Geography of resonances and Arnold diffusion in a priori unstable Hamiltonian systems

Amadeu Delshams† **and Gemma Huguet**‡

† Departament de Matemàtica Aplicada I, Universitat Politècnica de Catalunya, Diagonal 647, 08028 Barcelona, Spain

‡ Centre de Recerca Matemàtica, Apartat 50, 08193 Bellaterra (Barcelona), Spain

E-mail: Amadeu.Delshams@upc.edu, Gemma.Huguet@upc.edu

Abstract. In the present paper we consider the case of a general $\mathcal{C}^{r+2}$ perturbation, for r large enough, of an a priori unstable Hamiltonian system of $2+1/2$ degrees of freedom, and we provide explicit conditions on it, which turn out to be $\mathcal{C}^2$ generic and are verifiable in concrete examples, which guarantee the existence of Arnold diffusion.

This is a generalization of the result in Delshams et al., *Mem. Amer. Math. Soc.*, 2006, where it was considered the case of a perturbation with a finite number of harmonics in the angular variables.

The method of proof is based on a careful analysis of the geography of resonances created by a generic perturbation and it contains a deep quantitative description of the invariant objects generated by the resonances therein. The scattering map is used as an essential tool to construct transition chains of objects of different topology. The combination of quantitative expressions for both the geography of resonances and the scattering map provides, in a natural way, explicit computable conditions for instability.

AMS classification scheme numbers: 37J40, 37C29, 35B34, 34C29, 37C50

Submitted to: *Nonlinearity*

1. Introduction

The goal of this paper is to present a generalization of the geometric mechanism for global instability (popularly known as Arnold diffusion) in a priori unstable Hamiltonian systems introduced in [DLS06a]. That paper developed an argument to prove the existence of global instability in a-priori unstable nearly integrable Hamiltonian systems (the unperturbed Hamiltonian presents hyperbolicity, so that it can not be expressed globally in action-angle variables) and applied it to a model which presented the so called *large gap problem*. However, in that case, the perturbation was assumed to be a trigonometric polynomial in the angular variables. In this paper we perform an accurate process of truncation of the Fourier series of the perturbation and we present a deeper study of the *geography of resonances*. Using this, we are able to extend and simplify some of the results in [DLS06a] and apply them to an a priori unstable Hamiltonian system with a generic perturbation.

The phenomenon of global instability in Hamiltonian systems has attracted the attention of both mathematicians and physicists in the last years due to its remarkable importance for the applications. It deals, essentially, with the question of what is the effect on the dynamics when an autonomous mechanical system is submitted to a small periodic perturbation. More precisely, whether these perturbations accumulate over time giving rise to a long term effect or whether these effects average out.

The instability problem was formulated first by Arnold in 1964. In his celebrated paper [Arn64], Arnold constructed an example for which he proved the existence of trajectories that avoided the obstacles of KAM tori and performed long excursions. The mechanism is based on the existence of transition chains of whiskered tori, that is, sequences of tori such that the unstable manifold (whisker) of one of these tori intersects transversally the stable manifold (whisker) of the next one. By an obstruction argument, there is an orbit that follows this transition chain, giving rise to an unstable orbit.

The example proposed in [Arn64] turns out to be rather artificial because the perturbation was chosen in such a way that it preserved exactly the complete foliation of invariant tori existing in the unperturbed system. However, a generic perturbation of size ε creates gaps at most of size $\sqrt{\varepsilon}$ in the foliation of persisting primary KAM tori, whereas it moves the whiskers only by an amount ε. These gaps are centered around resonances, that is, resonant tori that are destroyed by the perturbation. This is what is known in the literature as the *large gap problem* (see, for instance, [Moe96] for a discussion about the large gap problem and, indeed, of the problem of diffusion).

In the last ten years there has been a notable progress in the comprehension of the mechanisms that give rise to the phenomenon of instability and a variety of methods has been suggested. As an example of this, we will mention that the *large gap problem* has been solved simultaneously by a variety of techniques: different geometrical methods [DLS00, DLS06a, DLS06b] (scattering map) and [Tre04, PT07] (separatrix map); topological methods [GL06b, GL06a] and variational methods [CY04a, CY04b]. For more information regarding the problem of Arnold diffusion in the absence of gaps as well as time estimates, the reader is referred to [DGLS08].

Of particular interest for the present paper are [DLS00, DLS06a, DLS06b]. The strategy in the mentioned papers is based on the incorporation of new invariant objects, created by the resonances, like secondary KAM tori and the stable and unstable manifolds of lower dimensional tori in the transition chain, together with the primary KAM tori. The scattering map, introduced by the same authors (see [DLS08] for a geometric study) is the essential tool for the heteroclinic connections between invariant objects of different topology.

In this paper we extend the geometric mechanism introduced in the mentioned papers to a wider class of model systems for which the perturbation does not need to have a finite number of harmonics in the angular variables. In particular, the Hamiltonian studied in this paper has the following form

$$H_\varepsilon(p,q,I,\varphi,t) = \pm\left(\frac{1}{2}p^2 + V(q)\right) + \frac{1}{2}I^2 + \varepsilon h(p,q,I,\varphi,t;\varepsilon), \qquad (1)$$

where $p \in (-p_0, p_0) \subset \mathbb{R}$, $I \in (I_-, I_+) \subset \mathbb{R}$ and $(q,\varphi,t) \in \mathbb{T}^3$.

The main result of this paper is Theorem 2.1, stated in section 2.2 with the concrete hypotheses for Hamiltonian (1), from which we can deduce the following short version:

Theorem 1.1. *Consider the Hamiltonian (1) and assume that V and h are $\mathcal{C}^{r+2}$ functions which are $\mathcal{C}^2$ generic, with $r > r_0$, large enough. Then there is $\varepsilon^* > 0$ such that for $0 < |\varepsilon| < \varepsilon^*$ and for any interval $[I_-^*, I_+^*] \in (I_-, I_+)$, there exists a trajectory $\widetilde{x}(t)$ of the system with Hamiltonian (1) such that for some $T > 0$*

$$I(\widetilde{x}(0)) \le I_-^*; \qquad I(\widetilde{x}(T)) \ge I_+^*.$$

Remark 1.2. A value of r_0 which follows from our argument is $r_0 = 242$ (see Remark 2.2).

Our strategy for the proof follows the geometric mechanism proposed in [DLS06a]. Indeed, in order to organize the different invariant objects that we will use to construct a transition chain, we will first identify the normally hyperbolic invariant manifold (NHIM) present in the system. This NHIM will have associated stable and unstable invariant manifolds that, generically, intersect transversally. Therefore, we can associate to this object two types of dynamics: the *inner* and the *outer*. The outer dynamics takes into account the asymptotic motions to the NHIM and is described by the scattering map. The inner dynamics is the one restricted to the NHIM and contains Cantor families of primary and secondary KAM tori. Since generically these families of KAM tori, invariant for the inner dynamics, are not invariant for the outer dynamics, the combination of both dynamics will allow us to construct a transition chain.

The results in [DLS06a] can be applied straightforwardly for the persistence of the NHIM and the transversality of its associated stable and unstable manifolds. The arguments presented in this paper focus on the inner dynamics and the study of the invariant objects present in the NHIM.

For Hamiltonian (1), resonances correspond to the places where the frequency $I = -l/k$ for $(k,l) \in \mathbb{Z}^2$ is rational and the associated Fourier coefficient $h_{k,l}$ of the perturbation h is nonzero. On these resonances, the foliation of KAM tori in the NHIM is destroyed and gaps between the Cantor family of invariant tori in the NHIM of size $\mathcal{O}(\varepsilon^{1/2}|h_{k_0,l_0}|^{1/2})$ are created, for (k_0,l_0) such that $l/k = l_0/k_0$ and $\gcd(k_0,l_0) = 1$ (see equation (86)). For a perturbation h which is a $\mathcal{C}^{r+2}$ function and $\mathcal{C}^2$ generic, when we restrict it to the NHIM and we write it in adequate coordinates we are left with a $\mathcal{C}^r$ perturbation (see the subsection "restriction to NHIM" in Section 2.3.3), so that $|h_{k,l}| \sim |(k,l)|^{-r}$, and therefore the above gaps are of size $\mathcal{O}(\varepsilon^{1/2}|(k_0,l_0)|^{-r/2})$. Moreover, other invariant objects, like secondary tori and lower dimensional tori, are created inside the gap. They correspond to invariant objects of different topology that were not present in the unperturbed system but are generated by the resonances.

In order to study their existence and give an approximate expression for them we will combine m steps of averaging plus a KAM Theorem. Notice that in our case, since the perturbation is generic, we will have an infinite number of resonances. Our approach for this study will be to consider an adequate truncation up to some order M, depending on ε, of the Fourier series of the perturbation h in such a way that we deal only with a finite number of harmonics $|(k,l)| \leq M$ and therefore of resonances.

Another remarkable difference with respect to the results obtained in [DLS06a] is that in that case the size of the gaps created in the foliation of invariant tori was uniform, whereas in our case, since the size is $\mathcal{O}(\varepsilon^{1/2}|(k_0,l_0)|^{-r/2})$, we have a heterogeneous sea of gaps of different sizes. Among them, we will distinguish between *small gaps* and *big gaps*, which are strongly related to the mentioned large gap problem. Indeed, big gaps are those of size bigger or equal than ε and therefore they are generated by resonances $-l_0/k_0$ of order one, such that $|(k_0,l_0)| < \varepsilon^{-1/r}$ or, equivalently, $|(k_0,l_0)|^{-r/2} \geq \varepsilon^{1/2}$ (see Section 3.3.3 for precise results).

