overline{\jmath}_i\ell_i+j)}{1+(\sum_{i=1}^{\nu}\overline{\jmath}_i\ell_i+j)^2}. \tag{A.12}$$

By Lemma A.1 (see (A.3)) if $|j|\geq\widehat{C}(S)$, which is easily computed in terms of $\mathtt{r}$, (A.12) is bounded from below by $\mathtt{r}/4$. By (A.11), $|\lambda(j)-\lambda(k)|\leq C(S)$. Since $0<|\ell|\leq 3$ and substituting $\mathfrak{l}_j=c(\omega)+\kappa_j/\lambda(j)$ (recall Lemma 7.10), where $c(\omega),\kappa_j$ are defined in (7.2), (7.5))

$$|\overline{\omega}\cdot\ell+\varepsilon^2\mathbb{A}\xi\cdot\ell+(1+\varepsilon^2c(\omega))(\lambda(j)-\lambda(k))+\varepsilon^2(\kappa_j-\kappa_k)|\geq\mathtt{r}/8$$

for ε small enough (depending on $\mathtt{r}$) and by using the fact that $\kappa_j-\kappa_k$ is uniformly bounded in j,k. This implies that $T_{\ell jk}=\emptyset$ for ε small enough.

We are left to deal with the case $|j|\leq\widehat{C}(S)$. We write (A.12) as $P(\overline{\jmath}_i,j)/Q(\overline{\jmath}_i,j)$ where P,Q are polynomials with integer coefficients and Q has no real zeros. We remark that $1<Q<C(S)$ due to the condition $|j|\leq\widehat{C}(S)$. If $P\neq 0$ then $|P|\geq 1$ and again (A.12) is larger than some $K(S)$. We conclude that $T_{\ell jk}=\emptyset$ by reasoning as in the case j large. Now we study the case in which $P=0$. Fixed ℓ in (A.12), then P has degree four in j and so the condition $P=0$ fixes at most four choices of j that we call $\widehat{\jmath}_1,\widehat{\jmath}_2,\widehat{\jmath}_3,\widehat{\jmath}_4$. For $P=0$ (which is (A.12)$=0$) we have

$$\overline{\omega}\cdot\ell+\varepsilon^2\mathbb{A}\xi\cdot\ell+\lambda(j)(1+\varepsilon^2\mathfrak{l}_j)-\lambda(k)(1+\varepsilon^2\mathfrak{l}_k)=\varepsilon^2(\mathbb{A}\xi\cdot\ell-(\overline{\omega}\cdot\ell)\vec{v}\cdot\xi+(w_j-w_k)\cdot\xi), \tag{A.13}$$

where $\vec{v}$ is in (8.28) and $\kappa_j=w_j\cdot\xi$ with κ_j in (7.5). These are a finite number (depending only on ν) of linear functions of ξ. We compute the derivative in ξ which is

$$(\mathbb{A}^T+\vec{v}\overline{\omega}^T)\ell+(w_{\widehat{\jmath}}-w_k), \tag{A.14}$$

where $\widehat{\jmath}\in\{\widehat{\jmath}_1,\widehat{\jmath}_2,\widehat{\jmath}_3,\widehat{\jmath}_4\}$. Now (A.5) implies that the quantity (A.14) is bounded from below by a constant depending on S. This lower bound and Fubini Theorem imply that $|T_{\ell jk}|\leq C(S)\varepsilon^{2(\nu-1)}\gamma$ for some $C(S)>0$ depending on S. By the discussion above we have

$$|\Omega_\varepsilon\backslash\mathcal{G}_0^{(1)}|\leq\sum_{|\ell|\leq 3,|j|,|k|\leq K(S)}|T_{\ell jk}|\leq C(S)\varepsilon^{2(\nu-1)}\gamma,$$

where $K(S)>0$ and $C(S)>0$ are constant depending on the set S. This implies the thesis. □

Appendix B. Normal form identification

Proof of Theorem 7.9. The core of Theorem 7.9 is to show that the terms in the l.h.s. of (7.78), which are obtained through a rather complicated sets of bounded changes of coordinates, coincide with the ones obtained by a purely formal full Birkhoff Normal Form procedure. In [32] it has been shown that, at purely formal level, the latter is well-defined and not resonant, i.e. the resonant Hamiltonian is supported only on trivial resonances as in (3.11). We procede as follows.

Step 1 The first step is to show that resonant terms at order ε^2 of the Full Birkhoff normal form coincide with the ones obtained by using the *weak* BNF procedure in Sect. 3, passing to action-angle variables and finally using a *formal linear* BNF.
Step 2 In order to conclude we note that the bounded maps we applied in Sects. 7.1 and 7.2 are, as functions of ε, C^3 with values in $\mathcal{L}(H^s, H^{s-3})$. Therefore the Taylor expansion of the Hamiltonian associated to the operators in (7.69) coincides with the Lie series of the generator up to order ε^2 (see also (7.71)). Then we show that the Lie series coincides (up to order ε^2) to the one obtained in step 1. Even though we taylor our proof to the particular set of changes of variables that we use in Sect. 7, the argument is quite general and is essentially that the *linear* BNF up to order ε^2 is coordinate independent.

Formal equivalence between weak + linear *and* full *BNF procedure.* One step of formal BNF means that we apply the formal change of variables generated by

$$\mathfrak{F}^{(3)} := [\mathrm{ad}_{H^{(2)}}]^{-1} H^{(3)} \tag{B.1}$$

which removes completely $H^{(3)}$ and conjugates H (using (2.5)) to $H \circ \Phi_{(\mathfrak{F}^{(3)})^{-1}} = H^{(2)} + \frac{1}{2}\{\mathfrak{F}^{(3)}, H^{(3)}\}+$ h.o.t.
The scecond step of formal BNF removes all the non-resonant terms of degree four thus we get

$$H^{(2)} + \frac{1}{2}\Pi_{\mathrm{Ker}(H^{(2)})}\{\mathfrak{F}^{(3)}, H^{(3)}\} + \text{h.o.t.}$$

In [32] it has been proved that $\left(\Pi_{\mathrm{Ker}(H^{(2)})} - \Pi_{\mathrm{triv}}\right)\{\mathfrak{F}^{(3)}, H^{(3)}\} = 0$ (see the notation of Definition 7.8) and hence

$$\Pi^{d_z=2}\Pi_{\mathrm{Ker}(H^{(2)})}\{\mathfrak{F}^{(3)}, H^{(3)}\} = \Pi^{d_z=2}\Pi_{\mathrm{triv}}\{\mathfrak{F}^{(3)}, H^{(3)}\}. \tag{B.2}$$

We claim that

$$\begin{aligned}\Pi_{\mathrm{triv}}\{\mathfrak{F}^{(3,1)}, H^{(3,3)}\} &= \Pi_{\mathrm{triv}}\{\mathfrak{F}^{(3,3)}, H^{(3,1)}\} = \Pi_{\mathrm{triv}}\{\mathfrak{F}^{(3,2)}, H^{(3,0)}\} \\ &= \Pi_{\mathrm{triv}}\{\mathfrak{F}^{(3,0)}, H^{(3,2)}\} = 0.\end{aligned} \tag{B.3}$$

This implies that

$$\begin{aligned}&\Pi_{\mathrm{Ker}(H^{(2)})}\Pi^{d_z=2}\{\mathfrak{F}^{(3)}, H^{(3)}\} \\ &= \Pi_{\mathrm{triv}}\Pi^{d_z=2}\Big(\{\mathfrak{F}^{(3,\le 1)}, H^{(3,\le 1)}\} + \{\mathfrak{F}^{(3,\le 1)}, H^{(3,>1)}\}\Big) \\ &\quad + \Pi_{\mathrm{triv}}\Pi^{d_z=2}\Big(\{\mathfrak{F}^{(3,2)}, H^{(3,0)}\} + \{\mathfrak{F}^{(3,2)}, H^{(3,2)}, \} + \{\mathfrak{F}^{(3,3)}, H^{(3,1)}\}\Big) \\ &= \Pi_{\mathrm{triv}}\Pi^{d_z=2}\Big(\{\mathfrak{F}^{(3,\le 1)}, H^{(3,\le 1)}\} + \{\mathfrak{F}^{(3,2)}, H^{(3,2)}\}\Big).\end{aligned} \tag{B.4}$$

Proof of the claim (B.3). Let us consider the term $\{\mathfrak{F}^{(3,1)}, H^{(3,3)}\}$, where

$$\mathfrak{F}^{(3,1)} = \sum_{\substack{j_1, j_2 \in S,\, j_3 \in S^c \\ j_1+j_2+j_3=0}} C_{j_1 j_2 j_3} u_{j_1} u_{j_2} u_{j_3}, \qquad H^{(3,3)} = \sum_{\substack{j_1, j_2, j_3 \in S^c \\ j_1+j_2+j_3=0}} \widetilde{C}_{j_1 j_2 j_3} u_{j_1} u_{j_2} u_{j_3},$$

for some coefficients $C_{j_1 j_2 j_3}, \widetilde{C}_{j_1 j_2 j_3} \in \mathbb{C}$. Using (2.2) one gets

$$\{\mathfrak{F}^{(3,1)}, H^{(3,3)}\} = \sum_{\substack{j_1, j_2 \in S,\, k_1, k_2 \in S^c \\ j_1 + j_2 + k_1 + k_2 = 0}} P_{j_1 j_2 k_1 k_2} u_{j_1} u_{j_2} u_{k_1} u_{k_2}, \qquad P_{j_1 j_2 k_1 k_2} \in \mathbb{C}. \tag{B.5}$$

A monomial $u_{j_1} u_{j_2} u_{k_1} u_{k_2}$ is supported on trivial resonances (3.11) only if $j_1 = -j_2$ and $k_1 = -k_2$, since $j_i \in S$, $k_i \in S^c$, $i = 1, 2$. On the other hand, by the momentum condition, we have that $j_1 + j_2 = -(k_1 + k_2)$, which is not possible since $0 \notin S \cup S^c$ (see (1.9)). Hence the (B.3) holds. The others equalities in (B.3) in follows in the same way. □

Let us now perform the same Birkhoff procedure by first cancelling the terms of degree ≤ 1 in z (weak BNF) and then the terms of degree two (linear BNF).

By the discussion in Sect. 3, recall the notations of Proposition 3.6, we have that, after two steps of weak Birkhoff normal form, the Hamiltonian of degree less or equal than 4 is

$$\Pi^{d \leq 4}(H \circ \Phi_2^{-1}) = H^{(2)} + Z_2^{(4,0)} + H_2^{(3,\geq 2)} + H_2^{(4,\geq 2)}. \tag{B.6}$$

Here $Z_2^{(4,0)}$ is defined in (3.23),

$$H_2^{(3,2)} = H_1^{(3,2)} = H^{(3,2)}, \qquad H_2^{(4,\geq 2)} = H_1^{(4,\geq 2)}, \tag{B.7}$$

where $H_1^{(3)}$, $H_1^{(4)}$ are defined in (3.19). The monomials of degree greater than 4 will be not involved in this computation, so we omit them. It is important to notice that, by direct inspection, $\mathfrak{F}^{(3,\leq 1)} = F^{(3,\leq 1)}$ defined in formula (3.21). By Proposition 3.6, we know that the same change of variables puts one of the constants of motion (lets say K_3 and drop the subindex 3) into normal form,

$$\Pi^{d \leq 4}(K \circ \Phi_2^{-1}) = K^{(2)} + W_2^{(4,0)} + K_2^{(3,\geq 2)} + K_2^{(4,\geq 2)}.$$

The step of formal linear BNF entails applying the formal change of variables generated by

$$F^{(3,2)} := [\mathrm{ad}_{H^{(2)}}]^{-1} H^{(3,2)}. \tag{B.8}$$

Again, by direct inspection, one can note that $F^{(3,2)} \equiv \mathfrak{F}^{(3,2)} := \Pi^{d_x = 2} \mathfrak{F}^{(3)}$, where $\mathfrak{F}^{(3)}$ is given in (B.1). Thus obtaining the Hamiltonian

$$H^{(2)} + Z_2^{(4,0)} + H_2^{(3,3)} + \frac{1}{2}\{\mathfrak{F}^{(3,2)}, H_2^{(3,\geq 2)}\} + H_2^{(4,\geq 2)} + h.o.t. \tag{B.9}$$