From a more technical point of view (see Section 3.2 for details), we would like to remark that the main difficulties arise from the fact that in order to perform a resonant averaging procedure, we need to isolate resonances corresponding to $|(k,l)| \leq M$, for M depending on ε. Consequently, the width L of the resonant domain can not be chosen independently of ε, as it was the case in [DLS06a]. Moreover, along the averaging procedure we need to keep track of the $\mathcal{C}^\ell$ norms of the averaged terms and the remainders, and these blow up as a negative power of L. Hence, we will see that a good choice for L around a resonance $I = -l/k$ will be $L = L_k \sim \varepsilon^{1/n}/|k|$ (see hypotheses of Theorem 3.11), where n is the required regularity to apply KAM Theorem after the averaging procedure. Notice that L is not uniform along the resonances but depends on the value $|k|$ of the resonance.

Finally, after m steps of averaging, we will show that the remainder tail, that is, the Fourier coefficients $h_{k,l}$ such that $|(k,l)| > M$ can be neglected. This will be ensured by a fast enough decreasing rate of the coefficients and therefore a large enough regularity r of the perturbation. Thus, the required regularity r will be determined according to the number m of steps of averaging performed, as well as the needed regularity n to apply KAM Theorem after the averaging procedure.

We are using a version of the KAM theorem that requires to have the Hamiltonian system written in action angle variables. Since near the resonances we approximate the system by one which is close to a pendulum, the action variable becomes singular on the separatrix. This fact, together with the requirement to have the invariant objects close enough (at a distance smaller than ε) implies that the perturbation of the averaged Hamiltonian has to be extremely small in the resonant regions. The immediate consequence of this fact is that, in the case we are studying, one has to perform at least $m = 10$ steps of averaging (see Theorem 3.28). The needed regularity n to apply KAM Theorem after m averaging steps is $n = 2m+6$ (see Proposition 3.24). Since the regularity r required to ensure that the remainder tail is smaller than the averaging remainder turns out to be $r > (n-2)m+2$, see Remark 3.20, one has to impose $r > r_0 = 242$.

We do not claim that this is an optimal result. Actually, another version of the KAM theorem that allowed us to avoid the change into action-angle variables like [LGJV05, FLS07] could improve the results in terms of the needed regularity (see also [LHS08] for a numerical implementation). However, it is worth mentioning that we managed to decrease the required steps of averaging in the resonant domains with respect to the results in [DLS06a]. Since in the resonances the behavior of KAM tori is different depending on how close they are to the separatrix (tori are flatter as they are further from the separatrix), we consider different regions where we perform different scalings. This strategy, which was already introduced in [DLS06a], has been improved in this paper introducing a new sequence of domains in Theorem 3.30. When applied to the case with a finite number of resonances as in [DLS06a], $m = 9$ steps of averaging and $r \geq 26$ are enough (see Remark 3.32). This clearly improves the needed regularity r which was $r \geq 60$ in [DLS06a] because m was chosen $m = 26$.

Sections 3.3.3, 3.3.4 and 3.3.5 contain a quantitative description of the *geography of resonances* and a detailed study of the invariant objects generated by the resonances. The effect of the resonances in a system constitutes a fundamental problem not only for diffusion but also for many other physical applications and it has been an important object of study in the physical literature, see for instance [Chi79, Ten82]. The study performed in this paper contributes to understand better the different types of resonances and the geometric objects that one can find therein and can be very helpful in many physical problems.

Moreover, we think that this study can be extended to a class of models that present multiple resonances, see [DLS07].

We would like to emphasize that in our case, and this is different from the results in [DLS06a], only the resonances of order 1, that is, the ones that appear at the first step of averaging, create big gaps; whereas in [DLS06a], both resonances of order 1 and 2 could generate big gaps. This is because we are dealing with a perturbation that generically will have all the harmonics different from zero. This means that the effect of the resonances associated to the biggest Fourier coefficients (low frequencies) will be detected at the first step of averaging. Since the size of the gap depends on both the order of the resonance and the size of the Fourier coefficient associated to that resonance, the ones that appear at the second step of averaging already correspond to small Fourier coefficients and the size of their gap will be smaller than ε. The immediate consequence of this fact it that in the forthcoming Theorem 2.1, we can give all conditions *explicitly* in terms of the original perturbation h.

The paper is organized in the following way. In Section 2 we state Theorem 2.1, which establishes the existence of diffusing orbits for the model considered under precise conditions. Since the required hypotheses are checked to be $\mathcal{C}^2$ generic, Theorem 1.1 follows straightforwardly. The proof of Theorem 2.1 is given in Section 2, except for two technical results, Theorem 3.1 and Proposition 4.1, which are postponed to the following sections.

Thus, in Section 3, we prove Theorem 3.1, which provides a quantitative existence of invariant objects for the inner dynamics in the NHIM following the steps indicated in Section 2. In Section 4, we use the scattering map to prove Proposition 4.1 about the existence of heteroclinic connections between the invariant objects obtained in the previous section.

We would like to remark that, in contrast to [DLS06a], and thanks to the new results about the scattering map obtained in [DLS08], we use the Hamiltonian function generating the deformation of the scattering map instead of the scattering map itself, in order to compute the images of the leaves of a certain foliation under the scattering map.

Finally, in Section 5 we have included for illustration a concrete example, for which we sketch how the hypotheses of Theorem 2.1 can be checked. We plan to come back to this example in a future paper for a more detailed description of the mechanism. In the Appendix, we have brought some technical results used in the paper.

2. Statement of results

Before stating the main result in this paper we need to introduce some notation.

2.1. Notation and preliminaries

Let r be a positive integer and $\mathcal{D} \subset \mathbb{R}^n$ a compact set with nonempty interior $\mathring{\mathcal{D}}$. We will denote the set of $\mathcal{C}^r$ functions from $\mathring{\mathcal{D}}$ to $\mathbb{R}^m$ and continuous on $\mathcal{D}$ by $\mathcal{C}^r(\mathcal{D}, \mathbb{R}^m)$. When $m = 1$, we simply write $\mathcal{C}^r(\mathcal{D})$ instead of $\mathcal{C}^r(\mathcal{D}, \mathbb{R}^m)$. Given $f \in \mathcal{C}^r(\mathcal{D}, \mathbb{R}^m)$, we consider the standard $\mathcal{C}^r$ norm,

$$|f|_{\mathcal{C}^r(\mathcal{D})} = \sum_{i=1}^{m} \sum_{\ell=0}^{r} \sum_{|\alpha|=\ell} \sup_{x\in\mathcal{D}} \frac{|D^\alpha f_i(x)|}{\alpha!}, \tag{2}$$

where f_i denotes the i-th component of the function f, for $i = 1, \ldots, m$. We omit the domain in the notation when it does not lead to confusion.

We use the standard multi-index notation: if $\alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbb{N}^n$ and $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ one sets

$$|\alpha| = \alpha_1 + \cdots + \alpha_n$$

$$\alpha! = \alpha_1!\alpha_2! \cdots \alpha_n!$$

$$D^\alpha = \frac{\partial^{|\alpha|}}{\partial x_1^{\alpha_1} \ldots \partial x_n^{\alpha_n}}$$

In the case that the function f depends only on a few of the variables, we will denote it in the same way, that is $|f|_{\mathcal{C}^r} = |f|_{\mathcal{C}^r(\mathcal{D})}$, and consider it as a function of more variables defined in the appropriate domain.

Note that we denote $|f|_{\mathcal{C}^0} = \sup_{x\in\mathcal{D}} |f(x)|$, which is the standard supremum norm, so the $|\cdot|_{\mathcal{C}^r(\mathcal{D})}$ norm can be expressed, equivalently, as

$$|f|_{\mathcal{C}^r(\mathcal{D})} := \sum_{i=1}^{m}\sum_{\ell=0}^{r}\sum_{|\alpha|=\ell} \frac{|D^\alpha f_i|_{\mathcal{C}^0(\mathcal{D})}}{\alpha!}.$$

The space of $\mathcal{C}^r(\mathcal{D})$ functions endowed with the $\mathcal{C}^r$ norm is a Banach algebra (see [Con90]), that is, it is a Banach space with the property that given any two functions f, g in $\mathcal{C}^r(\mathcal{D})$, they satisfy

$$|fg|_{\mathcal{C}^\ell} \le |f|_{\mathcal{C}^\ell}\, |g|_{\mathcal{C}^\ell}\,.$$

Since we will also deal with $\mathcal{C}^r$ functions defined on a compact domain $\mathcal{D} = \mathcal{I}\times\mathbb{T}^n$, where $\mathcal{I}\subset\mathbb{R}^n$ is a compact set with non empty interior, we can also consider the following seminorm, that takes into account the different regularities and the estimates for the derivatives in each type of variable:

$$|f|_{\ell_1,\ell_2} := \sum_{m_1=0}^{\ell_1}\sum_{m_2=0}^{\ell_2}\sum_{\substack{\alpha_1,\alpha_2\in\mathbb{N}^n\\ |\alpha_1|=m_1,|\alpha_2|=m_2}} \frac{1}{\alpha_1!\alpha_2!}\sup_{(I,\varphi)\in\mathcal{D}}\left|\frac{\partial^{m_1+m_2} f(I,\varphi)}{\partial I^{\alpha_1}\partial\varphi^{\alpha_2}}\right|, \tag{3}$$

for $0\le \ell_1+\ell_2\le r$.

Note that $|f|_{\mathcal{C}^\ell} = \sum_{m=0}^{\ell} |f|_{m,\ell-m}$, for $0\le\ell\le r$.

We will use the following notation, which is rather usual. Given $\alpha = \alpha(\varepsilon)$ and $\beta = \beta(\varepsilon)$, we will write $\alpha \preceq \beta$ and also $\alpha = \mathcal{O}(\beta)$ if there exists ε_0 and a constant C independent of ε, such that $|\alpha(\varepsilon)| \le C|\beta(\varepsilon)|$, for $|\varepsilon|\le\varepsilon_0$. When we have $\alpha\preceq\beta$ and $\beta\preceq\alpha$ we will write $\alpha\sim\beta$. However, in some informal discussions we will abuse notation and we will say that α is of order $\varepsilon^p \Leftrightarrow \alpha\sim\varepsilon^p$.

We will say that a function $f = \mathcal{O}_{\mathcal{C}^r(\mathcal{D})}(\beta)$ when

$$|f|_{\mathcal{C}^r(\mathcal{D})} \preceq \beta.$$

2.2. Set up and main result

We consider a 2π-periodic in time perturbation of a pendulum and a rotor described by the non-autonomous Hamiltonian (1),

$$\begin{aligned} H_\varepsilon(p,q,I,\varphi,t) &= H_0(p,q,I) + \varepsilon h(p,q,I,\varphi,t;\varepsilon) \\ &= P_\pm(p,q) + \frac{1}{2}I^2 + \varepsilon h(p,q,I,\varphi,t;\varepsilon), \end{aligned} \tag{4}$$

where

$$P_\pm(p,q) = \pm\left(\frac{1}{2}p^2 + V(q)\right) \tag{5}$$

and $V(q)$ is a 2π-periodic function. We will refer to $P_\pm(p,q)$ as the *pendulum*.

The term $\frac{1}{2}I^2$ describes a *rotor* and the final term εh is the perturbation term and depends periodically on time, so that h can be expressed via its Fourier series in the variables (φ, t)

$$h(p,q,I,\varphi,t;\varepsilon) = \sum_{(k,l)\in\mathbb{Z}^2} h_{k,l}(p,q,I;\varepsilon)e^{\mathrm{i}(k\varphi+lt)}. \tag{6}$$

It will be convenient to consider the autonomous system by introducing the extra variables (A, s):

$$\begin{aligned} \widetilde{H}_\varepsilon(p,q,I,\varphi,A,s) &= A + H_0(p,q,I) + \varepsilon h(p,q,I,\varphi,s;\varepsilon) \\ &= A + P_\pm(p,q) + \frac{1}{2}I^2 + \varepsilon h(p,q,I,\varphi,s;\varepsilon) \end{aligned} \tag{7}$$

where the pairs $(p,q) \in \mathbb{R} \times \mathbb{T}$, $(I,\varphi) \in \mathbb{R} \times \mathbb{T}$ and $(A,s) \in \mathbb{R} \times \mathbb{T}$ are symplectically conjugate.

The extra variable A does not play any dynamical role; it is symplectically conjugate to the variable s and simply makes the system autonomous. So, we are only interested in studying the dynamics of variables (p,q,I,φ,s), given by the system of equations:

$$\begin{aligned} \dot{p} &= \mp V'(p) - \varepsilon \frac{\partial h}{\partial q}(p,q,I,\varphi,s;\varepsilon) \\ \dot{q} &= \pm p + \varepsilon \frac{\partial h}{\partial p}(p,q,I,\varphi,s;\varepsilon) \\ \dot{I} &= - \varepsilon \frac{\partial h}{\partial \varphi}(p,q,I,\varphi,s;\varepsilon) \\ \dot{\varphi} &= I + \varepsilon \frac{\partial h}{\partial I}(p,q,I,\varphi,s;\varepsilon) \\ \dot{s} &= 1 \end{aligned} \tag{8}$$

The domain of definition we consider is a compact set of type

$$\mathcal{D} := \mathcal{S} \times [I_-, I_+] \times \mathbb{T}^2 \times [-\varepsilon_0, \varepsilon_0],$$

where $\mathcal{S} \subset \mathbb{R} \times \mathbb{T}$ is a neighborhood of the separatrix $(P_\pm^{-1}(0))$ of the pendulum.

Then, the main Theorem of this paper is:

Theorem 2.1. *Consider a Hamiltonian of the form (1) where V and h are $\mathcal{C}^{r+2}$ in $\mathcal{D}$, with $r > r_0$, sufficiently large. Assume also that,*

H1 *The potential $V : \mathbb{T} \to \mathbb{R}$ has a unique global maximum, say at $q = 0$, which is non-degenerate (i.e. $V''(0) < 0$). We denote by $(p_0(t), q_0(t))$ an orbit of the pendulum $P_\pm(p,q)$ in (1), homoclinic to $(0,0)$.*

H2 *Consider the* Poincaré function, *also called* Melnikov potential, *associated to h (and to the homoclinic orbit (p_0, q_0)):*

$$\begin{aligned} \mathcal{L}(I,\varphi,s) = - \int_{-\infty}^{+\infty} & (h(p_0(\sigma), q_0(\sigma), I, \varphi + I\sigma, s + \sigma; 0) \\ & - h(0,0,I,\varphi + I\sigma, s + \sigma; 0)) d\sigma \end{aligned} \tag{9}$$

H2' *Given real numbers $I_- < I_+$, assume that, for any value of $I \in (I_-, I_+)$, there exists an open set $\mathcal{J}_I \in \mathbb{T}^2$, with the property that when $(I, \varphi, s) \in H_+$, where*

$$H_+ = \bigcup_{I \in (I_-, I_+)} \{I\} \times \mathcal{J}_I \subset (I_-, I_+) \times \mathbb{T}^2, \tag{10}$$

the map

$$\tau \in \mathbb{R} \mapsto \mathcal{L}(I, \varphi - I\tau, s - \tau)$$

has a non-degenerate critical point τ which is locally given by the implicit function theorem in the form $\tau = \tau^*(I, \varphi, s)$, with τ^* a smooth function.

H2" *Introduce the* reduced Poincaré function $\mathcal{L}^*$ *defined by*

$$\mathcal{L}^*(I, \varphi) := \mathcal{L}(I, \varphi - I\tau^*(I, \varphi, 0), -\tau^*(I, \varphi, 0)), \tag{11}$$

on

$$H_+^* = \{(I, \widetilde{\theta}) : \widetilde{\theta} = \varphi - Is, (I, \varphi, s) \in H_+\} = \bigcup_{I \in (I_-, I_+)} \{I\} \times \mathcal{J}_I^*, \tag{12}$$

and assume that

$$\widetilde{\theta} \mapsto \frac{\partial \mathcal{L}^*}{\partial \widetilde{\theta}}(I, \widetilde{\theta})$$

for $\widetilde{\theta} = \varphi - Is \in \mathcal{J}_I^*$ is non-constant and positive (respectively negative).

H3 *Fix $1/(r/6 - 1) < \nu \leq 1/16$, for any $0 < \varepsilon < 1$ and for any $(k_0, l_0) \in \mathbb{Z}^2$ with $\gcd(k_0, l_0) = 1$ and $|(k_0, l_0)| < M_{\mathrm{BG}}$, where $|(k_0, l_0)| = \max(|k_0|, |l_0|)$ and $M_{\mathrm{BG}} = \varepsilon^{-(1+\nu)/r}$, introduce the 2π-periodic function*

$$U^{k_0, l_0}(\theta) = \sum_{\substack{t \in \mathbb{Z} - \{0\}, \\ |t||(k_0, l_0)| < M}} h_{tk_0, tl_0}(0, 0, -l_0/k_0; 0) e^{\mathrm{i}t\theta},$$

where $\theta = k_0\varphi + l_0 s$ and $M = \varepsilon^{-1/(26+\delta)}$, for δ small, for which we assume:

H3' *The function U^{k_0, l_0} has a non-degenerate global maximum.*

H3" *For $|(k_0, l_0)| \prec \varepsilon^{-1/r}$, we assume that the $2\pi k_0$-periodic function f given by*

$$f(\theta) = \frac{k_0 U'^{k_0, l_0}(\theta) \frac{\partial \mathcal{L}^*}{\partial \widetilde{\theta}} \left(\frac{-l_0}{k_0}, \frac{\theta}{k_0} \right) + 2U^{k_0, l_0}(\theta) \frac{\partial^2 \mathcal{L}^*}{\partial \widetilde{\theta}^2} \left(\frac{-l_0}{k_0}, \frac{\theta}{k_0} \right)}{2 \frac{\partial^2 \mathcal{L}^*}{\partial \widetilde{\theta}^2} \left(\frac{-l_0}{k_0}, \frac{\theta}{k_0} \right)} \tag{13}$$

is non-constant.

H3"' *For $|(k_0, l_0)| \sim \varepsilon^{-1/r}$, we assume the non-degeneracy condition stated explicitly in equation (152).*

Then, there exists $\varepsilon^* > 0$ such that for $0 < |\varepsilon| < \varepsilon^*$ and for any interval $[I_-^*, I_+^*] \in (I_-, I_+)$, there exists a trajectory $\widetilde{x}(t)$ of the system (1) such that for some $T > 0$

$$I(\widetilde{x}(0)) \leq I_-^*; \qquad I(\widetilde{x}(T)) \geq I_+^*.$$

(respectively:

$$I(\widetilde{x}(0)) \geq I_+^*; \qquad I(\widetilde{x}(T)) \leq I_-^*).$$

Remark 2.2. r_0 depends on the number m of some averaging steps performed in the proof: $r_0 = 2(m+1)^2$ and $m \geq 10$ (see hypotheses of Theorem 3.1 in Section 3). If we take just $m = 10$ then $r_0 = 242$ is enough.