Since (B.8) puts in normal form also $K \circ \Phi_2^{-1}$ following the same reasoning as in [32] and Proposition 3.6, we get that

$$\begin{aligned}
&\Pi^{d_z = 2} \Pi_{\mathrm{Ker}(H^{(2)})} \Big(\frac{1}{2}\{\mathfrak{F}^{(3,2)}, H_2^{(3,\geq 2)}\} + H_2^{(4,\geq 2)} \Big) \\
&= \Pi^{d_z = 2} \Pi_{\mathrm{triv}} \Big(\frac{1}{2}\{\mathfrak{F}^{(3,2)}, H_2^{(3,\geq 2)}\} + H_2^{(4,\geq 2)} \Big) \\
&= \Pi^{d_z = 2} \Pi_{\mathrm{triv}} \Big(\frac{1}{2}\{\mathfrak{F}^{(3,2)}, H_2^{(3,2)}\} + \frac{1}{2}\{\mathfrak{F}^{(3,\leq 1)}, H^{(3,\leq 1)}\} + \{\mathfrak{F}^{(3,1)}, H^{(3,3)}\} \Big) \\
&\overset{(B.3)}{=} \frac{1}{2} \Pi^{d_z = 2} \Pi_{\mathrm{triv}} \Big(\{\mathfrak{F}^{(3,2)}, H_2^{(3,2)}\} + \frac{1}{2}\{\mathfrak{F}^{(3,\leq 1)}, H^{(3,\leq 1)}\} \Big)
\end{aligned}$$

$$\stackrel{(B.4),(B.2)}{=} \frac{1}{2}\Pi^{d_z=2}\Pi_{\mathrm{triv}}\{\mathfrak{F}^{(3)}, H^{(3)}\}. \tag{B.10}$$

We now want to pass to the action-angle variables introduced in (4.7). Since the rescaling with the parameter ε is covariant under the change of variables that we use, we consider instead of A_ε the symplectic change of variables $A_1 := A_{\varepsilon_{|\varepsilon=1}}$. By recalling that

$$\mathtt{H}_0 := H^{(2)} \circ A_{1|_{\substack{y=0\\ \theta=\varphi}}}, \quad \mathtt{H}_1 := H_2^{(3,2)} \circ A_{1|_{\substack{y=0\\ \theta=\varphi}}}, \quad \Pi^{d_z=2}\big(\mathtt{H}_2 + \mathtt{H}_{\mathcal{R}_2}\big) := H_2^{(4,2)} \circ A_{1|_{\substack{y=0\\ \theta=\varphi}}}$$

and setting

$$\widetilde{F}^{(3,2)} := \mathfrak{F}^{(3,2)} \circ A_{1|_{\substack{y=0\\ \theta=\varphi}}}, \tag{B.11}$$

we have that (B.10) reads

$$\Pi^{d_z=2}\Pi_{\mathrm{Ker}(\mathtt{H}_0)}\Big(\frac{1}{2}\{\widetilde{F}^{(3,2)}, \mathtt{H}_1\} + \mathtt{H}_2 + \mathtt{H}_{\mathcal{R}_2}\Big) = \frac{1}{2}\Big[\Pi^{d_z=2}\Pi_{\mathrm{triv}}\{\mathfrak{F}^{(3)}, H^{(3)}\}\Big] \circ A_{1|_{\substack{y=0\\ \theta=\varphi}}}. \tag{B.12}$$

The rigorous *procedure of subsections* 7.1, 7.2 *and the* linear *BNF.* Since the r.h.s. of (B.12) is the r.h.s. of (7.78) it remains to show that the ε^2-terms of the Hamiltonian associated to the operator $\mathcal{L}_7$ in (7.69) coincides with the l.h.s. in (B.12). We remark that the operator $\mathcal{L}_7$ has been constructed through a rigorous procedure providing also "tame" estimates of the remainder of higher order in ε. As already explained the maps $\mathcal{B}$, Υ_i, $i = 1, 2$, are, as functions of ε, C^3 with values in $\mathcal{L}(H^s, H^{s-3})$. Therefore the Taylor expansion of the Hamiltonian associated to the operators in (7.69) coincides with the Lie series of the generator up to order ε^2 (see also (7.71)).

Let us then Taylor expand the Hamiltonian $\mathcal{K}^{(2)} = \mathtt{H} \circ \mathcal{B}^{-1} \circ \Upsilon_1^{-1} \circ \Upsilon_2{}^{-1}$ (see (7.56), (7.57), (7.65) and (7.76)) up to order ε^2. It is sufficient to consider $\mathcal{B}_1$ the flow generated by the Hamiltonian in (7.20) and Υ_1 the flow of the Hamiltonian $H_{\mathtt{A}_1}$ (which has the form (C.47) and satisfies (7.66)). We have that

$$\Pi_{\mathrm{Ker}(\mathtt{H}_0)}\Big(\mathtt{Z}_0 + \mathcal{K}_2^{(1)}\Big) \stackrel{(7.76),(7.77)}{=} \Pi_{\mathrm{Ker}(\mathtt{H}_0)}\mathtt{H} \circ \mathcal{B}_1^{-1} \circ \Upsilon_1^{-1} + O(\varepsilon^3) \tag{B.13}$$

$$\begin{aligned}
&\stackrel{(7.43),(7.65)}{=} \mathtt{H}_0 + \varepsilon^2\Pi_{\mathrm{Ker}(\mathtt{H}_0)}\Big(\mathtt{H}_2^{(2)} + \frac{1}{2}\{H_{\mathtt{A}_1}, \{H_{\mathtt{A}_1}, \mathtt{H}_0\}_e\}_e + \{H_{\mathtt{A}_1}, \mathcal{K}_1\}_e\Big) + O(\varepsilon^3)\\
&\stackrel{(7.30),(7.43),(7.22)}{=} \mathtt{H}_0 + \varepsilon^2\Pi_{\mathrm{Ker}(\mathtt{H}_0)}\Big(\mathtt{H}_2^{(1)} + \frac{1}{2}\{H_{\mathtt{A}_1}, \{H_{\mathtt{A}_1}, \mathtt{H}_0\}_e\}_e\\
&\quad + \{H_{\mathtt{A}_1}, \mathtt{H}_1\}_e + \{H_{\mathtt{A}_1}, \{S_1, \mathtt{H}_0\}_e\}_e\Big) + O(\varepsilon^3).\\
&\stackrel{(7.23)}{=} \mathtt{H}_0 + \varepsilon^2\Pi_{\mathrm{Ker}(\mathtt{H}_0)}\Big(\frac{1}{2}\{S_1, \{S_1, \mathtt{H}_0\}_e\}_e + \{S_1, \mathtt{H}_1\}_e + \frac{1}{2}\{S_2, \mathtt{H}_0\}_e + \mathtt{H}_2 + \mathtt{H}_{\mathcal{R}_2}\\
&\qquad + \frac{1}{2}\{H_{\mathtt{A}_1}, \{H_{\mathtt{A}_1}, \mathtt{H}_0\}_e\}_e + \{H_{\mathtt{A}_1}, \mathtt{H}_1\}_e + \{H_{\mathtt{A}_1}, \{S_1, \mathtt{H}_0\}_e\}_e\Big) + O(\varepsilon^3).
\end{aligned} \tag{B.14}$$

Using the Jacobi identity we have

$$\frac{1}{2}\{S_1, \{S_1, \mathtt{H}_0\}_e\}_e + \frac{1}{2}\{H_{\mathtt{A}_1}, \{H_{\mathtt{A}_1}, \mathtt{H}_0\}_e\}_e + \{H_{\mathtt{A}_1}, \{S_1, \mathtt{H}_0\}_e\}_e$$

$$= \frac{1}{2}\{S_1 + H_{\mathbb{A}_1}, \{S_1 + H_{\mathbb{A}_1}, \mathsf{H}_0\}_e\}_e + \frac{1}{2}\{\mathsf{H}_0, \{S_1, S_1 + H_{\mathbb{A}_1}\}\}_e. \tag{B.15}$$

Moreover, setting $\mathsf{F}_1 := S_1 + H_{\mathbb{A}_1}$, we note that

$$\{\mathsf{F}_1, \mathsf{H}_0\}_e = \{H_{\mathbb{A}_1}, \mathsf{H}_0\}_e + \{S_1, \mathsf{H}_0\}_e \overset{(7.22),(7.43),(7.66)}{=} -\mathsf{H}_1. \tag{B.16}$$

Since $\Pi_{\mathrm{Ker}(\mathrm{H}_0)}\{S_2, \mathsf{H}_0\}_e \equiv 0$, by (B.13), (B.15), and (B.16) we get

$$\Pi^{d_w=2}\Pi_{\mathrm{Ker}(\mathrm{H}_0)}\Big(\mathbb{Z}_0 + \mathcal{K}_2^{(1)}\Big) = \Pi^{d_w=2}\Pi_{\mathrm{Ker}(\mathrm{H}_0)}\Big(\frac{1}{2}\{\mathsf{F}_1, \mathsf{H}_1\}_e + \mathsf{H}_2 + \mathsf{H}_{\mathcal{R}_2}\Big). \tag{B.17}$$

Since $\mathrm{Ker}(H^{(2)})$ is trivial on cubic monomials, we deduce that

$$\mathsf{F}_1 = -\mathrm{ad}_{\mathsf{H}_0}^{-1}(\mathsf{H}_1) = \widetilde{F}^{(3,2)} \tag{B.18}$$

and this concludes the proof. □

Appendix C. Technical Lemmata

C.1 Flows of hyperbolic PDEs. In this subsection we study some properties of the flow of (7.11). We start by studying the time-one flow map of the pseudo differential PDE

$$\partial_\tau \Psi^\tau(u) = (J \circ b)\Psi^\tau(u), \qquad \Psi^0 u = u, \tag{C.1}$$

and how this differs from Φ^τ. Proposition 3.1 in [33] guarantees that the flow of (C.1) is the composition of the diffeomorphism of the torus

$$\begin{aligned} &\mathcal{A}^\tau h(\varphi, x) := (1 + \tau\beta_x(\varphi, x))h(\varphi, x + \tau\beta(\varphi, x)), \qquad \varphi \in \mathbb{T}^\nu, x \in \mathbb{T}, \\ &(\mathcal{A}^\tau)^{-1} h(\varphi, y) := (1 + \tilde\beta_y(\tau, \varphi, y))\, h(\varphi, y + \tilde\beta(\tau, \varphi, y)), \quad \varphi \in \mathbb{T}^\nu, y \in \mathbb{T}, \end{aligned} \tag{C.2}$$

where $\tilde\beta(\tau; x, \xi)$ is such that

$$x \mapsto y = x + \tau\beta(\varphi, x) \;\Leftrightarrow\; y \mapsto x = y + \tilde\beta(\tau, \varphi, x), \quad \tau \in [0, 1],$$

with a pseudo differential operator of order -1 up to smoothing remainders belonging to the class $\mathfrak{L}_{\rho,p}$ (see Definition 2.8 in [33]) for any $\rho \in \mathbb{N}$, $\rho \geq 3$, $p \geq s_0$. For completeness we restate the definition of $\mathfrak{L}_{\rho,p}$ in Appendix C.2 (see Definition C.5). The class $\mathfrak{L}_{\rho,p}$ has the property of being closed w.r.t. changes of variables as Φ^τ (see Lemma $B.10$ in [33]). Moreover, in Lemma C.8, we show that $\mathfrak{L}_{\rho,p}$ is contained into $\mathfrak{C}_{-1}$ which is another class of "tame" operators. Such class is introduced in Definition C.6 and is included in -1-tame operators, see Lemma C.10.

We refer to Proposition 3.1, Corollary 3.2 and Appendix A in [33] for properties and estimates of Ψ^τ and $\mathcal{A}^\tau$ in (C.2).

Lemma C.1. *Fix $\rho \geq 3$ and $p \geq s_0$. There exist $\delta \ll 1$, $\sigma_1 := \sigma_1(\rho, p)$ such that if*

$$\|\beta\|_{s_0+\sigma_1} \leq \delta \tag{C.3}$$

then the following holds. Let Ψ^τ be the flow of the system (C.1), then the flow of (7.11) Φ^τ is well defined for $|\tau| \leq 1$ and one has $\Phi^1 = \Pi_S^\perp \Psi^1 \Pi_S^\perp \circ (\mathrm{I} + \mathcal{R})$ where $\mathcal{R}$ is an operator with the form (6.16). Moreover $\mathcal{R}$ belongs to $\mathfrak{L}_{\rho,p}(\mathcal{O})$ and satisfies

$$\mathbb{M}_{\mathcal{R}}^\gamma(s, \mathfrak{b}) \lesssim_s \|\beta\|_{s+\sigma_1}^{\gamma,\mathcal{O}}, \qquad \mathbb{M}_{\Delta_{12}\mathcal{R}}(p, \mathfrak{b}) \lesssim_p \|\Delta_{12}\beta\|_{p+\sigma_1}. \tag{C.4}$$

Proof. Proposition 3.1 in [33] provides δ, σ_1 such that if the smallness condition (C.3) is satisfied with such parameters, then the flow Ψ^τ is well-defined for $|\tau| \leq 1$. Let us define Υ^τ as the flow of the following Cauchy problem

$$\partial_\tau \Upsilon^\tau u = -\big(\Psi_*^\tau \mathtt{Z}\big)\Upsilon^\tau u, \qquad \Upsilon^0 u = u \tag{C.5}$$

with

$$\begin{aligned}
\mathtt{Z}u &:= (J \circ b)\Pi_S[u] + \Pi_S(J \circ b)\Pi_S^\perp[u] = \sum_{j\in S} \big(g_j(\tau), u\big)_{L^2}\, \chi_j(\tau) \\
&\quad + \sum_{j\in S} \big(\tilde g_j(\tau), u\big)_{L^2}\, \tilde\chi_j(\tau), \\
g_j = \tilde\chi_j &:= e^{\mathrm{i}jx}, \quad \chi_j = J(b(\tau)\, e^{\mathrm{i}jx}), \quad \tilde g_j := \lambda(j)\Pi_S^\perp[b(\tau)\, e^{\mathrm{i}jx}].
\end{aligned}$$

Equation (C.5) is well posed on H^s since its vector field is finite rank. By the following computations

$$\begin{aligned}
&(\partial_\tau \Psi^\tau)(\Upsilon^\tau u) + \Psi^\tau(\partial_\tau \Upsilon^\tau u) \\
&\quad = (J \circ b)\Pi_S^\perp[\Psi^\tau(\Upsilon^\tau u)] - \Pi_S[(J \circ b)\Pi_S^\perp[\Psi^\tau(\Upsilon^\tau u)] \\
&\quad = (J \circ b)\Psi^\tau(\Upsilon^\tau u) - \Big((J \circ b)\Pi_S[\Psi^\tau(\Upsilon^\tau u)] + \Pi_S[(J \circ b)\Pi_S^\perp[\Psi^\tau(\Upsilon^\tau u)]\Big)
\end{aligned}$$

we have that $\Phi^\tau = \Pi_S^\perp \circ \Psi^\tau \circ \Upsilon^\tau$ is well-defined on H^s and solves (7.11).
Now we show that $\Upsilon^\tau - \mathrm{I}$ is of the form (6.16). By Taylor expansion at $\tau = 0$ we get

$$\Upsilon^\tau u - u = -\tau\big(\Psi_*^\tau \mathtt{Z}(\Upsilon^\tau u)\big)_{|_{\tau=0}} + \int_0^\tau (1-t)\big(\partial_{tt}\Upsilon^t(u)\big)\, dt. \tag{C.6}$$

Note that

$$\Psi_*^\tau \mathtt{Z}(\Upsilon^\tau u) = \sum_{j\in S} \big((\Phi^\tau)^* g_j(\tau), u\big)_{L^2}\, (\Psi^\tau)^{-1}\chi_j(\tau) + \sum_{j\in S} \big((\Phi^\tau)^* \tilde g_j(\tau), u\big)_{L^2}\, (\Psi^\tau)^{-1}\tilde\chi_j(\tau)$$

has already the form (6.16) and $\big(\Psi_*^\tau \mathtt{Z}(\Upsilon^\tau u)\big)_{|_{\tau=0}} = \mathtt{Z}u$. We denoted by $(\Psi^\tau)^*$, $(\Phi^\tau)^*$ the flows of the adjoint PDEs

$$\partial_\tau (\Psi^\tau)^* u = -bJ((\Psi^\tau)^* u), \qquad \partial_\tau (\Phi^\tau)^* u = -\Pi_S^\perp\, b\, J\, \Pi_S^\perp[(\Phi^\tau)^* u].$$

We have

$$\begin{aligned}
\int_0^\tau (1-t)\partial_{tt}\Upsilon^t(u)\, ds &= \sum_{j\in S} \int_0^\tau (1-t)\big(\mathtt{g}_j(t), u\big)_{L^2}\, \mathtt{f}_j(t)\, dt \\
&\quad + \sum_{j\in S} \int_0^\tau (1-t)\big(\tilde{\mathtt{g}}_j(t), u\big)_{L^2}\, \tilde{\mathtt{f}}_j(t)\, dt
\end{aligned}$$

where

$$\begin{aligned}
\mathtt{g}_j &:= \big(- (\Psi_*^s \mathtt{Z})^*(\Phi^s)^* - (\Upsilon^s)^* b\, J((\Psi^s)^*)\big)(g_j), \\
\mathtt{f}_j &:= -(J \circ b)((\Psi^s)^{-1})\chi_j + (\Psi^s)^{-1} J(b'(s) e^{\mathrm{i}jx}),
\end{aligned}$$

$$\tilde{\mathtt{g}}_j := \big(- (\Psi^s_* \mathtt{Z})^*(\Phi^s)^* - (\Upsilon^s)^* b\, J((\Psi^s)^*)\big)(\tilde{g}_j) + (\Phi^s)^* \lambda(j) \Pi_S^\perp [b'(s) e^{\mathrm{i}jx}],$$
$$\tilde{\mathtt{f}}_j := -(J \circ b)((\Psi^s)^{-1}) \tilde{\chi}_j .$$

Thus by (C.6), for $\tau = 1$ we get

$$\Upsilon^1 u - u = \mathcal{R} u, \qquad \mathcal{R} u := \mathcal{R}_1 u + \mathcal{R}_2 u, \tag{C.7}$$

$$\mathcal{R}_1 u := -\mathtt{Z} u, \qquad \mathcal{R}_2 u := \int_0^1 (1-t) \partial_{tt} \Upsilon^t(u)\, dt, \tag{C.8}$$

where $\mathcal{R}_1$ has the finite dimensional form (6.16) and $\mathcal{R}_2$ has the form (6.16). Hence by Lemma C.7 we have

$$\mathbb{M}^\gamma_{\mathcal{R}_1}(s, \mathtt{b}) \lesssim_s \sum_{|j| \le C} \sup_{\tau \in [0,1]} (\|\mathtt{f}_j(\tau)\|^{\gamma, \mathcal{O}_0}_s \|\mathtt{g}_j(\tau)\|^{\gamma, \mathcal{O}_0}_{s_0} + \|\mathtt{f}_j(\tau)\|^{\gamma, \mathcal{O}_0}_{s_0} \|\mathtt{g}_j(\tau)\|^{\gamma, \mathcal{O}_0}_s), \tag{C.9}$$

$$\mathbb{M}^\gamma_{\mathcal{R}_2}(s, \mathtt{b}) \lesssim_s \sum_{|j| \le C} \sup_{\tau \in [0,1]} (\|\tilde{\mathtt{f}}_j(\tau)\|^{\gamma, \mathcal{O}_0}_s \|\tilde{\mathtt{g}}_j(\tau)\|^{\gamma, \mathcal{O}_0}_{s_0} + \|\tilde{\mathtt{f}}_j(\tau)\|^{\gamma, \mathcal{O}_0}_{s_0} \|\tilde{\mathtt{g}}_j(\tau)\|^{\gamma, \mathcal{O}_0}_s). \tag{C.10}$$

By using the estimates in Proposition 3.1 and Corollary 3.2 in [33] we have

$$\|\mathtt{f}_j\|^{\gamma, \mathcal{O}_0}_s, \|\tilde{\mathtt{g}_j}\|^{\gamma, \mathcal{O}_0}_s \lesssim_s \|b\|^{\gamma, \mathcal{O}_0}_{s+s_0+2} + \|\partial_\tau b(\tau)\|^{\gamma, \mathcal{O}_0}_{s+1}, \quad \|\tilde{\mathtt{f}_j}\|^{\gamma, \mathcal{O}_0}_s, \|\mathtt{g}_j\|^{\gamma, \mathcal{O}_0}_s \lesssim_s \|b\|^{\gamma, \mathcal{O}_0}_{s+s_0+2}. \tag{C.11}$$

In the same way, the bounds for the variation on the i-variable (the second in (C.4)) follows by the estimates on the derivatives of the coefficients $\mathtt{g}_j, \tilde{\mathtt{g}_j}, \chi_j, \tilde{\chi}_j$ whose depend on the variation Δ_{12} of the flows Φ^τ, Ψ^τ and their adjoints. We have proved that $\Upsilon^1 u = u + \mathcal{R} u$ and hence $\Phi^1 u = \Psi^1 \circ (\mathrm{I} + \mathcal{R})\mathrm{u}$. By (C.9), (C.10) and (C.11) we obtain (C.4). □

The system (7.11) is a Hyperbolic PDE, thus we shall use a version of Egorov Theorem to study how pseudo differential operators change under the flow Φ^τ. This is the content of Theorem 3.4 in [33] which provides precise estimates for the transformed pseudo differential operators.

The following proposition is the counter-part of Proposition 3.5 in [33]. It describes the structure of an operator like $\mathcal{L}_\omega$ conjugated by a flow of a system like (7.11).

Proposition C.2. (Conjugation). *Let $\mathcal{O}$ be a subset of $\mathbb{R}^\nu$. Fix $\rho \ge 3$, $\alpha \in \mathbb{N}$, $p \ge s_0$ and consider a linear operator*

$$\mathcal{L} := \Pi_S^\perp \Big(\omega \cdot \partial_\varphi - J \circ (m + a(\varphi, x)) + \mathcal{Q} \Big) \tag{C.12}$$

where $m = m(\omega)$ is a real constant, $a = a(\omega, i(\omega)) \in C^\infty(\mathbb{T}^{\nu+1})$ is real valued, both are Lipschitz in $\omega \in \mathcal{O}$ and a is Lipschitz in the variable i. Moreover $\mathcal{Q} = \mathrm{Op}(\mathtt{q}(\varphi, x, \xi)) + \widehat{\mathcal{Q}}$ with $\widehat{\mathcal{Q}} \in \mathfrak{L}_{\rho, p}(\mathcal{O})$ and $\mathtt{q} = \mathtt{q}(\omega, i(\omega)) \in S^{-1}$ satisfying

$$|\mathtt{q}|^{\gamma, \mathcal{O}}_{-1, s, \alpha} \lesssim_{s, \alpha} \mathtt{k}_1 + \mathtt{k}_2 \|f\|^{\gamma, \mathcal{O}}_{s+\sigma}, \tag{C.13}$$

$$|\Delta_{12} \mathtt{q}|_{-1, p, \alpha} \lesssim_{p, \alpha} \mathtt{k}_3 (\|\Delta_{12} f\|_{p+\sigma} + \|\Delta_{12} f\|_{p+\sigma} \|f\|_{p+\sigma}). \tag{C.14}$$