Remark 2.3. The truncation order M in hypotheses **H3** depends on the regularity n required for the application of the KAM Theorem along the proof: $M = \varepsilon^{-1/(n+\delta)}$, for $n = 2m + 6$ and $0 < \delta < 1/m$, where m is the number of averaging steps performed in the proof and is such that $m \geq 10$ (see hypotheses of Theorems 3.11 and 3.1 and Remark 3.20). Hence, we choose $m = 10$ and therefore $M = \varepsilon^{-1/(26+\delta)}$ in hypotheses **H3**.

Remark 2.4. Notice that for every fixed ε we have one condition **H3** for every (k_0, l_0) such that $|(k_0, l_0)| < M_{\mathrm{BG}}$, that depends explicitly on (k_0, l_0). Hence, the number of non-degeneracy conditions **H3** is finite but grows with ε.

Remark 2.5. Notice that by the definition of $\tau^*(I, \varphi, s)$, the function

$$f(I, \varphi, s) = \mathcal{L}(I, \varphi - I\tau^*(I, \varphi, s), s - \tau^*(I, \varphi, s))$$

satisfies the equation

$$I\partial_\varphi f(I, \varphi, s) + \partial_s f(I, \varphi, s) = 0.$$

Therefore it is of the form $f(I, \varphi, s) = \mathcal{L}^*(I, \varphi - Is)$, so we can alternatively define

$$\mathcal{L}^*(I, \varphi - Is) = \mathcal{L}(I, \varphi - I\tau^*(I, \varphi, s), s - \tau^*(I, \varphi, s)).$$

Remark 2.6. The main feature of Theorem 2.1, as already said in the Introduction, is that h is not required to be a trigonometric polynomial in the variables (φ, s), which is a non-generic assumption, as it was the case in [DLS06a].

Before proving Theorem 2.1 let us see that Theorem 1.1 stated in the Introduction is just a consequence of Theorem 2.1. Indeed, for every fixed ε, conditions **H1** and **H2** are open and dense, that is they hold for an open and dense set of Hamiltonians in the $\mathcal{C}^2$ topology.

For every fixed ε, the number of non-degeneracy conditions **H3** is finite but grows with ε (the number of conditions depends on $(k_0, l_0) \in \mathbb{Z}^2$ such that $\gcd(k_0, l_0) = 1$ and $|(k_0, l_0)| \preceq \varepsilon^{-1/r}$). When ε tends to 0 we have a countable number of conditions in terms of the functions

$$U_\infty^{k_0,l_0}(\theta) = \sum_{t\in\mathbb{Z}-\{0\}} h_{tk_0,tl_0}(0, 0, -l_0/k_0; 0)e^{\mathrm{i}t\theta},$$

which are the same as those in hypotheses **H3** but without any truncation. This countable number of conditions involve only derivatives up to order 2 of the Hamiltonian, hence the set of Hamiltonians satisfying them is a residual set in the $\mathcal{C}^2$ topology, that is, a countable intersection of open and dense sets in the $\mathcal{C}^2$ topology.

Therefore the hypotheses of the Theorem are $\mathcal{C}^2$ generic in the set of $\mathcal{C}^{r+2}$ Hamiltonians of the form (1). So, the short version of Theorem 2.1 stated in Theorem 1.1 in the Introduction follows straightforwardly.

2.3. Proof of Theorem 2.1

The proof of this theorem follows the geometric mechanism stated in [DLS06a] and it is organized in four parts that we first sketch now:

Part 1 The first part deals with the existence of a Normally Hyperbolic Invariant Manifold (NHIM), which jointly with its associated stable and unstable manifolds, organizes all the dynamics, and is a consequence of hypothesis **H1**. By hypothesis **H2'**, its associated stable and unstable manifolds will intersect transversally, so we can associate to this object two types of dynamics: the inner and the outer.

Part 2 The outer dynamics, which is the one that takes into account the asymptotic motions to the NHIM, is studied in the second part. We will see that we can associate a scattering map to the NHIM and give formulas for the Hamiltonian function which determines the deformation of this scattering map.

Part 3 The third part of the proof consists of studying the inner dynamics, that is the one restricted to the NHIM. The goal is to show that, by hypotheses **H3'**, there exists a discrete foliation of invariant tori, which are closely spaced. Among these tori, some of them are primary, so they are just a continuation of the ones that existed for the integrable system ($\varepsilon = 0$), and some of them are secondary, these ones are contractible to a periodic orbit, so they correspond to motions with topologies that were not present in the unperturbed system but they are created by the resonances. The method of proof will be a combination of an averaging procedure and a quantitative version of KAM Theorem, which requires the Hamiltonian to be differentiable enough.

Part 4 The last part of the proof consists of showing that the combination of both types of dynamics give rise to a construction of a transition chain, that is, a sequence of whiskered tori in which the stable manifold of one torus intersects transversally the unstable manifold of the next one. To this end, one needs to show that the discrete foliation of whiskered tori which are invariant under the (inner) flow is not invariant under the scattering map or outer map. This is ensured by hypotheses **H2"**, **H3"** and **H3"'** in Theorem 2.1, which indeed provide the transversality of this discrete foliation to the scattering map. Finally we prove, using a standard obstruction property, that there is an orbit that follows this transition chain.

Next we give a proof of Theorem 2.1 organized in the four parts that we have mentioned. The first two parts follow readily from [DLS06a] and Theorems stated in [DLS06a] apply straightforwardly because hypotheses **H1** and **H2'** required for the proof of the mentioned results are the same as in our case. Moreover, for the second part we use the symplectic properties developed in [DLS08] to generalize the computation of the scattering map using its Hamiltonian function. So, for these parts, we only refer in Section 2.3.1 and 2.3.2, to the results in [DLS06a] and [DLS08] that we are using.

However, for the third part, the results obtained in [DLS06a] do not apply directly because in the present paper we are not assuming that the perturbation has a finite number of harmonics. Therefore, it has been necessary to develop a new methodology in order to prove that when we have a $\mathcal{C}^{r+2}$ perturbation h, with r large enough, and hypotheses **H3'** are fulfilled, for every ε we can truncate adequately its Fourier series and deal only with a finite number of harmonics and therefore a finite number of resonances to get a discrete foliation of tori closely spaced. Moreover, explicit approximate expressions for these tori are obtained as the level sets of a certain function. The mentioned results are stated and proved rigorously in Section 3, giving rise to Theorem 3.1 and they constitute the essential result of this paper. In Section 2.3.3 we just refer to the results in Section 3 needed to prove part 3 of Theorem 2.1.

Once we have fixed in part 3, for every ε, the number of resonances, part 4 follows readily from the finite hypotheses **H2"**, **H3"** and **H3"'** as in [DLS06a]. The main difference is that, in contrast to [DLS06a] and thanks to the new results about the symplectic properties of the scattering map obtained in [DLS08], we can use the Hamiltonian function generating the deformation of the scattering map instead of the scattering map itself, in order to compute the images of the leaves of a certain foliation under the scattering map. The results with their proof are stated in Section 4. In Section 2.3.4 we just refer to the results in Section 4 needed to prove part 4 of Theorem 2.1.

2.3.1. First Part: Existence of a NHIM and its associated stable and unstable manifolds
The method of proof is based on the existence of an invariant object, a NHIM (see, for instance, [HPS77, Fen74, Fen77, Fen79, Lla00, Wig90] for the standard theory of NHIMs used in this paper), which jointly with its associated stable and unstable manifolds, organizes all the dynamics around it.

We start by discussing the geometric features of the unperturbed case which will survive under the perturbation. For the case $\varepsilon = 0$, Hamiltonian (1) is integrable and consists of two uncoupled systems: a rotor and a pendulum. So, the cartesian product of invariant objects of each of these subsystems will give an invariant object of the full system. Then, by hypothesis **H1**, if we consider the product of the hyperbolic fixed point $(p, q) = (0, 0)$ of the pendulum $P_\pm(p, q)$ in (5) with all the other variables, we have that for the values I_-, I_+ given in Theorem 2.1, the set

$$\widetilde{\Lambda} = \{\widetilde{x} = (p, q, I, \varphi, s) \in (\mathbb{R} \times \mathbb{T})^2 \times \mathbb{T} : p = q = 0, I \in [I_-, I_+]\} \qquad (14)$$

is a 3-dimensional invariant manifold and normally hyperbolic for the flow of the Hamiltonian system (8) for $\varepsilon = 0$. The associated stable and unstable invariant manifolds of $\widetilde{\Lambda}$ are the ones inherited from the separatrices of the pendulum (stable and unstable manifolds of the hyperbolic fixed point) and they agree:

$$W^s\widetilde{\Lambda} = W^u\widetilde{\Lambda} = \{(p_0(\tau), q_0(\tau), I, \varphi, s) : \tau \in \mathbb{R}, I \in [I_-, I_+], (\varphi, s) \in \mathbb{T}^2\} \qquad (15)$$

where $(p_0(\tau), q_0(\tau))$ is the chosen orbit of the pendulum $P_\pm$, provided by hypothesis **H1**, which is homoclinic to the hyperbolic fixed point $(0, 0)$.

The Hamiltonian system (8) for $\varepsilon = 0$ restricted to the manifold $\widetilde{\Lambda}$ is given simply by

$$\dot{I} = 0, \quad \dot{\varphi} = I, \quad \dot{s} = 1.$$

The dynamics on this manifold is very simple: all the solutions lie on a 2-dimensional invariant torus $I = \text{cte}$. Therefore, the normally hyperbolic invariant manifold is foliated by a one-parameter family of 2-dimensional invariant tori indexed by I, with associated frequency $(I, 1)$.