Here $\mathtt{k}_1, \mathtt{k}_2, \mathtt{k}_3, \sigma > 0$ *are constants depending on* $\mathtt{q}$ *while* $f = f(\omega, i(\omega)) \in C^\infty(\mathbb{T}^{\nu+1})$*, is Lipschitz in* ω *and in the variable* i *. There exist* $\tilde\sigma = \tilde\sigma(\rho) > 0$ *and* $\delta_* := \delta_*(\rho)$ *such that, if*

$$\|\beta\|^{\gamma,\mathcal{O}}_{s_0+\tilde\sigma} + \|a\|^{\gamma,\mathcal{O}}_{s_0+\tilde\sigma} + \mathtt{k}_2\|f\|^{\gamma,\mathcal{O}}_{s_0+\tilde\sigma} + \mathtt{k}_1 + \mathbb{M}^\gamma_{\widehat{\mathcal{Q}}}(s_0, \mathtt{b}) \leq \delta_*, \tag{C.15}$$

the following holds for $p + \sigma \leq s_0 + \tilde\sigma$*. Consider* $\Phi := \Phi^1$ *the flow at time one of the system* (7.11)*, where b is defined in* (7.10)*. Then we have*

$$\mathcal{L}_+ := \Phi\mathcal{L}\Phi^{-1} = \Pi^\perp_S\Big(\omega\cdot\partial_\varphi - J\circ(m + a_+(\varphi, x)) + \mathcal{Q}_+\Big) \tag{C.16}$$

where

$$m+a_+(\varphi, x) := -(\omega\cdot\partial_\varphi\tilde\beta)(\varphi, x+\beta(\varphi, x))+(m+a(\varphi, x+\beta(\varphi, x)))(1+\tilde\beta_x(\varphi, x+\beta(\varphi, x))) \tag{C.17}$$

with $\tilde\beta$ *the function such that* $x + \tilde\beta(\varphi, x)$ *is the inverse of the diffeomorphism of the torus* $x \mapsto x + \beta(\varphi, x)$*. The operator* $\mathcal{Q}_+ := \mathrm{Op}(\mathtt{q}_+(\varphi, x, \xi)) + \widehat{\mathcal{Q}}_+$*, with*

$$\begin{aligned} |\mathtt{q}_+|^{\gamma,\mathcal{O}}_{-1,s,\alpha} &\lesssim_{s,\alpha,\rho} \mathtt{k}_1 + \mathtt{k}_2\|f\|^{\gamma,\mathcal{O}}_{s+\tilde\sigma} + \|\beta\|^{\gamma,\mathcal{O}}_{s+\tilde\sigma} + \|a\|^{\gamma,\mathcal{O}}_{s+\tilde\sigma},\\ |\Delta_{12}\mathtt{q}_+|_{-1,p,\alpha} &\lesssim_{p,\alpha,\rho} \mathtt{k}_3(\|\Delta_{12}f\|_{p+\tilde\sigma} + \|\Delta_{12}f\|_{p+\tilde\sigma}\|f\|_{p+\tilde\sigma})\\ &\quad + \|\Delta_{12}\beta\|_{p+\tilde\sigma} + \|\Delta_{12}a\|_{p+\tilde\sigma} \end{aligned} \tag{C.18}$$

and $\widehat{\mathcal{Q}}_+ \in \mathfrak{L}_{\rho,p}(\mathcal{O})$ *with, for* $s_0 \leq s \leq \mathcal{S}$*,*

$$\mathbb{M}^\gamma_{\widehat{\mathcal{Q}}_+}(s, \mathtt{b}) \lesssim_{s,\rho} \mathbb{M}^\gamma_{\widehat{\mathcal{Q}}}(s, \mathtt{b}) + \|\beta\|^{\gamma,\mathcal{O}}_{s+\tilde\sigma} + \mathtt{k}_1 + \mathtt{k}_2\|f\|^{\gamma,\mathcal{O}}_{s+\tilde\sigma} + \|a\|^{\gamma,\mathcal{O}}_{s+\tilde\sigma}, \tag{C.19}$$

for any $0 \leq \mathtt{b} \leq \rho - 2$ *and*

$$\begin{aligned} \mathbb{M}_{\Delta_{12}\widehat{\mathcal{Q}}_+}(p, \mathtt{b}) &\lesssim_{p,\rho} \mathbb{M}_{\Delta_{12}\widehat{\mathcal{Q}}}(p, \mathtt{b})\\ &\quad + \mathtt{k}_3(\|\Delta_{12}f\|_{p+\tilde\sigma} + \|\Delta_{12}f\|_{p+\tilde\sigma}\|f\|_{p+\tilde\sigma}) + \|\Delta_{12}\beta\|_{p+\tilde\sigma} + \|\Delta_{12}a\|_{p+\tilde\sigma} \end{aligned} \tag{C.20}$$

for any $0 \leq \mathtt{b} \leq \rho - 3$*.*

Proof. The strategy is the following.

- We conjugate
$$\mathcal{L}^0 := \omega\cdot\partial_\varphi - J\circ(m + a(\varphi, x)) + \mathcal{Q} \tag{C.21}$$
by the flow Ψ^τ in (C.1). In order to do this we use Proposition 3.5 in [33].
- The operator $\mathcal{L}^0$ differs from $\mathcal{L}$ by a infinitely regularizing operator of the form (6.16). By using this fact and Lemma C.1 we estimate the difference between $\Phi^\tau\mathcal{L}(\Phi^\tau)^{-1}$ and $\Psi^\tau\mathcal{L}^0(\Psi^\tau)^{-1}$.

Let Ψ^τ be the flow in (C.1). The (C.13)-(C.14) imply that $\mathcal{L}_0$ in (C.21) satisfies the hypotheses of Proposition 3.5 in [33]. Hence, if $y := x + \beta(\varphi, x)$, then we have

$$\begin{aligned} \mathcal{L}^0_+h &:= \Psi\mathcal{L}^0\Psi^{-1} = \omega\cdot\partial_\varphi h - J\{\big(m + a_+(\varphi, x)\big)\, h\} + \mathrm{Op}(\mathtt{q}_+) + \widehat{\mathcal{Q}}_\star h,\\ m + a_+(\varphi, x) &:= -\omega\cdot\partial_\varphi\tilde\beta(\varphi, y) + (m + a(\varphi, y))(1 + \tilde\beta_y(\varphi, y)). \end{aligned}$$

In particular $\mathfrak{q}_+$ and $\widehat{\mathcal{Q}}_\star$ satisfy bounds like (C.18)-(C.20). By Lemma C.1 we have (recall (C.7), (C.8))

$$\widehat{\mathcal{Q}}_{\star\star} := \Phi\mathcal{L}\Phi^{-1} - \Pi_S^\perp\mathcal{L}_+^0\Pi_S^\perp = \Pi_S^\perp\Psi\mathcal{L}^0\Psi^{-1}\Pi_S + \Pi_S\Psi\mathcal{L}^0\Psi^{-1} + \Psi\mathcal{L}\Gamma\Psi^{-1} + \Psi\mathcal{L}\mathcal{R}\Psi^{-1} + \Psi\mathcal{R}\mathcal{L}\Gamma\Psi^{-1} - \Psi\mathcal{L}^0\Pi_S\Psi^{-1} - \Psi\Pi_S\mathcal{L}\Pi_S^\perp\Psi^{-1}.$$

We define the remainder $\widehat{\mathcal{Q}}_+ := \widehat{\mathcal{Q}}_\star + \widehat{\mathcal{Q}}_{\star\star}$. To conclude the proof we show that $\widehat{\mathcal{Q}}_{\star\star}$ satisfies the bounds (C.19) and (C.20). We note that

$$\begin{aligned}
&\Pi_S^\perp\Psi\mathcal{L}^0\Psi^{-1}\Pi_S h = \sum_{j\in S}(h, g_j^{(1)})_{L^2}\chi_j^{(1)}, \quad g_j^{(1)} := e^{\mathrm{i}jx}, \quad \chi_j^{(1)} := \Psi\mathcal{L}^0\Psi^{-1}e^{\mathrm{i}jx},\\
&\Pi_S\Psi\mathcal{L}^0\Psi^{-1}h = \sum_{j\in S}(h, g_j^{(2)})_{L^2}\chi_j^{(2)}, \quad g_j^{(2)} := \Psi\mathcal{L}^0\Psi^{-1}e^{\mathrm{i}jx}, \quad \chi_j^{(2)} = e^{\mathrm{i}jx},\\
&\Psi\mathcal{L}^0\Pi_S\Psi^{-1} = \sum_{j\in S}(h, g_j^{(3)})_{L^2}\chi_j^{(3)}, \quad g_j^{(3)} := (\Psi^{-1})^* e^{\mathrm{i}jx}, \quad \chi_j^{(3)} := \Psi\mathcal{L}^0 e^{\mathrm{i}jx},\\
&\Psi\Pi_S\mathcal{L}\Pi_S^\perp\Psi^{-1}h = \sum_{j\in S}(h, g_j^{(4)})_{L^2}\chi_j^{(4)}, \quad g_j^{(4)} := (\mathcal{L}\Pi_S^\perp\Psi^{-1})^* e^{\mathrm{i}jx}, \quad \chi_j^{(4)} := \Psi e^{\mathrm{i}jx}.
\end{aligned} \tag{C.22}$$

Thus by Lemma C.7, Corollary 3.2 in [33] (for the estimates on Ψ) and (C.21) we get the bounds (C.19) and (C.20) for the operators (C.22). The bounds on $\Psi\mathcal{L}\Gamma\Psi^{-1}$, $\Psi\mathcal{L}\mathcal{R}\Psi^{-1}$, $\Psi\mathcal{R}\mathcal{L}\Gamma\Psi^{-1}$ follow by Proposition 3.1 in [33] and (C.4). □

Remark C.3. Assume that the symbols $a, \mathfrak{q}$ and the smoothing operator $\widehat{\mathcal{Q}}$ in (C.12) admits an expansion in ε (actually in degree of homogeneity of $\overline{v}$ in (6.25)) of the form

$$a = \sum_{i=1}^{3}\varepsilon^i a^{(i)} + a^{(\geq 4)}, \quad \mathfrak{q} = \sum_{i=1}^{3}\varepsilon^i \mathfrak{q}^{(i)} + \mathfrak{q}^{(\geq 4)}, \quad \widehat{\mathcal{Q}} = \sum_{i=1}^{3}\varepsilon^i \widehat{\mathcal{Q}}^{(i)} + \widehat{\mathcal{Q}}^{(\geq 4)}, \tag{C.23}$$

where $a^{(i)}, \mathfrak{q}^{(i)}$ have the form respectively (6.36), (6.34), and $a^{(\geq 4)}, \mathfrak{q}^{(\geq 4)}$ satisfy estimates like (7.42). The remainders $\widehat{\mathcal{Q}}^{(i)}$ are *almost-diagonal* (see Def. 6.4) and $\widehat{\mathcal{Q}}^{(\geq 4)}$ satisfies estimates like (7.42). Assume also that β in (7.10) has an expansion as in (C.23). Then the symbols $a^+, \mathfrak{q}^+$ and the operator $\mathcal{Q}^+$ in (C.16) admit the same expansion in ε as in (C.23). This fact can be deduced by following the proof of Proposition 3.5 in [33]. More precisely one reasons as follows. First of all, by linearity, the conjugate of a sum as in (C.23) is the sum of the conjugates. The conjugate of a $\widehat{\mathcal{Q}}^{(i)}$ in (C.23) under the flow of (7.11) is a smoothing remainder by applying Lemma B.10 in [33]. Of course in order to obtain homogeneous terms of degree ≤ 3 in ε we must Taylor expand the flow, following Remark 7.2 this implies that the remainders are in $\mathfrak{L}_{\rho-3,p}$ (of course since ρ is arbitrary this is not a problem).

The conjugation of the pseudo differential operators $J \circ a(\varphi, x)$ and $\mathrm{Op}(\mathfrak{q}(\varphi, x, \xi))$ is based on the Egorov Theorem 3.4 in [33]. This is a constructive perturbation scheme so we can Taylor expand up to order three. In conclusion (C.18) holds for each term in the homogeneity expansion, possibly with a larger $\tilde{\sigma}$. On the other hand expanding the remainder $\mathcal{Q}^+$ gives an estimate as (C.19) but with $\rho \rightsquigarrow \rho - 3$.

Remark C.4. We point out that the remainder in Proposition C.2 is of the order of β, i.e. of the generator of the torus diffeomorphism. This will create problems in fulfilling the smallness conditions in the KAM reducibility scheme of Sect. 7.3, where a term is perturbative if it is small w.r.t. $\gamma^{3/2}$ (and $\gamma \ll \varepsilon^2$).

C.2 Classes of "smoothing" operators. In the first step of our reduction procedure (see Theorem 7.1) we need to work with operators which are pseudo differential up to a remainder in the class $\mathfrak{L}_{\rho,p}$ defined as follows. This class of smoothing (in space) operators has been introduced in [33].