For $0 < |\varepsilon| \ll 1$, by the theory of NHIM (see the references above), the manifold $\widetilde{\Lambda}$ persists, giving rise to another manifold $\widetilde{\Lambda}_\varepsilon$ with associated local stable and unstable manifolds $W^{s,\text{loc}}\widetilde{\Lambda}_\varepsilon$ and $W^{u,\text{loc}}\widetilde{\Lambda}_\varepsilon$, which can be prolonged to $W^s\widetilde{\Lambda}_\varepsilon$ and $W^u\widetilde{\Lambda}_\varepsilon$, respectively. Both $\widetilde{\Lambda}_\varepsilon$ and its local stable and unstable manifolds, $W^{s,\text{loc}}\widetilde{\Lambda}_\varepsilon$ and $W^{u,\text{loc}}\widetilde{\Lambda}_\varepsilon$, are ε-close in the $\mathcal{C}^r$ sense to the unperturbed ones:

$$\widetilde{\Lambda}_\varepsilon = \widetilde{\Lambda} + \mathcal{O}_{\mathcal{C}^r}(\varepsilon); \quad W^{s,\text{loc}}\widetilde{\Lambda}_\varepsilon = W^{s,\text{loc}}\widetilde{\Lambda} + \mathcal{O}_{\mathcal{C}^r}(\varepsilon); \quad W^{u,\text{loc}}\widetilde{\Lambda}_\varepsilon = W^{u,\text{loc}}\widetilde{\Lambda} + \mathcal{O}_{\mathcal{C}^r}(\varepsilon). \quad (16)$$

The result of the persistence of the NHIM $\widetilde{\Lambda}_\varepsilon$ and its stable and unstable manifolds is formulated in Theorem 7.1 of [DLS06a], where the perturbation h in (1) was assumed to be a trigonometric polynomial. However, the only assumption required for the proof was the fact that the perturbation h and the potential V were $\mathcal{C}^{r+2}$, so Theorem 7.1 can be applied straightforwardly in our case.

2.3.2. Second Part: Outer Dynamics The outer dynamics, which is the one that takes into account the asymptotic motion to the NHIM $\widetilde{\Lambda}_\varepsilon$, is described by the scattering map. It is possible to construct a scattering map associated to the NHIM $\widetilde{\Lambda}_\varepsilon$, as long as its stable and unstable manifolds intersect transversally.

In Proposition 9.2 in [DLS06a] it is proved that if hypothesis **H2'** in Theorem 2.1 is satisfied, then the stable and unstable manifolds $W^s\widetilde{\Lambda}_\varepsilon$ and $W^u\widetilde{\Lambda}_\varepsilon$ of the NHIM intersect transversally along a homoclinic manifold Γ_ε, which is also called a *homoclinic channel* (see [DLS08] for more details, in particular for the definition of the *wave operators*, needed for the construction of the scattering map). So, we will be able to locally define the scattering map associated to Γ_ε and compute it in first order perturbation theory using the results in [DLS08]. Again, hypothesis **H2'** required for Proposition 9.2 in [DLS08] does not depend on the number of harmonics of the perturbation h, so the results stated also hold for the case we are considering in this paper.

Therefore, the manifold $\widetilde{\Lambda}_\varepsilon$ defined in (14) has a scattering map associated to a homoclinic manifold Γ_ε, defined in the following way

$$\begin{array}{rccc} S_\varepsilon: & H_+ \subset \widetilde{\Lambda}_\varepsilon & \to & \widetilde{\Lambda}_\varepsilon \\ & x_- & \mapsto & x_+ \end{array} \quad (17)$$

such that $x_+ = S(x_-) \Leftrightarrow \exists\ z \in \Gamma_\varepsilon$ such that

$$\text{dist}(\Phi_{t,\varepsilon}(z), \Phi_{t,\varepsilon}(x_\pm)) \to 0 \qquad \text{for}\, t \to \pm\infty,$$

where $\Phi_{t,\varepsilon}$ is the flow of Hamiltonian (1). Indeed,

$$|\Phi_{t,\varepsilon}(z) - \Phi_{t,\varepsilon}(x_\pm)| \leq \text{cte}\, e^{-\mu|t|/2} \qquad \text{for}\, t \to \pm\infty,$$

where $\mu = \sqrt{-V''(0)} > 0$ is the characteristic exponent of the saddle point $(0,0)$ of the pendulum $P_\pm(p,q)$ in (5).

Heuristically, the scattering map maps points of the manifold $\widetilde{\Lambda}_\varepsilon$ to points of the manifold $\widetilde{\Lambda}_\varepsilon$, such that the motion of z synchronizes with that of x_- (and x_+) in the past (and in the future).

Moreover, in Proposition 9.2 in [DLS06a] it is given a perturbative formula for the difference of the actions I of the points $x_+ = S_\varepsilon(x_-)$ and x_-. Concretely, expressing the points $x_\pm$ in terms of the parametrization of $\widetilde{\Lambda}_\varepsilon$, given in Theorem 7.1 in [DLS06a] we have that

$$I(x_\pm) = I + \mathcal{O}_{\mathcal{C}^1}(\varepsilon),\ \varphi(x_\pm) = \varphi + \mathcal{O}_{\mathcal{C}^1}(\varepsilon),\ s(x_\pm) = s,$$

and

$$I(x_+) - I(x_-) = \varepsilon \frac{\partial \mathcal{L}^*}{\partial \widetilde{\theta}}(I, \widetilde{\theta}) + \mathcal{O}_{\mathcal{C}^1}(\varepsilon^2), \tag{18}$$

for $\widetilde{\theta} = \varphi - Is$, where $\mathcal{L}^*$ is the reduced Poincaré function defined in hypothesis **H2”**.

Remark 2.7. Notice that there is a wrong sign in formula (9.9) in [DLS06a].

The method used in [DLS06a], based on the fact that I is a slow variable, allowed only to compute the leading term of the action component of the scattering map, but not the φ component since it is not a slow variable.

In a more recent paper [DLS08] the authors showed that the scattering map is exact symplectic and introduced geometric methods that allow to compute perturbatively an expression for both fast and slow variables.

Thus, using the method proposed in Section 5 in [DLS08], we can give perturbative formulas for the Hamiltonian $\mathcal{S}_\varepsilon$ generating the deformation of the scattering map S_ε.

It follows straightforwardly from Theorem 31 in [DLS08] that the reduced Poincaré function $\mathcal{L}^*$ introduced in (11) is equal to the Hamiltonian $-\mathcal{S}_0$, so that we obtain

$$\mathcal{S}_\varepsilon(I, \varphi, A, s) = -\mathcal{L}^*(I, \widetilde{\theta}) + \mathcal{O}(\varepsilon), \tag{19}$$

with $\widetilde{\theta} = \varphi - Is$.

Hence, the first order perturbative term of the scattering map is given by

$$S_\varepsilon(I, \varphi, A, s) = (I, \varphi, A, s) + \varepsilon \mathrm{J}\nabla \mathcal{S}_0(I, \varphi, A, s) + \mathcal{O}(\varepsilon^2), \tag{20}$$

where J is the canonical matrix of the symplectic form $\omega = \mathrm{d}I \wedge \mathrm{d}\varphi + \mathrm{d}A \wedge \mathrm{d}s$ and $\nabla = (\partial_I, \partial_\varphi, \partial_A, \partial_s)$. The extra variable A, conjugated to the angle s, was introduced to make apparent the symplectic character of the scattering map.

Notice that equation (18) is just the I component of equation (20).

We would like to remark that $S_\varepsilon = \mathrm{Id} + \mathcal{O}(\varepsilon)$. In particular, one iteration of S_ε can only jump distances of order ε in the action direction I.

Remark 2.8. For the mechanism of diffusion we are interested in comparing the inner dynamics in $\widetilde{\Lambda}_\varepsilon$ with the outer dynamics provided by the scattering map S_ε. Although the computation up to first order of the scattering map for the I component is enough for our purposes, it is more natural to study the action of the scattering map in terms of the Hamiltonian $\mathcal{S}_\varepsilon$.

2.3.3. Third Part: Inner Dynamics In this section we study the inner dynamics, that is, the dynamics of the flow of Hamiltonian (1) restricted to the NHIM $\widetilde{\Lambda}_\varepsilon$. The main result is Theorem 3.1, which states that there exists a discrete sequence of invariant tori $\mathcal{T}_i$ in the NHIM $\widetilde{\Lambda}_\varepsilon$, which are distributed along the actions in the interval (I_-, I_+) introduced in Theorem 2.1 and which are $\mathcal{O}(\varepsilon^{1+\eta})$-closely spaced in terms of the action variable, for some $\eta > 0$. Moreover, Theorem 3.1 provides explicit approximate expressions for the invariant tori, which are of two types depending on the region of the phase of space where invariant tori lie: the *big gaps region* and the *flat tori region.*

The *big gaps region* is defined as

$$\mathcal{D}_{\mathrm{BG}} = \{(I, \varphi, s) \in (I_-, I_+) \times \mathbb{T}^2 : |I + l/k| \leq L/|k|, |(k,l)| < M_{\mathrm{BG}}\} \quad (21)$$

where L is defined in (54) and is going to be introduced precisely along this third Part and M_{BG} was introduced in hypothesis **H3** of Theorem 2.1. For the purpose of this exposition it is enough to know now that $L = \mathcal{O}(\varepsilon^{1/n})$ and $M_{\mathrm{BG}} = \mathcal{O}(\varepsilon^{-1/r})$, where n is the regularity of the Hamiltonian required for the application of KAM Theorem ($n = 26$ will be enough, see hypotheses of Theorem 3.1) and r ($r > n$) is the regularity of the Hamiltonian required for Theorem 2.1. The flat tori region is the complementary region of the big gaps region.