Definition C.5. Fix $s_0 \geq (\nu+1)/2$ and $p, \mathcal{S} \in \mathbb{N}$ with $s_0 \leq p < \mathcal{S}$ with possibly $\mathcal{S} = +\infty$. Fix $\rho \in \mathbb{N}$, with $\rho \geq 3$ and consider any subset $\mathcal{O}$ of $\mathbb{R}^\nu$. We denote by $\mathfrak{L}_{\rho,p} = \mathfrak{L}_{\rho,p}(\mathcal{O})$ the set of the linear operators $A = A(\omega)\colon H^s(\mathbb{T}^{\nu+1}) \to H^s(\mathbb{T}^{\nu+1})$, $\omega \in \mathcal{O}$ with the following properties:

- The operator A is Lipschitz in ω,
- The operators $\partial_\varphi^{\vec{\mathfrak{b}}} A, [\partial_\varphi^{\vec{\mathfrak{b}}} A, \partial_x]$, for all $\vec{\mathfrak{b}} = (\mathfrak{b}_1, \dots, \mathfrak{b}_\nu) \in \mathbb{N}^\nu$ with $0 \leq |\vec{\mathfrak{b}}| \leq \rho - 2$ have the following properties, for any $s_0 \leq s \leq \mathcal{S}$, with possibly $\mathcal{S} = +\infty$:

(i) For any $m_1, m_2 \in \mathbb{R}, m_1, m_2 \geq 0$ and $m_1 + m_2 = \rho - |\vec{\mathfrak{b}}|$ one has that the operator $\langle D_x\rangle^{m_1} \partial_\varphi^{\vec{\mathfrak{b}}} A \langle D_x\rangle^{m_2}$ is Lip-0-tame according to Definition 2.3 and we set

$$\mathfrak{M}^\gamma_{\partial_\varphi^{\vec{\mathfrak{b}}} A}(-\rho + |\vec{\mathfrak{b}}|, s) := \sup_{\substack{m_1+m_2=\rho-|\vec{\mathfrak{b}}| \\ m_1,m_2\geq 0}} \mathfrak{M}^\gamma_{\langle D_x\rangle^{m_1} \partial_\varphi^{\vec{\mathfrak{b}}} A \langle D_x\rangle^{m_2}}(0, s)\,; \tag{C.24}$$

(ii) For any $m_1, m_2 \in \mathbb{R}$, $m_1, m_2 \geq 0$ and $m_1 + m_2 = \rho - |\vec{\mathfrak{b}}| - 1$ one has that $\langle D_x\rangle^{m_1} [\partial_\varphi^{\vec{\mathfrak{b}}} A, \partial_x] \langle D_x\rangle^{m_2}$ is Lip-0-tame according to Definition 2.3 and we set

$$\mathfrak{M}^\gamma_{[\partial_\varphi^{\vec{\mathfrak{b}}} A, \partial_x]}(-\rho + |\vec{\mathfrak{b}}| + 1, s) := \sup_{\substack{m_1+m_2=\rho-|\vec{\mathfrak{b}}|-1 \\ m_1,m_2\geq 0}} \mathfrak{M}^\gamma_{\langle D_x\rangle^{m_1} [\partial_\varphi^{\vec{\mathfrak{b}}} A, \partial_x] \langle D_x\rangle^{m_2}}(0, s). \tag{C.25}$$

We define for $0 \leq \mathfrak{b} \leq \rho - 2$

$$\mathbb{M}^\gamma_A(s, \mathfrak{b}) := \max_{0\leq|\vec{\mathfrak{b}}|\leq \mathfrak{b}} \max \left(\mathfrak{M}^\gamma_{\partial_\varphi^{\vec{\mathfrak{b}}} A}(-\rho + |\vec{\mathfrak{b}}|, s), \mathfrak{M}^\gamma_{\partial_\varphi^{\vec{\mathfrak{b}}} [A, \partial_x]}(-\rho + |\vec{\mathfrak{b}}| + 1, s) \right). \tag{C.26}$$

Assume now that the set $\mathcal{O}$ and the operator A depend on $i = i(\omega)$, and are well defined for $\omega \in \mathcal{O}_0 \subseteq \Omega_\varepsilon$ for all i satisfying (6.12). We consider $i_1 = i_1(\omega)$, $i_2 = i_2(\omega)$ and for $\omega \in \mathcal{O}(i_1) \cap \mathcal{O}(i_2)$ we define

$$\Delta_{12} A := A(i_1) - A(i_2). \tag{C.27}$$

We require the following:

- The operators $\partial_\varphi^{\vec{\mathfrak{b}}} \Delta_{12} A$, $[\partial_\varphi^{\vec{\mathfrak{b}}} \Delta_{12} A, \partial_x]$, for $0 \leq |\vec{\mathfrak{b}}| \leq \rho - 3$, have the following properties, for any $s_0 \leq s \leq \mathcal{S}$, with possibly $\mathcal{S} = +\infty$:

(iii) For any $m_1, m_2 \in \mathbb{R}$, $m_1, m_2 \geq 0$ and $m_1 + m_2 = \rho - |\vec{\mathfrak{b}}| - 1$ one has that $\langle D_x\rangle^{m_1} \partial_\varphi^{\vec{\mathfrak{b}}} \Delta_{12} A \langle D_x\rangle^{m_2}$ is bounded on H^p into itself. More precisely there is a positive constant $\mathfrak{N}_{\partial_\varphi^{\vec{\mathfrak{b}}} \Delta_{12} A}(-\rho + |\vec{\mathfrak{b}}| + 1, p)$ such that, for any $h \in H^p$, we have

$$\sup_{\substack{m_1+m_2=\rho-|\vec{\mathfrak{b}}|-1 \\ m_1,m_2\geq 0}} \|\langle D_x\rangle^{m_1} \partial_\varphi^{\vec{\mathfrak{b}}} \Delta_{12} A \langle D_x\rangle^{m_2} h\|_p \leq \mathfrak{N}_{\partial_\varphi^{\vec{\mathfrak{b}}} \Delta_{12} A}(-\rho + |\vec{\mathfrak{b}}| + 1, p) \|h\|_p\,; \tag{C.28}$$

(iv) For any $m_1, m_2 \in \mathbb{R}$, $m_1, m_2 \geq 0$ and $m_1 + m_2 = \rho - |\vec{\mathtt{b}}| - 2$ one has that $\langle D_x\rangle^{m_1}[\partial_{\vec{\varphi}}^{\vec{\mathtt{b}}}\Delta_{12}A, \partial_x]\langle D_x\rangle^{m_2}$ is bounded on H^p into itself. More precisely there is a positive constant $\mathfrak{N}_{[\partial_{\vec{\varphi}}^{\vec{\mathtt{b}}}\Delta_{12}A,\partial_x]}(-\rho+|\vec{\mathtt{b}}|+2, p)$ such that for any $h \in H^p$ one has

$$\sup_{\substack{m_1+m_2=\rho-|\vec{\mathtt{b}}|-2\\ m_1,m_2\geq 0}} \|\langle D_x\rangle^{m_1}[\partial_{\vec{\varphi}}^{\vec{\mathtt{b}}}\Delta_{12}A, \partial_x]\langle D_x\rangle^{m_2}h\|_p \leq \mathfrak{N}_{[\partial_{\vec{\varphi}}^{\vec{\mathtt{b}}}\Delta_{12}A,\partial_x]}(-\rho+|\vec{\mathtt{b}}|+2, p)\|h\|_p. \tag{C.29}$$

We define for $0 \leq \mathtt{b} \leq \rho - 3$

$$\mathbb{M}_{\Delta_{12}A}(p, \mathtt{b}) := \max_{0\leq|\vec{\mathtt{b}}|\leq\mathtt{b}} \max\left(\mathfrak{N}_{\partial_{\vec{\varphi}}^{\vec{\mathtt{b}}}\Delta_{12}A}(-\rho+|\vec{\mathtt{b}}|+1, p), \mathfrak{N}_{\partial_{\vec{\varphi}}^{\vec{\mathtt{b}}}[\Delta_{12}A,\partial_x]}(-\rho+|\vec{\mathtt{b}}|+2, p)\right). \tag{C.30}$$

By construction one has that $\mathbb{M}_A^\gamma(s, \mathtt{b}_1) \leq \mathbb{M}_A^\gamma(s, \mathtt{b}_2)$ if $\mathtt{b}_1 \leq \mathtt{b}_2 \leq \rho - 2$ and $\mathbb{M}_{\Delta_{12}A}(p, \mathtt{b}_1) \leq \mathbb{M}_{\Delta_{12}A}(p, \mathtt{b}_2)$ if $\mathtt{b}_1 \leq \mathtt{b}_2 \leq \rho - 3$.

We shall also deal with "tame" operators in the following class.

Definition C.6. Fix $\mathcal{S} \in \mathbb{N}$ with possibly $\mathcal{S} = +\infty$. Fix $\mathtt{b} = s_0 + 6\tau + 6$ and consider $\mathcal{O} \subseteq \mathbb{R}^\nu$. We denote by $\mathfrak{C}_{-1} := \mathfrak{C}_{-1}(\mathcal{O})$ the set of the linear operators $A = A(\omega)\colon H^s(\mathbb{T}^{\nu+1}) \to H^s(\mathbb{T}^{\nu+1})$, $\omega \in \mathcal{O}$ which satisfy the following for any $s_0 \leq s \leq \mathcal{S}$:

- $\langle D_x\rangle^{1/2}A\langle D_x\rangle^{1/2}$, $\langle D_x\rangle^{1/2}\partial_{\varphi_m}^{\mathtt{b}}A\langle D_x\rangle^{1/2}$, $\langle D_x\rangle^{1/2}[\partial_{\varphi_m}^{\mathtt{b}}A, \partial_x]\langle D_x\rangle^{1/2}$, for $m = 1, \ldots, \nu$, $0 \leq \mathtt{b}_1 \leq \mathtt{b}$ are Lip-0-tame operators (see Definition 2.3) and we define

$$\begin{aligned}
\mathfrak{M}^\gamma_{\partial_{\varphi m}^{\mathtt{b}_1}A}(-1, s) &:= \mathfrak{M}^\gamma_{\langle D_x\rangle^{1/2}\partial_{\varphi m}^{\mathtt{b}_1}A\langle D_x\rangle^{1/2}}(0, s),\\
\mathfrak{M}^\gamma_{\partial_{\varphi m}^{\mathtt{b}_1}[A,\partial_x]}(-1, s) &:= \mathfrak{M}^\gamma_{\langle D_x\rangle^{1/2}\partial_{\varphi m}^{\mathtt{b}_1}[A,\partial_x]\langle D_x\rangle^{1/2}}(0, s), \qquad\qquad (C.31)\\
\mathbb{B}_A^\gamma(s) &:= \max_{\substack{0\leq\mathtt{b}_1\leq\mathtt{b}\\ m=1,\ldots,\nu}} \max\left(\mathfrak{M}^\gamma_{\partial_{\varphi m}^{\mathtt{b}_1}A}(-1, s), \mathfrak{M}^\gamma_{\partial_{\varphi m}^{\mathtt{b}_1}[A,\partial_x]}(-1, s)\right). \qquad (C.32)
\end{aligned}$$

Assume now that the set $\mathcal{O}$ and the operator A depend on $i = i(\omega)$, and are well defined for $\omega \in \mathcal{O} \subseteq \mathbb{R}^\nu$ for all i satisfying (6.12). We consider $i_1 = i_1(\omega), i_2 = i_2(\omega)$ and for $\omega \in \mathcal{O}(i_1) \cap \mathcal{O}(i_2)$ let $\Delta_{12}A$ as in (C.27). We require the following:

- $\langle D_x\rangle^{1/2}\partial_{\varphi_m}^{\mathtt{b}_1}\Delta_{12}A\langle D_x\rangle^{1/2}$, $\langle D_x\rangle^{1/2}[\partial_{\varphi_m}^{\mathtt{b}_1}\Delta_{12}A, \partial_x]\langle D_x\rangle^{1/2}$ for $m = 1, \ldots, \nu$, $0 \leq \mathtt{b}_1 \leq \mathtt{b}$ are bounded operators on H^{s_0} into itself. More precisely there are positive constants $\mathfrak{N}_{\partial_{\varphi m}^{\mathtt{b}_1}\Delta_{12}A}(-1, s_0)$, and $\mathfrak{N}_{[\partial_{\varphi m}^{\mathtt{b}_1}\Delta_{12}A,\partial_x]}(-1, s_0)$ such that, for any $h \in H^{s_0}$, we have

$$\begin{aligned}
\|\langle D_x\rangle^{1/2}\partial_{\varphi_m}^{\mathtt{b}_1}\Delta_{12}A\langle D_x\rangle^{1/2}h\|_{s_0} &\leq \mathfrak{N}_{\partial_{\varphi m}^{\mathtt{b}_1}\Delta_{12}A}(-1, s_0)\|h\|_{s_0},\\
\|\langle D_x\rangle^{1/2}[\partial_{\varphi_m}^{\mathtt{b}_1}\Delta_{12}A, \partial_x]\langle D_x\rangle^{1/2}h\|_{s_0} &\leq \mathfrak{N}_{[\partial_{\varphi m}^{\mathtt{b}_1}\Delta_{12}A,\partial_x]}(-1, s_0)\|h\|_{s_0}. \qquad (C.33)
\end{aligned}$$

We define

$$\mathbb{B}_{\Delta_{12}A}(s_0) := \max_{\substack{0\leq\mathtt{b}_1\leq\mathtt{b}\\ m=1,\ldots,\nu}} \max\left(\mathfrak{N}_{\partial_{\varphi m}^{\mathtt{b}_1}\Delta_{12}A}(-1, s_0), \mathfrak{N}_{\partial_{\varphi m}^{\mathtt{b}_1}[\Delta_{12}A,\partial_x]}(-1, s_0)\right). \tag{C.34}$$

The next Lemma shows that the finite rank operators of the form (6.16) are in $\mathfrak{L}_{\rho,p}$.

Lemma C.7. *Fix $\rho \geq 3$. Let $\mathcal{R}$ be an operator of the form (6.16), where the functions $g_j(\tau)$, $\chi_j(\tau)$ belong to H^s for $\tau \in [0,1]$ and depend in a Lipschitz way on the parameter $\omega \in \mathcal{O} \subset \mathbb{R}^\nu$. Then there exists $\sigma_1 = \sigma_1(\rho) > 0$ such that $\mathcal{R}$ belongs to $\mathfrak{L}_{\rho,p}$ and*

$$\mathbb{M}^\gamma_{\mathcal{R}}(s,\mathtt{b}) \lesssim_s \sum_{|j|\leq C} \sup_{\tau\in[0,1]} (\|\chi_j(\tau)\|^{\gamma,\mathcal{O}}_{s+\sigma_1}\|g_j\|^{\gamma,\mathcal{O}}_{s_0+\sigma_1} + \|\chi_j(\tau)\|^{\gamma,\mathcal{O}}_{s_0+\sigma_1}\|g_j(\tau)\|^{\gamma,\mathcal{O}}_{s+\sigma_1}), \tag{C.35}$$

$$\mathbb{M}_{\Delta_{12}\mathcal{R}}(p,\mathtt{b}) \leq_p \sum_{|j|\leq C} \sup_{\tau\in[0,1]} \Big(\|\Delta_{12}\chi_j(\tau)\|_{p+\sigma_1}\|g_j\|_{p+\sigma_1} + \|\chi_j(\tau)\|_{p+\sigma_1}\|\Delta_{12}g_j\|_{p+\sigma_1}\Big). \tag{C.36}$$

Proof. The Lemma follows by reasoning as in the proof of Lemma B.2 in [33] and using the explicit formula (6.16). □

We conclude this section by showing the connection between the class $\mathfrak{L}_{\rho,p}$ and the class $\mathfrak{C}_{-1}$ in Definition C.6.

Lemma C.8. *Consider $\mathtt{b} \in \mathbb{N}$ and $\rho \in \mathbb{N}$ with $\rho \geq \mathtt{b}+3$. The following holds.*

(i) *If $A \in \mathfrak{L}_{\rho,p}$ (see Definition C.5) then $A \in \mathfrak{C}_{-1}$ (see Definition C.6) with*

$$\mathbb{B}^\gamma_A(s,\mathtt{b}) \leq_{\rho,s} \mathbb{M}^\gamma_A(s,\rho-2), \qquad \mathbb{B}_{\Delta_{12}A}(p,\mathtt{b}) \leq_{\rho,p} \mathbb{M}_{\Delta_{12}A}(p,\rho-3). \tag{C.37}$$

(ii) *Consider a symbol $a = a(\omega, i(\omega))$ in S^m with $m \leq -1$ depending on $\omega \in \mathcal{O}_0 \subset \mathbb{R}^\nu$ in a Lipschitz way and on i in a Lipschitz way and let $A := \mathrm{Op}(a(x,\xi))$. Then one has that $A \in \mathfrak{C}_{-1}$ with*

$$\mathbb{B}^\gamma_A(s,\mathtt{b}) \leq_s |a|^{\gamma,\mathcal{O}_0}_{m,s+\mathtt{b}+2,0}, \quad \mathbb{B}_{\Delta_{12}A}(p,\mathtt{b}) \leq_s |\Delta_{12}a|_{m,p+\mathtt{b}+3,0}. \tag{C.38}$$

(iii) *Let $A, B \in \mathfrak{C}_{-1}$. Then $A \circ B \in \mathfrak{C}_{-1}$ with*

$$\mathbb{B}^\gamma_{A\circ B}(s,\mathtt{b}) \leq_s \mathbb{B}^\gamma_A(s,\mathtt{b})\mathbb{B}^\gamma_A(s_0,\mathtt{b}) + \mathbb{B}^\gamma_A(s_0,\mathtt{b})\mathbb{B}^\gamma_A(s,\mathtt{b}) \tag{C.39}$$

$$\mathbb{B}_{\Delta_{12}(A\circ B)}(p,\mathtt{b}) \leq_{p,\rho} \mathbb{B}_{\Delta_{12}A}(p,\mathtt{b})\mathbb{B}_B(p,\mathtt{b}), + \mathbb{B}_{\Delta_{12}B}(p,\mathtt{b})\mathbb{B}_A(p,\mathtt{b}). \tag{C.40}$$

Proof. Let us check item (i). The fact that $\langle D_x\rangle^{1/2} A \langle D_x\rangle^{1/2}$ is Lip-0-tame follows by (C.24) since $\rho \geq 1$. Indeed $\langle D_x\rangle^{-\rho+1}$ is bounded in x and for any $h \in H^s$

$$\begin{aligned}\|\langle D_x\rangle^{\frac12} A \langle D_x\rangle^{\frac12} h\|^{\gamma,\mathcal{O}_0}_s &\leq \|\langle D_x\rangle^{-\rho+1}\big(\langle D_x\rangle^{\rho-\frac12} A \langle D_x\rangle^{\frac12}\big)h\|^{\gamma,\mathcal{O}_0}_s \\ &\leq_s \mathfrak{M}^\gamma_A(-\rho,s)\|h\|^{\gamma,\mathcal{O}_0}_{s_0} + \mathfrak{M}^\gamma_A(-\rho,s_0)\|h\|^{\gamma,\mathcal{O}_0}_s.\end{aligned}$$

By studying the tameness constant of the operators $\partial^{\vec{\mathtt{b}}}_\varphi A$, $[A,\partial_x]$, $[\partial^{\vec{\mathtt{b}}}_\varphi A,\partial_x]$ $\Delta_{12}A$, $\partial^{\vec{\mathtt{b}}}_\varphi \Delta_{12}A$, $[\Delta_{12}A,\partial_x]$, $[\partial^{\vec{\mathtt{b}}}_\varphi \Delta_{12}A,\partial_x]$ for $\vec{\mathtt{b}} \in \mathbb{N}^\nu$, $|\vec{\mathtt{b}}| = \mathtt{b}$, following the same reasoning as above, one gets the (C.37).
In order to prove item (ii) one can follow almost word by word the proof of Lemma A.4 in [33].
Let us check (C.39). Let $\mathtt{b} \in \mathbb{N}$ and consider $0 \leq \mathtt{b}_1 \leq \mathtt{b}$, $m = 1,\dots,\nu$. One has

$$\langle D_x\rangle^{\frac12}\partial^{\mathtt{b}_1}_{\varphi_m}(A\circ B)\langle D_x\rangle^{\frac12} = \sum_{\mathtt{c}_1+\mathtt{c}_2=\mathtt{b}_1} C(\mathtt{c}_1,\mathtt{c}_2)\langle D_x\rangle^{\frac12}(\partial^{\mathtt{c}_1}_{\varphi_m}A)(\partial^{\mathtt{c}_2}_{\varphi_m}B)\langle D_x\rangle^{\frac12}. \tag{C.41}$$

We show that each summand in (C.41) is a Lip-0-tame operator. We have for $h \in H^s$

$$\|\langle D_x\rangle^{\frac12}(\partial^{\mathtt{c}_1}_{\varphi_m}A)(\partial^{\mathtt{c}_2}_{\varphi_m}B)\langle D_x\rangle^{\frac12} h\|^{\gamma,\mathcal{O}_0}_s$$

$$\leq \|\langle D_x\rangle^{\frac12}(\partial_{\varphi_m}^{\mathtt{c}_1}A)\langle D_x\rangle^{\frac12}\langle D_x\rangle^{-1}\langle D_x\rangle^{\frac12}(\partial_{\varphi_m}^{\mathtt{c}_2}B)\langle D_x\rangle^{\frac12}h\|_s^{\gamma,\mathcal{O}_0}$$
$$\leq_s (\mathbb{B}_A^\gamma(s,\mathtt{b})\mathbb{B}_B^\gamma(s_0,\mathtt{b})+\mathbb{B}_A^\gamma(s_0,\mathtt{b})\mathbb{B}_B^\gamma(s,\mathtt{b}))\|h\|_{s_0}^{\gamma,\mathcal{O}_0}+\mathbb{B}_A^\gamma(s_0,\mathtt{b})\mathbb{B}_B^\gamma(s_0,\mathtt{b})\|h\|_s^{\gamma,\mathcal{O}_0}. \tag{C.42}$$

In (C.42) we used the fact that $\langle D_x\rangle^{\frac12}(\partial_{\varphi_m}^{\mathtt{c}_1}A)\langle D_x\rangle^{\frac12}$ and $\langle D_x\rangle^{\frac12}(\partial_{\varphi_m}^{\mathtt{c}_2}B)\langle D_x\rangle^{\frac12}$ are 0-tame by hypothesis (see Definition (C.6)). This proves (C.39) for the operators $A\circ B$ and $\partial_{\varphi_m}^{\mathtt{b}_1}(A\circ B)$ for any $0\leq\mathtt{b}_1\leq\mathtt{b}$, $m=1,\dots,\nu$. One concludes the proof of (C.39) and (C.40) followings the same ideas used above. For further details we refer to the proof of Lemma $B.1$ in [33]. □

Lemma C.9. *Let $X,Y\in\mathfrak{C}_{-1}$ then $\mathrm{ad}_X^k[Y]\in\mathfrak{C}_{-1}$ for any $k\geq1$ (recall (7.68)). Moreover for any $k\geq1$ we have*

$$\mathbb{B}^\gamma_{\mathrm{ad}_X^k[Y]}(s)\lesssim_s \mathbb{B}_X^\gamma(s)\big(\mathbb{B}_X^\gamma(s_0)\big)^{k-1}\mathbb{B}_Y^\gamma(s_0)+\big(\mathbb{B}_X^\gamma(s_0)\big)^k\mathbb{B}_Y^\gamma(s). \tag{C.43}$$

Moreover if X,Y depend on some parameter i we have

$$\mathbb{B}_{\Delta_{12}\mathrm{ad}_X^k[Y]}(s_0)\lesssim\sum_{j_1+j_2=k-1}\mathbb{B}^{j_1}_{X(i_2)}(s_0)\,\mathbb{B}_{\Delta_{12}X}(s_0)\,\mathbb{B}^{j_2}_{X(i_1)}(s_0)\,\mathbb{B}_{Y(i_1)}(s_0)+\mathbb{B}^k_{X(i_2)}\mathbb{B}_{\Delta_{12}Y}(s_0). \tag{C.44}$$

Proof. It follows by using the formula

$$\Delta_{12}\mathrm{ad}_X^k[Y]=\sum_{j_1+j_2=k-1}\mathrm{ad}^{j_1}_{X(i_2)}\,\mathrm{ad}_{\Delta_{12}X}\,\mathrm{ad}^{j_2}_{X(i_1)}[Y(i_1)]+\mathrm{ad}^k_{X(i_2)}[\Delta_{12}Y]$$

and applying iteratively the estimates of Lemma C.8. □

Lemma C.10. *Let $A\in\mathfrak{C}_{-1}$ then A is a Lip--1-modulo tame operator according to Definition 2.7. Moreover*

$$\mathfrak{M}_A^{\sharp,\gamma^{3/2}}(s)\leq\max_{m=1,\dots,\nu}\mathfrak{M}^{\gamma^{3/2}}_{\partial_{\varphi_m}^{s_0}[A,\partial_x]}(-1,s),\quad \mathfrak{M}_A^{\sharp,\gamma^{3/2}}(s,\mathtt{b}_0)\leq\max_{m=1,\dots,\nu}\mathfrak{M}^{\gamma^{3/2}}_{\partial_{\varphi_m}^{s_0+\mathtt{b}_0}[A,\partial_x]}(-1,s), \tag{C.45}$$

$$\|\langle D_x\rangle^{1/2}\underline{\Delta_{12}A}\langle D_x\rangle^{1/2}\|_{\mathcal{L}(H^{s_0})},\ \|\langle D_x\rangle^{1/2}\underline{\Delta_{12}\langle\partial_\varphi\rangle^{\mathtt{b}_0}A}\langle D_x\rangle^{1/2}\|_{\mathcal{L}(H^{s_0})}\leq\mathbb{B}_{\Delta_{12}A}(s_0,\mathtt{b}_0). \tag{C.46}$$

Proof. It follows arguing as in the proof of Lemma $A.4$ in Appendix A of [33] using the definition of $\mathfrak{C}_{-1}$ in Definition C.6. □

C.3 Linear Birkhoff normal form. The next lemma regards almost diagonal vector fields which belong to $\mathfrak{C}_{-1}$.