In the flat tori region, there exists a Cantorian foliation of primary KAM tori, which are just a continuation of invariant tori $I = \mathrm{cte}$ present in $\widetilde{\Lambda}_0$ for the unperturbed Hamiltonian (1) for $\varepsilon = 0$.

The big gaps region is formed by gaps of size bigger or equal than ε in the Cantorian foliation of primary KAM tori. These gaps are bigger than the size ε of the heteroclinic jumps provided by the scattering map (20). This is what is known in the literature as the *large gap problem.* Inside these regions, apart from primary KAM tori which are bent, there appear other invariant objects, which were not present in the unperturbed case, like secondary KAM tori and lower-dimensional tori, which are not detected by a direct application of KAM Theorem, but require a more careful analysis based on an averaging procedure.

In order to prove Theorem 3.1 we will restrict Hamiltonian (1) to the NHIM $\widetilde{\Lambda}_\varepsilon$ and perform an averaging procedure before applying a quantitative version of KAM Theorem. The fundamental difference with respect to [DLS06a] is that for every fixed ε it will be necessary to truncate adequately the perturbation in order to deal with a finite number of harmonics depending on ε. The phase space of the truncated Hamiltonian possesses an heterogeneous sea of a finite number of big gaps of different sizes, depending on the size of the harmonics of the perturbation.

Restriction to the NHIM $\widetilde{\Lambda}_\varepsilon$

Following the same arguments given in Sections 8.1 and 8.2 in [DLS06a], we have that the flow restricted to $\widetilde{\Lambda}_\varepsilon$ is Hamiltonian. More precisely, by Proposition 8.2 of [DLS06a], we can construct a $\mathcal{C}^r$ system of coordinates (J, φ, s) on $\widetilde{\Lambda}_\varepsilon$, where

$$J = \mathcal{J}(I, \varphi, s; \varepsilon) = I + \mathcal{O}_{\mathcal{C}^{r-1}}(\varepsilon), \quad (22)$$

such that the symplectic form on any $\Lambda_\varepsilon^s = \{(J', \varphi', s') \in \widetilde{\Lambda}_\varepsilon : s' = s\}$ has the standard expression $\omega_{|\Lambda_\varepsilon^s} = dJ \wedge d\varphi$. Since $\widetilde{\Lambda}_\varepsilon = \widetilde{\Lambda}$ for $\varepsilon = 0$ according to equation (16), by Proposition 8.4 in [DLS06a], the restriction of the Hamiltonian H_ε in (1) to $\widetilde{\Lambda}_\varepsilon$ expressed in these action angle coordinates (J, φ, s) has the form

$$k(J, \varphi, s; \varepsilon) = Z(J) + \varepsilon R(J, \varphi, s; \varepsilon) \tag{23}$$

with

$$Z(J) = J^2/2 \quad \text{and} \quad R(J, \varphi, s; 0) = h(0, 0, J, \varphi, s; 0), \tag{24}$$

where h is the perturbation in H_ε given in (6) and R is $\mathcal{O}_{\mathcal{C}^r}(1)$.

Remark 2.9. Notice that, by expression (24), $R_{k,l}(J; 0) = h_{k,l}(0, 0, J; 0)$, where $h_{k,l}$ and $R_{k,l}$ are the Fourier coefficients in the angle variables (φ, s) of the perturbation h and its restriction R to $\widetilde{\Lambda}_\varepsilon$, respectively.

Averaging procedure

We start performing an averaging procedure to the restricted Hamiltonian (23), as it was done in [DLS06a], which follows the argument used in the proof of KAM theorem in [Arn63], but paying attention to resonant regions. In [DLS06a] the perturbation was assumed to be a trigonometric polynomial, so there was only a finite number of resonances. However, in Hamiltonian (1) the perturbation h has an infinite number of harmonics, in the same way as R in equation (23), which give rise to an infinite number of resonances, so the results in [DLS06a] do not apply directly.

The main result for the implementation of an averaging procedure for a generic perturbation will be Theorem 3.11 in Section 3.2. This theorem makes precise the hypotheses required to truncate the Fourier series of the perturbation R in (23) with respect to the angle variables and develop a global averaging procedure that casts the Hamiltonian (23) into a global normal form that has different expressions in the non-resonant and resonant regions. The main property of the normal form is that it is almost ready to apply on it a quantitative version of KAM Theorem.

The precise statement and rigorous proof of Theorem 3.11 are postponed to Section 3.2. In the following we only describe its main features and the results needed to apply KAM Theorem.

There are three parameters that play an important role in the averaging procedure of Theorem 3.11. One is the number of steps of averaging m to be performed, which imposes a restriction on the differentiability r of the perturbation: $r > 2(m+1)^2$. This number of averaging steps is chosen later in the application of KAM Theorem. The other two are M, which is the order of truncation of the Fourier series and L, which determines the size of the resonant regions. Both of them are chosen to depend on ε in the following way: $M \sim \varepsilon^{-\rho}$ and $L \sim \varepsilon^{\alpha}$ where $\rho, \alpha > 0$ are going to be chosen conveniently during this averaging procedure.

For every fixed ε, we truncate the Fourier series of the perturbation R in equation (23) with respect to the angle variables (φ, s) up to order M in the following way

$$R = R^{[\leq M]} + R^{[>M]},$$

where

$$R^{[\leq M]}(J,\varphi,s;\varepsilon)=\sum_{\substack{(k,l)\in\mathbb{Z}^2,\\ |k|+|l|\leq M}} R_{k,l}(J;\varepsilon)e^{\mathrm{i}(k\varphi+ls)}, \tag{25}$$

and

$$R^{[>M]}(J,\varphi,s;\varepsilon)=\sum_{\substack{(k,l)\in\mathbb{Z}^2,\\ |k|+|l|>M}} R_{k,l}(J;\varepsilon)e^{\mathrm{i}(k\varphi+ls)}, \tag{26}$$

and we deal only with $R^{[\leq M]}$, which is the trigonometric polynomial of degree M, as a perturbation. The error introduced in Hamiltonian (23) coming from the neglected tail of the Fourier series will have to be estimated later on.

Since the truncated Hamiltonian $R^{[\leq M]}$ has a finite number of harmonics, an averaging procedure of m steps has to take into account a finite number of resonances, which are the set of rational numbers $J=-l/k$ with $|l|+|k|\leq mM$ (see Definitions 3.6 and 3.4).

This averaging procedure divides the phase space (J,φ,s) in two types of domains. On the one hand, the **non-resonant regions up to order** m $\mathcal{D}^m_{\mathrm{nr}}$, which are the set of points (J,φ,s) such that its action variable J is at a distance greater than $2L_k$ of any resonance $J=-l/k$, where $L_k=L/|k|$. On the other hand, the **resonant regions up to order** m $\mathcal{D}^m_{\mathrm{r}}$, which are the set of points (J,φ,s) such that its action variable J is at a distance smaller than L_k of any resonance $J=-l/k$ (see Definitions 3.7 and 3.9).

To avoid overlapping between all the resonant domains, the distance between a resonance $-l_0/k_0$ and any other $-l/k$ must be greater than $2(L_{k_0}+L_k)$. Since the resonances considered satisfy $|k|\leq mM$ we need to impose $4L<1/mM$, which requires $\rho\leq\alpha$ in terms of exponents of ε and this corresponds to the left hand side inequality of hypothesis (55) in Theorem 3.11.

Along the averaging procedure, one needs to control the $\mathcal{C}^\ell$ norms of the averaged terms and the remainders, for $0\leq\ell\leq n$ and $2m<n<r$, where n is the regularity which will be needed for the KAM Theorem and r is the regularity of the perturbation R in Hamiltonian (23). It turns out that the estimates for the $\mathcal{C}^\ell$ norm blow up as a negative power of $L\sim\varepsilon^\alpha$. Since the averaged terms and the remainder contain a power of ε in front of them, bounds for them can be kept small provided that α is small enough, that is for $\alpha<1/n$. This corresponds to the right hand side inequality of hypothesis (55) in Theorem 3.11 and also implies $\rho<1/n$, which is formula (51) in the hypotheses of Theorem 3.11.

In all this averaging procedure, there was an initial error coming from the neglected tail of the truncation of order M of the perturbation R in Hamiltonian (23), whose $\mathcal{C}^\ell$ norm can be bounded by $\varepsilon/M^{r-\ell-2}$, where r is the regularity of the perturbation R. To keep it smaller than the $\mathcal{C}^\ell$ norm of the remainder after m steps of averaging, one has to impose a lower bound on ρ, which implies $r\geq(1/\rho-2)m+2$ in order to make compatible lower and upper bounds for ρ, and this is hypothesis (52) in Theorem 3.11.

These conditions on m, ρ, α and r are stated in the hypotheses of Theorem 3.11. In it, it is proved that one can develop a global averaging procedure that casts the Hamiltonian (23) into a global normal form (56), that has different expressions in the non-resonant and resonant regions (these correspond to expressions (57) and (58) in theses of Theorem 3.11). In the non-resonant regions one can perform non resonant averaging transformations in such a way that the averaged Hamiltonian is very close to a rotor. On the other hand, near the resonances, the resonant averaging transformations cast the system to a one d.o.f. Hamiltonian, which is close to an integrable pendulum, provided that the perturbation satisfies some non-degeneracy conditions like **H3'**.

Summing up, we end up with a Hamiltonian (56) that consists of an integrable part $\bar{Z}^m$ (the averaged Hamiltonian) plus a perturbation $\varepsilon^{m+1}\bar{R}^m$ which is $\mathcal{O}_{\mathcal{C}^\ell}(\varepsilon^{m+1-\alpha(\ell+2m)})$, for $\ell = 0, \dots, n-2m$, where m is the number of steps of averaging performed. Recall that the integrable Hamiltonian $\bar{Z}^m$ has different expressions in resonant regions and non-resonant regions.

The integrable part of Hamiltonian (56) gives us an approximate equation $\bar{Z}^m = \text{cte}$ for the invariant tori. The next step is to show which tori survive and what is the distance between them when we add the perturbation term $\varepsilon^{m+1}\bar{R}^m$ in equation (56).

Quantitative version of KAM Theorem

The main tool for this section will be KAM Theorem 3.22, which is a result about the existence of invariant tori of a periodic perturbation of a Hamiltonian expressed in action-angle variables. It is a direct adaptation of Theorem 8.12 in [DLS06a]. We will use Theorem 3.22 to show that there exists a discrete foliation of invariant tori which are $\mathcal{O}(\varepsilon^{1+\eta})$-closely spaced, for some $\eta > 0$, and give approximate explicit expressions for them.

Since the integrable Hamiltonian (56) after m steps of averaging has different expressions in resonant and non-resonant regions (up to order m) introduced along the averaging procedure, we perform this study separately. In the end, we will show that all these regions can be grouped in two according to the expressions for the invariant tori obtained in each one, which are the big gaps region (21) and its complementary the flat tori region, already mentioned at the beginning of this subsection. Notice that the big gaps region (21) is formed by the resonances $J = -l/k$ of order 1, such that $|(k, l)| \leq M_{\mathrm{BG}}$, whereas flat tori region is composed by the non resonant regions up to order m and the resonant regions up to order m such that $J = -l/k$ and $|(k, l)| > M_{\mathrm{BG}}$, where M_{BG} is explicitly chosen in hypotheses **H3** as $M_{\mathrm{BG}} = \varepsilon^{-(1+\nu)/r}$, for any $1/(r/6 - 1) < \nu \leq 1/16$.

Non-resonant regions are studied in Section 3.3.2. In Proposition 3.24, we apply Theorem 3.22 directly to Hamiltonian (56)-(57), which is already written in action-angle variables, and we conclude that for these regions there exist flat primary KAM tori given in (79) as the level sets of a flat function $F = I + \mathcal{O}(\varepsilon^{1+\eta})$, which are $\mathcal{O}(\varepsilon^{1+\eta})$-closely spaced, for some $\eta > 0$, provided that $m \geq 2$ and $n \geq 2m + 6$.

Resonant regions are studied in Section 3.3.3. As we already said, for these regions

Hamiltonian (56)-(58) is not written in action angle variables but it is close to an integrable pendulum (58) provided that hypotheses **H3'** are satisfied. The integrable pendulum has rotational and librational orbits as well as separatrices, which separate these two types of motion. Rotational orbits have the same topology as the primary tori in the integrable Hamiltonian $Z(J) = J^2/2$ in Hamiltonian (23) and librational orbits are contractible to a periodic orbit, so they correspond to motions with topologies that were not present in the unperturbed Hamiltonian $Z(J)$ and they are called secondary tori. Librational orbits cover all the region inside the separatrix loop of Hamiltonian (58), giving rise to a gap between primary tori, and the size of this gap depends on the order of the corresponding resonance and the size of the Fourier coefficient associated to it.

When gaps are of size smaller than ε, which is the size of the heteroclinic jumps provided by the scattering map (17), they are called small gaps. In section 3.3.4, we study the resonant regions with small gaps $\mathcal{D}_{\mathrm{SG}}$ and in Proposition 3.26 we show that we can apply the same argument as in the case of non resonant regions to conclude that for these regions there exist flat primary KAM tori given in (87) as the level sets of a flat function $F = I + \mathcal{O}(\varepsilon^{1+\eta})$, which is the same as in the non-resonant case, and which are $\mathcal{O}(\varepsilon^{1+\eta})$-closely spaced, for some $\eta > 0$, provided that $m \geq 2$ and $n \geq 2m + 6$.

Notice that tori in the non resonant regions and resonant regions with small gaps are given by the level sets of the same function $F = I + \mathcal{O}(\varepsilon^{1+\eta})$ and they are flat up to $\mathcal{O}(\varepsilon^{1+\eta})$, for some $\eta > 0$. Both regions form the flat tori region.

Resonant regions with big gaps $\mathcal{D}_{\mathrm{BG}}$ are studied in Section 3.3.5. They correspond to resonances $J = -l/k$ such that $|(k,l)| < M_{\mathrm{BG}}$, where $M_{\mathrm{BG}} = \varepsilon^{-(1+\nu)/r}$, for $1/(r/6-1) < \nu \leq 1/16$. The size of the gap for these resonances is $C\varepsilon^{1/2}|(k,l)|^{-r/2}$, where C is a constant independent of ε and (k,l). Note that there is no uniform size of the gaps since it runs from order $\varepsilon^{1/2}$ for resonances with low $|(k,l)|$ to $\varepsilon^{1+\nu/2}$ for resonances with $|(k,l)| \sim M_{\mathrm{BG}}$.

Our criterium for the choice of the big gaps has been motivated by the size of the heteroclinic jumps provided by the scattering map (20): small gaps are of size smaller than ε, so they can always be traversed just connecting two primary tori by the scattering map, whereas this is not the case for big gaps. For these big gaps, we will show that we can find other invariant objects, like secondary tori, which fill the region inside the gaps and they get rather close to the frontier of the gaps among the primary KAM tori.

Remark 2.10. We would like to remark that our result about resonances that create big gaps is remarkably different of the one obtained in [DLS06a], where it was considered the case of a perturbation h with a finite number of harmonics. In that case there was a uniform size for the gaps created by the resonances of order 1 which was $C\varepsilon^{1/2}$. Moreover, for resonances of order 2 the uniform size of the associated gap was $C\varepsilon$. Hence, both resonances of order 1 and 2 were considered as big gaps.

In the case of resonances with big gaps, we will need to write the integrable pendulum $\bar{Z}^m$ given in (58) into action-angle variables before applying KAM Theorem 3.22. Since this change of coordinates becomes singular on the separatrix of the pendulum, we will need to define different action-angle variables inside and outside the separatrix, and we will exclude a thin neighborhood of the separatrix.

Moreover, since the behavior of the tori outside is different depending on their distance to the separatrix (tori are flatter as they are further from the separatrix) we consider different regions in the outside part of the separatrix, where we perform different scalings. This strategy, which was already introduced in [DLS06a], has been improved introducing a new sequence of domains in Theorem 3.30, which reduce the differentiability requirements.

The main result for the implementation of the above strategy for resonances with big gaps is Theorem 3.30 jointly with Corollary 3.31 which make explicit the relationship between the minimum distance between the surviving tori and the number m of steps of averaging performed.

In Theorem 3.28 we use both Theorem 3.30 and Corollary 3.31 to show that many of the invariant tori (both primary and secondary) of the integrable averaged Hamiltonian persist under the perturbation forming a sequence of tori given in (94) as the level sets of a function F, close to the averaged Hamiltonian with a distance between consecutive tori of order $\varepsilon^{1+\eta}$, for some $\eta > 0$, in terms of the action variable, provided that $m \geq 10$ and $n \geq 2m + 6$.

Propositions 3.24 and 3.26 and Theorem 3.28 can be joined in a unique result about the existence of nearby invariant tori for the inner dynamics, which is Theorem 3.1. This Theorem also gives explicit approximate expressions for the invariant tori, which are of two types depending on the region of the phase of space where invariant tori lie: the big gaps region and the flat tori region.

We refer to Sections 3.2 and 3.3 for the referenced theorems where one can find the complete proof.

2.3.4. Fourth Part: Construction of a transition chain and obstruction property In order to finish the proof, it remains to check that the finite sequence of invariant tori provided by Theorem 3.1 form a transition chain along the NHIM $\widetilde{\Lambda}_\varepsilon$, traversing both big gaps and flat tori regions, and to show that there are orbits that follow it closely. These are the orbits claimed in Theorem 2.1.

The scattering map S_ε associated to the homoclinic channel Γ_ε, defined in (17), is the main tool to detect that there exist transverse heteroclinic connections between these tori, which are objects of different topology. Indeed, by Lemma 10.4 in [DLS06a], we know that two submanifolds, like the invariant tori $\mathcal{T}_i$, of a NHIM $\widetilde{\Lambda}_\varepsilon$, have a transverse heteroclinic intersection if they are transversal under the scattering map as submanifolds of $\widetilde{\Lambda}_\varepsilon$.

The main result of this section is Proposition 4.1 where it is proved the existence of transition chains, that is chains of invariant tori $\mathcal{T}_i$, both primary and secondary, such that their image under the scattering map S_ε in (20) intersects transversally $\mathcal{T}_{i+1}$ on $\widetilde{\Lambda}_\varepsilon$, that is

$$S_\varepsilon(\mathcal{T}_i) \pitchfork_{\widetilde{\Lambda}_\varepsilon} \mathcal{T}_{i+1}. \tag{27}$$

In Section 2.3.2 we have obtained an explicit expression (20) up to first order for the scattering map S_ε using the first order calculation of the Hamiltonian function $\mathcal{S}_\varepsilon$. In Section 2.3.3 we have shown that on the NHIM $\widetilde{\Lambda}_\varepsilon$ there exists a discrete foliation of KAM tori $\mathcal{T}_i$ (primary and secondary) which are $\mathcal{O}(\varepsilon^{1+\eta})$-closely spaced, for some $\eta > 0$. Moreover, we have obtained explicit expressions for tori $\mathcal{T}_i$, both primary and secondary, and we have seen that these invariant objects are given approximately by the level sets of the averaged Hamiltonian.