Lemma C.11. *Let $X_A:=JA\in\mathfrak{C}_{-1}$ with A almost diagonal in the sense of Definition 6.4 . Then $|A^{j',l'}_{j,l}|\leq C\langle j,j'\rangle^{-2}$ for some constant $C>0$.*

Proof. The operator $B:=\langle D_x\rangle^{1/2}JA\langle D_x\rangle^{1/2}$ belongs to $\mathcal{L}(H^s)$ for any s. Then $|B^{j',l'}_{j,l}|\leq C$ for all $j,j'\in\mathbb{Z},l,l'\in\mathbb{Z}^\nu$. The thesis follows by the fact that

$$|B^{j',l'}_{j,l}|=|A^{j',l'}_{j,l}|\langle j\rangle^{1/2}\langle j'\rangle^{1/2}|\lambda(j)|\gtrsim|A^{j',l'}_{j,l}|\langle j\rangle^{1/2}\langle j'\rangle^{1/2}|j|$$

and $|j|\gtrsim\langle j\rangle^{1/2}\langle j'\rangle^{1/2}$, since A is almost diagonal. □

We now study the flows generated by operators belonging to the class $\mathfrak{C}_{-1}$ given in Definition C.6.
In Sect. 7.2 we look for symplectic changes of variable $\Upsilon_i : H^s_{S^\perp}(\mathbb{T}^{\nu+1}) \to H^s_{S^\perp}(\mathbb{T}^{\nu+1})$, $i = 1, 2, 3$, that are the time-1 flow of quadratic Hamiltonians

$$H_{\mathtt{A}_i}(u) := \varepsilon \sum_{j, j' \in S^c} (\mathtt{A}_i)^{j'}_{j}(\varphi)\, u_{j'}\, \overline{u}_j, \tag{C.47}$$

where $\mathtt{A}_i(\varphi)$ is a self-adjoint operator $\forall \varphi \in \mathbb{T}^\nu$ and thus

$$\Upsilon_i := \exp(\varepsilon\, J\mathtt{A}_i) = \mathrm{I}_{H^\perp_S} + \varepsilon\, J\mathtt{A}_i + \varepsilon^2 \frac{(J\mathtt{A}_i)^2}{2} + \varepsilon^3 R_i, \quad R_i := \sum_{k \geq 3} \frac{\varepsilon^{k-3}}{k!} (J\mathtt{A}_i)^k. \tag{C.48}$$

Given linear operators $\mathtt{B}_i$, $i = 1, 2, 3$ define the matrices $\mathtt{A}_i$, $i = 1, 2, 3$ as

$$\begin{aligned} (\mathtt{A}_i)^{j',l'}_{j,l} &= (\mathtt{A}_i)^{j'}_{j}(l - l') = -\frac{(\mathtt{B}_i)^{j'}_{j}(l - l')}{\delta_{ljj'}}, \\ &\delta_{ljj'} \neq 0, \quad |j - j'| \leq 2\, i\, \overline{\jmath}_1, \quad |l - l'| \leq i, \quad i = 1, 2, \end{aligned} \tag{C.49}$$

$$\begin{aligned} (\mathtt{A}_3)^{j',l'}_{j,l} &= (\mathtt{A}_3)^{j'}_{j}(l - l') = -\frac{(\mathtt{B}_3)^{j'}_{j}(l - l')}{\delta^*_{ljj'}}, \\ &\delta^*_{ljj'} \neq 0, \quad |j - j'| \leq 6\, \overline{\jmath}_1, \quad |l - l'| \leq 3, \end{aligned} \tag{C.50}$$

where, recall (1.8), (4.6),(5.6),

$$\delta_{\ell jj'} := \overline{\omega} \cdot \ell + \lambda(j) - \lambda(j'), \quad \delta^*_{\ell jj'} := \delta_{\ell jj'} + \varepsilon^2 (\mathbb{A}\xi \cdot \ell + \lambda(j')\mathfrak{l}_{j'} - \lambda(j)\mathfrak{l}_j). \tag{C.51}$$

Lemma C.12. *Let $j, j' \in S^c$, $j \neq j'$. If $\sum_{i=1}^{\nu} \overline{\jmath}_i\, \ell_i + j - j' = 0$, $0 < |\ell| \leq 2$, $\delta_{\ell jj'} \neq 0$, where $\delta_{\ell jj'}$ are given in (C.51), then there exists a constant C depending on the set S such that $|\delta_{\ell jj'}| \geq C$.*

Proof. If $|\ell| = 1$ we have by the preservation of momentum

$$\delta_{\ell, j, j'} = \lambda(j - j') - \lambda(j) + \lambda(j') = \frac{3jj'(j - j')[3 + jj' + (j - j')^2]}{(1 + j^2)(1 + j'^2)(1 + (j - j')^2)}.$$

It is easy to verify that (recall that $|j - j'| \leq 2\overline{\jmath}_1$)

$$\begin{aligned} &|j\ j'||j - j'||3 + jj' + (j - j')^2| \geq |j\ j'|^2, \quad (1 + j^2)(1 + j'^2)(1 + (j - j')^2) \\ &\leq K |j\ j'|^2 \overline{\jmath}_1^2. \end{aligned}$$

Which implies the thesis for $|\ell| = 1$. Now suppose $|\ell| = 2$ and consider $j_1, j_2 \in S$. We can write $\delta_{\ell jj'} = \lambda(j_1) + \lambda(j_2) + \lambda(j) - \lambda(j')$ and by the conservation of momentum

$$\lambda(j_1) + \lambda(j_2) + \lambda(j) - \lambda(j') = (j_1 + j_2)(j_1 + j)(j_2 + j)\, P(j_1, j_2, j) \tag{C.52}$$

and P is the rational function

$$P(x, y, z) := \frac{3 + x^2 + y^2 + z^2 + xy + xz + yz + xyz(x + y + z)}{(1 + x^2)(1 + y^2)(1 + z^2)(1 + (x + y + z)^2)}. \tag{C.53}$$

If $|j| > N$, where $N = N(S)$ is a large constant to be fixed and which depends on the set S, then

$$|(j_1 + j_2)(j_1 + j)(j_2 + j)| \geq C_1 \, j^2$$

for some constant $C_1 := C_1(N) > 0$ (possibly small), provided that N is large enough. Moreover

$$|(1 + j_1 j_2) j^2 + (j_1^2 j_2 + j_1 j_2^2) j + 3 + j_1^2 + j_2^2| \geq C_2 \, j^2,$$
$$|(1 + j_1^2)(1 + j_2^2)(1 + j^2)(1 + (j_1 + j_2 + j)^2)| \leq C_3 \, j^4$$

for some small constant $C_2 := C_2(N) > 0$, provided that N is large enough, and some big constant $C_3 := C_3(S) > 0$. Thus $|\lambda(j_1) + \lambda(j_2) + \lambda(j) - \lambda(j')| \geq C_1 C_2 C_3^{-1} > 0$. Now consider $|j| \leq N$. Then we have

$$|(j_1 + j_2)(j_1 + j)(j_2 + j)||(1 + j_1 j_2) j^2 + (j_1^2 j_2 + j_1 j_2^2) j + 3 + j_1^2 + j_2^2| \geq M_1$$
$$|(1 + j_1^2)(1 + j_2^2)(1 + j^2)(1 + (j_1 + j_2 + j)^2| \leq M_2$$

for some constant $M_1, M_2 > 0$ depending on S. Set $C_4 := M_1/M_2$.
Therefore $|\lambda(j_1) + \lambda(j_2) + \lambda(j) - \lambda(j')| \geq C_4 > 0$. At the end we choose $C \geq \max\{C_4, C_1 C_2 / C_3\}$. □

Lemma C.13. *Assume that* $\mathtt{B}_i$, $i = 1, 2, 3$ *are such that* $J\mathtt{B}_i \in \mathfrak{C}_{-1}$ *and that they are* almost diagonal *(see Definition 6.4). Then, for any* $\omega \in \mathcal{G}_0^{(1)}$ *(see* (5.5)*), the following holds true:*
(i) The linear vector fields $X_{\mathtt{A}_i} := J\mathtt{A}_i$, *with* $\mathtt{A}_i$ *defined as in* (C.49),(C.50), *belongs to the class* $\mathfrak{C}_{-1}$, *in particular it satisfies the following:*

$$\mathbb{B}^{\gamma}_{\varepsilon^i X_{\mathtt{A}_i}}(s) \leq C(s)\varepsilon^i \quad i = 1, 2, \qquad \mathbb{B}^{\gamma}_{\varepsilon^i X_{\mathtt{A}_3}}(s) \leq C(s)\varepsilon^3 \gamma^{-1}, \qquad \forall s \geq s_0. \tag{C.54}$$

Note that $X_{\mathtt{A}_i}$ *does not depend on* $i(\omega)$.
(ii) The transformation $\Upsilon_i \colon H^s(\mathbb{T}^{\nu+1}) \to H^s(\mathbb{T}^{\nu+1})$, $i = 1, 2, 3$ *defined in* (C.48) *is invertible and satisfies, for any* $u = u(\omega) \in H^s$ *Lipschitz in* $\omega \in \mathcal{O}_\infty^{2\gamma}$,

$$\|(\Upsilon_i^{\pm 1} - \mathrm{I})u\|_s^{\gamma, \mathcal{O}_\infty^{2\gamma}} \lesssim \varepsilon^i C(s_0)\|u\|_s^{\gamma, \mathcal{O}_\infty^{2\gamma}} + \varepsilon^i C(s)\|u\|_{s_0}^{\gamma, \mathcal{O}_\infty^{2\gamma}}, \quad i = 1, 2, \tag{C.55}$$
$$\|(\Upsilon_3^{\pm 1} - \mathrm{I})u\|_s^{\gamma, \mathcal{O}_\infty^{2\gamma}} \lesssim \varepsilon^3 \gamma^{-1} C(s_0)\|u\|_s^{\gamma, \mathcal{O}_\infty^{2\gamma}} + \varepsilon^3 \gamma^{-1} C(s)\|u\|_{s_0}^{\gamma, \mathcal{O}_\infty^{2\gamma}}. \tag{C.56}$$

Proof. First of all notice that, by Lemmata C.11, C.12 and the fact that $\omega \in \mathcal{G}_0^{(1)}$ (see (5.5)) we will have

$$|(\mathtt{A}_i)^{j', l'}_{j, l}| \leq \frac{C}{\langle j, j' \rangle^2} \quad i = 1, 2, \qquad |(\mathtt{A}_3)^{j', l'}_{j, l}| \leq \frac{C\gamma}{\langle j, j' \rangle^2}, \qquad \forall j, j' \in S^c, \ l, l' \in \mathbb{Z}^\nu \tag{C.57}$$

for some constant $C > 0$ depending on the set S.
Proof of item (i). First we note that that $B := \langle D_x \rangle^{1/2} J\mathtt{A}_i \langle D_x \rangle^{1/2}$ maps H^s to itself for all $s \geq 0$. Indeed it is sufficient to exploit the fact that the matrix entries $B^{j'}_j(l - l')$ are uniformly bounded by a constant and B is almost diagonal. The matrix elements of $\partial^{\mathtt{b}}_{\varphi_m} X_{\mathtt{A}_1}$, $[X_{\mathtt{A}_1}, \partial_x]$, $[\partial^{\mathtt{b}}_{\varphi_m} X_{\mathtt{A}_1}, \partial_x]$ are respectively

$$\langle \ell_m - \ell'_m \rangle^{\mathtt{b}} \lambda(j)(\mathtt{A}_1)^{j'}_j (\ell - \ell'), \quad (j - j')\lambda(j)(\mathtt{A}_1)^{j'}_j (\ell - \ell'),$$

$$\langle \ell_m - \ell'_m \rangle^{\mathtt{b}} (j - j') \lambda(j) (\mathtt{A}_1)_j^{j'} (\ell - \ell').$$

Note that by the definition of $\mathtt{A}_1$ in (C.49)

$$\langle \ell_m - \ell'_m \rangle^{\mathtt{b}}, |j - j'| \leq C$$

for some constant C depending on the set S. Thus arguing as above one can easily prove that $\partial_{\varphi_m}^{\mathtt{b}} X_{\mathtt{A}_1}$, $[X_{\mathtt{A}_1}, \partial_x]$, $[\partial_{\varphi_m}^{\mathtt{b}} X_{\mathtt{A}_1}, \partial_x]$ are -1-Lip-tame operators. This concludes the proof of the (C.54)

Proof of item (ii). By (C.48) we have

$$(\Upsilon_i - \mathrm{I})u = \sum_{k \geq 1} \frac{\varepsilon^{ik} X_{\mathtt{A}_i}^k u}{k!}, \qquad (\Upsilon_1^{-1} - \mathrm{I})u = \sum_{k \geq 1} (-1)^k \frac{\varepsilon^{ik} X_{\mathtt{A}_i}^k u}{k!}. \tag{C.58}$$

By using iteratively the property (iii) of Lemma C.8 and item (i) we have that

$$\| X_{\mathtt{A}_i}^k u \|_s^{\gamma, \mathcal{O}_\infty^{2\gamma}} \leq \mathbb{B}_{X_{\mathtt{A}_i}}^{\gamma}(s) (\mathbb{B}_{X_{\mathtt{A}_i}}^{\gamma}(s_0))^{k-1} \| u \|_{s_0}^{\gamma, \mathcal{O}_\infty^{2\gamma}} + (\mathbb{B}_{X_{\mathtt{A}_i}}^{\gamma}(s_0))^k \| u \|_s^{\gamma, \mathcal{O}_\infty^{2\gamma}}.$$

By using this relation to estimate the Lip-Sobolev norm of (C.58) and by noting that $\varepsilon^n C(s_0)^n$ is a summable sequence, for ε small enough, we prove the thesis. □

References

1. Baldi, P.: Periodic solutions of fully nonlinear autonomous equations of Benjamin–Ono type. Ann. Inst. H. Poincaré (C) Anal. Non Linéaire **30**, 33–77 (2013)
2. Baldi, P., Berti, M., Haus, E., Montalto, R.: Time quasi-periodic gravity water waves in finite depth. Invent. Math. **214**(2), 739–911 (2018)
3. Baldi, P., Berti, M., Montalto, R.: KAM for quasi-linear and fully nonlinear forced perturbations of Airy equation. Math. Ann. **359**, 471–536 (2014)
4. Baldi, P., Berti, M., Montalto, R.: KAM for autonomous quasi-linear perturbations of KdV. Ann. I. H. Poincaré (C) Anal. Non Linéaire **33**, 1589–1638 (2016)
5. Bambusi, D., Maspero, A.: 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)
6. Berti, M., Biasco, L., Procesi, M.: KAM theory for the Hamiltonian derivative wave equation. Ann. Sci. de l'ENS **46**(2), 299–371 (2013)
7. Berti, M., Biasco, L., Procesi, M.: KAM theory for reversible derivative wave equations. Arch. Ration. Mech. Anal. **212**(3), 905–955 (2014)
8. Berti, M., Bolle, P.: Quasi-periodic solutions with Sobolev regularity of NLS on $\mathbb{T}^d$ with a multiplicative potential. Eur. J. Math. **15**, 229–286 (2013)
9. Berti, M., Bolle, P.: A Nash–Moser approach to KAM theory. In: Guyenne, P., Nicholls, D., Sulem, C. (eds.) Hamiltonian PDEs and Applications. Fields Institute Communications, vol. 75, pp. 255–284. Springer, Berlin (2014)
10. Berti, M., Delort, J.M.: Almost global solutions of capillarity-gravity water waves equations on the circle. UMI Lecture Notes (2017)
11. Berti, M., Corsi, L., Procesi, M.: An abstract Nash–Moser theorem and quasi-periodic solutions for NLW and NLS on compact Lie groups and homogeneous manifolds. Commun. Math. Phys. **334**(3), 1413–1454 (2015)
12. Berti, M., Feola, R., Pusateri, F.: Birkhoff normal form and long time existence for periodic gravity water waves (2018). Preprint arXiv:1810.11549
13. Berti, M., Feola, R., Pusateri, F.: Birkhoff normal form for gravity water waves. Water Waves (2020). https://doi.org/10.1007/s42286-020-00024-y
14. Berti, M., Kappeler, Th, Montalto, R.: KAM Tori for perturbations of the defocusing NLS equation. Astérisque **403**, viii+148 (2018)
15. Berti, M., Montalto, R.: Quasi-Periodic Standing Wave Solutions of Gravity-Capillary Water Waves. Memoirs of the American Mathematical Society, American Mathematical Society, Providence, RI (2016)

16. Bourgain, J.: Construction of quasi-periodic solutions for Hamiltonian perturbations of linear equations and applications to nonlinear PDE. Int. Math. Res. Not. **11**, 475–497 (1994)
17. Bourgain, J.: Quasi-periodic solutions of Hamiltonian perturbations of 2D linear Schrödinger equations. Ann. Math. (2) **148**(2), 363–439 (1998)
18. Camassa, R., Holm, D.: An integrable shallow water equation with peaked solution. Phys. Rev. Lett. **71**(11), 1661–1664 (1993)
19. Chierchia, L., You, J.: KAM Tori for 1D nonlinear wave equations with periodic boundary conditions. Commun. Math. Phys. **211**, 497–525 (2000)
20. Constantin, A.: Global existence and breaking waves for a shallow water equation: a geometric approach. Ann. Inst. Fourier (Grenoble) **50**, 321–362 (2000)
21. Constantin, A., Escher, J.: Wave breaking for nonlinear nonlocal shallow water equations. Acta Math. **181**(2), 229–243 (1998)
22. Constantin, A., Ivanov, R.: Dressing method for the Degasperis–Procesi equation. Stud. Appl. Math. **138**(2), 205–226 (2017)
23. Constantin, A., Lannes, D.: The hydrodynamical relevance of the Camassa–Holm and Degasperis–Procesi equations. Arch. Ration. Mech. Anal. **192**, 165–186 (2009)
24. Corsi, L., Feola, R., Procesi, M.: Finite dimensional invariant KAM Tori for tame vector fields. Trans. Am. Math. Soc. **372**, 1913–1983 (2019)
25. Corsi, L., Montalto, R.: Quasi-periodic solutions for the forced Kirchhoff equation on $\mathbb{T}^d$. Nonlinearity **31**(11), 5075–5109 (2018)
26. Craig, W., Wayne, C.E.: Newton's method and periodic solutions of nonlinear wave equations. Commun. Pure Appl. Math. **46**(11), 1409–1498 (1993)
27. Craig, W., Worfolk, P.: An integrable normal form for water waves in infinite depth. Physica D **84**(3–4), 513–531 (1995)
28. Degasperis, A., Holm, D.D., Hone, A.N.W.: A new integrable equation with peakon solutions. Theor. Math. Phys. **133**(2), 1463–1474 (2002)
29. Degasperis, A., Procesi, M.: Asymptotic integrability. In: Degasperis, A., Gaeta, G. (eds.) Symmetry and Perturbation Theory, pp. 23–37. World Scientific Publishing, Singapore (1999)
30. Eliasson, H., Kuksin, S.: KAM for the nonlinear Schrödinger equation. Ann. Math. (2) **172**(1), 371–435 (2010)
31. Feola, R., Giuliani, F., Montalto, R., Procesi, M.: Reducibility of first order linear operators on tori via Moser's theorem. J. Funct. Anal. **276**(3), 932–970 (2019)
32. Feola, R., Giuliani, F., Pasquali, S.: On the integrability of Degasperis–Procesi equation: control of the Sobolev norms and Birkhoff resonances. J. Differ. Equ. **266**(6), 3390–3437 (2018)
33. Feola, R., Giuliani, F., Procesi, M.: Reducibility for a class of weakly dispersive linear operators arising from the Degasperis Procesi equation. Dyn. Partial Differ. Equ. **16**(1), 25–94 (2019)
34. Feola, R., Grébert, B.: Reducibility of Schrödinger equation on the sphere. Int. Math. Res. Not. (2019). https://doi.org/10.1093/imrn/rnz344
35. Feola, R., Iandoli, F.: Local well-posedness for quasi-linear NLS with large Cauchy data on the circle. Annales de l'Institut Henri Poincaré C. Analyse non linéaire **36**(1), 119–164 (2019)
36. Feola, R., Iandoli, F.: Long time existence for fully nonlinear NLS with small Cauchy data on the circle. Annali della Scuola normale superiore di Pisa, Classe di scienze (2018). https://doi.org/10.2422/2036-2145.201811003
37. Feola, R., Procesi, M.: Quasi-periodic solutions for fully nonlinear forced reversible Schrödinger equations. J. Differ. Equ. **259**(7), 3389–3447 (2014)
38. Feola, R., Procesi, M.: KAM for quasi-linear autonomous NLS. Preprint 2017, arXiv:1705.07287
39. Giuliani, F.: Quasi-periodic solutions for quasi-linear generalized KdV equations. J. Differ. Equ. **262**(10), 5052–5132 (2017)
40. Giuliani, F.: KAM for quasi-linear PDEs. Ph.D. thesis. http://hdl.handle.net/20.500.11767/57306
41. Grébert, B., Kappeler, T.: The Defocusing NLS Equation and Its Normal Form, EMS Series of Lectures in Mathematics. European Mathematical Society Publishing House, Zürich (2014)
42. Hou, Y., Zhao, P., Fan, E., Qiao, Z.: Algebro-geometric solutions for the Degasperis–Procesi hierarchy. SIAM J. Math. Anal. **45**(3), 1216–1266 (2013)
43. Kappeler, T., Pöschel, J.: KAM and KdV. Springer, Berlin (2003)
44. Kuksin, S.: Hamiltonian perturbations of infinite-dimensional linear systems with imaginary spectrum. Funktsional. Anal. i Prilozhen. **21**(3), 192–205 (1987)
45. Kuksin, S.: A KAM theorem for equations of the Korteweg–de Vries type. Rev. Math. Phys. **10**(3), 1–64 (1998)
46. Kuksin, S., Perelman, G.: Vey theorem in infinite dimensions and its application to KdV. Discrete Contin. Dyn. Syst. **27**(1), 1–24 (2010)
47. Kuksin, S., Pöschel, J.: Invariant Cantor manifolds of quasi-periodic oscillations for a nonlinear Schrödinger equation. Ann. Math. **143**(1), 149–179 (1996)

48. Liu, Y., Yin, Z.: Global existence and blow-up phenomena for the Degasperis–Procesi equation. Commun. Math. Phys. **267**(3), 801–820 (2006)
49. Liu, J., Yuan, X.: A KAM theorem for Hamiltonian partial differential equations with unbounded perturbations. Commun. Math. Phys **307**, 629–673 (2011)
50. Montalto, R.: Quasi-periodic solutions of forced Kirchhoff equation. Nonlinear Differ. Equ. Appl. **24**(1), 9 (2017)
51. Pöschel, J.: A KAM-theorem for some nonlinear partial differential equations. Ann. Scuola Norm. Sup. Pisa Cl. Sci. (4) **23**(1), 119–148 (1996)
52. Procesi, C., Procesi, M.: A KAM algorithm for the non-linear Schrödinger equation. Adv. Math. **272**, 399–470 (2015)
53. Wayne, E.: Periodic and quasi-periodic solutions of nonlinear wave equations via KAM theory. Commun. Math. Phys. **127**, 479–528 (1990)
54. Yin, Z.: On the Cauchy problem for an integrable equation with peakon solutions. Ill. J. Math. **47**(3), 649–666 (2003)

Communicated by W. Schlag