In Lemma 4.2 in Section 4.1, we give an expression for the action of the scattering map S_ε on a foliation given by the level sets of a given function F, using the expression for the Hamiltonian function $\mathcal{S}_\varepsilon$ generating the deformation of the scattering map, introduced in Section 2.3.2. Moreover, we give conditions to assure transversality between the foliation in $\widetilde{\Lambda}_\varepsilon$ and its image under the scattering map S_ε.

As we have seen in the previous section, the different types of tori that appear in our problem have different quantitative properties and therefore the dominant terms in the expression of these invariant objects as the level sets of a certain function are different whether they lie in a flat tori region or a big gaps region. Lemma 4.2 is applied in Lemma 4.5 for the case of the flat tori region, and in Lemma 4.7 for the case of the big gaps region. In these Lemmas it is shown that the sufficient conditions on the perturbation of the Hamiltonian (1) for the transversality are hypotheses **H2"**, **H3"** and **H3"'** in Theorem 2.1.

Putting all these results together in Proposition 4.1, we have that, by hypothesis **H2"** and the non-degeneracy conditions **H3"** and **H3"'**, the scattering map S_ε maps pieces of these tori transversally in Λ_ε to other tori at a distance $\mathcal{O}(\varepsilon)$, that is $S_\varepsilon(\mathcal{T}_i) \pitchfork \mathcal{T}_{i+1}$, where $\mathcal{T}_i$ and $\mathcal{T}_{i+1}$ are invariant tori at a distance smaller than ε. Therefore, by Lemma 10.4 in [DLS06a] we have that $W^u_{\mathcal{T}_i} \pitchfork W^s_{\mathcal{T}_{i+1}}$ and we have constructed a transition chain.

Finally, we use the well known result that given a transition chain $\{\mathcal{T}_i\}_{i=0}^N$, we can find an orbit visiting all the elements of the chain. In our case, as it was the case in [DLS06a] we have incorporated in the chain objects with different topologies, so applying Lemma 11.1 in [DLS06a] to the transition chain obtained, we get that there is $\varepsilon^* > 0$ such that for $0 < |\varepsilon| < \varepsilon^*$, and for any interval $[I^*_-, I^*_+] \in (I_-, I_+)$, $\widetilde{x}(t)$ satisfies that, for some $T > 0$

$$I(\widetilde{x}(0)) \leq I^*_-; \qquad I(\widetilde{x}(T)) \geq I^*_+$$

(respectively:

$$I(\widetilde{x}(0)) \geq I^*_+; \qquad I(\widetilde{x}(T)) \leq I^*_-)$$

as we wanted to prove.

3. Inner Dynamics

The main goal of this section is to prove Theorem 3.1 about the existence of a sequence of invariant tori $\mathcal{T}_i$ in the NHIM $\widetilde{\Lambda}_\varepsilon$, which are distributed along all the actions in the interval (I_-, I_+) and are $\mathcal{O}(\varepsilon^{1+\eta})$-closely spaced, for some $\eta > 0$. The method of proof will consist of the combination of two parts: averaging methods and KAM Theorem.

In Section 3.2 we will consider the restricted Hamiltonian (23) and perform, in Theorem 3.11, a global averaging procedure that casts the Hamiltonian into a global normal form, which has different expressions in the non-resonant and resonant regions. In the non-resonant regions, averaging transformations cast the system to close to a rotor and, in general, in the non-resonant regions to close to an integrable pendulum.

In Section 3.3 we will use KAM Theorem 3.22 to show that many of the invariant tori of the averaged Hamiltonian persist when we add the error terms of the normal form and they are close enough in terms of the action variables. For the flat tori region, which consists of non-resonant regions and resonant regions with small gaps, we can apply KAM Theorem 3.22 almost straightforwardly and this is done in Propositions 3.24 and 3.26, respectively. For the big gaps region, we will show in Theorem 3.28 that we can apply KAM Theorem after we have written the Hamiltonian in action-angle coordinates.

3.1. Main result

The main result about the existence of invariant tori in the NHIM $\widetilde{\Lambda}_\varepsilon$ is stated in the following Theorem:

Theorem 3.1. *Consider a Hamiltonian of the form (1) and assume that* $r > 2(m+1)^2$, *with* $m \geq 10$ *and* $n = 2m+6$, *as well as hypothesis* **H3'**. *Choose* $\eta = \min((m-1-\alpha n)/2, \nu/2 - 3(1+\nu)/r)$, *where* $\alpha < 1/n$ *and* $1/(r/6-1) < \nu \leq 1/16$. *Then, for* ε *small enough, there exists a finite sequence of invariant tori* $\{\mathcal{T}_i\}_{i=0}^N$ *in* $\widetilde{\Lambda}_\varepsilon$ *which are distributed along all the actions in the interval* (I_-, I_+), *such that*

1. *They are defined by the equation* $F(I, \varphi, s; \varepsilon) \equiv E_i$, *where* F *is a* $\mathcal{C}^{4-\varrho}$ *function, for any* $\varrho > 0$, *which has the form (87) and (94) depending on the region where the invariant tori lie: the flat tori region or a connected component of the big gaps region defined in (82), respectively. In the flat tori region, the invariant tori are primary whereas in the big gaps region invariant tori can be primary or secondary. In the big gaps region, for values of* $E_i > 0$ *equation (94) provides two primary tori* $\mathcal{T}_{E_i}^{\pm}$, *whereas for* $E_i < 0$ *it gives a secondary tori* $\mathcal{T}_{E_i}$.
2. *They can be also written as a graph of the variable* I *over the angle variables* (φ, s): $I = \lambda_E(\varphi, s; \varepsilon)$ *with* λ_E *given in (88) for the flat tori region. In the case of the big gaps region, the equations for them are given for two different invariant tori* $\mathcal{T}_i^{\pm}$ *(two different components in the case of secondary KAM tori) in the form* $I = \lambda_E^{\pm}(\varphi, s; \varepsilon)$, *with* $\lambda_E^{\pm}$ *given in (95).*
3. *These tori are* $\mathcal{O}(\varepsilon^{1+\eta})$*-closely spaced in terms of the action variable* I. *In the connected component (82) of the big gaps region associated to the resonance* $-l_0/k_0$, *they are* $\mathcal{O}(\varepsilon^{3/2+\eta}|(k_0, l_0)|^{-r/2+1})$*-closely spaced in terms of energies* E_i, *where* $-l_0/k_0$ *is the associated resonance.*

4. $\mathcal{T}_0$ and $\mathcal{T}_N$ are $\mathcal{O}_{\mathcal{C}^2}(\varepsilon^{1+\eta})$-close to I_- and I_+, respectively.

The proof of Theorem 3.1 is a combination of an averaging procedure (Section 3.2) and a KAM Theorem (Section 3.3). In Section 3.4 we put the results obtained in the previous sections together to give a proof of Theorem 3.1.

3.2. Averaging procedure

In this section we proceed to obtain a suitable global normal form of the restricted Hamiltonian (23), according to the procedure described in Section 2.3.3. We use the standard formalism of Lie Series, so we are considering canonical transformations obtained as the time-one map of the flow of a Hamiltonian. A very pedagogical treatment of this method can be found in [LM88]. As we have already mentioned, we consider a truncation of the Fourier Series of the perturbation and we deal with trigonometric polynomials of a finite order. We first introduce a Banach space with a suitable norm, which allows an efficient study of the estimates for the different terms that appear in the averaging procedure.

3.2.1. Preliminaries. Functional Spaces We consider the space of functions defined on $\mathcal{I} \times \mathbb{T}^2$, $\mathcal{I} \subset \mathbb{R}$ compact set, which consists of trigonometric polynomials of order M on $(\varphi, s) \in \mathbb{T}^2$, and $\mathcal{C}^r$ with respect to $J \in \mathcal{I} \subset \mathbb{R}$. We denote this space $\mathcal{T}_M(\mathcal{I} \times \mathbb{T}^2)$. A function $u \in \mathcal{T}_M(\mathcal{I} \times \mathbb{T}^2)$ is of the form

$$u(J, \varphi, s) = \sum_{\substack{(k,l)\in\mathbb{Z}^2,\\ |k|+|l|\leq M}} u_{k,l}(J) e^{\mathrm{i}(k\varphi+ls)}. \tag{28}$$

Remark 3.2. Note that the product of two elements $u \in \mathcal{T}_M(\mathcal{I}\times\mathbb{T}^2)$ and $v \in \mathcal{T}_N(\mathcal{I}\times\mathbb{T}^2)$ is another trigonometric polynomial in the variables $(\varphi, s) \in \mathbb{T}^2$ but of degree $M + N$, that is, $uv \in \mathcal{T}_{M+N}(\mathcal{I} \times \mathbb{T}^2)$.

Clearly, the space $\mathcal{T}_M(\mathcal{I}\times\mathbb{T}^2)$ is a closed subset of $\mathcal{C}^r(\mathcal{I}\times\mathbb{T}^2)$. Therefore, $\mathcal{T}_M(\mathcal{I}\times\mathbb{T}^2)$ is a Banach space with the $\mathcal{C}^r$ norm introduced in (2).

Moreover, since the functions u are trigonometric polynomials in (φ, s), we can consider the expression (28) and deal with the Fourier norm:

$$\|u\|^{[\leq M]}_{\mathcal{C}^\ell(\mathcal{I}\times\mathbb{T}^2)} := \sum_{m=0}^{\ell}\sum_{n=0}^{m} 2^\ell \sum_{\substack{(k,l)\in\mathbb{Z}^2,\\ |k|+|l|\leq M}} |u_{k,l}|_{\mathcal{C}^n(\mathcal{I})}|(k,l)|^{m-n} \tag{29}$$

where $|u_{k,l}|_{\mathcal{C}^n(\mathcal{I})}$ is defined in (2) and $|(k, l)| = \max(|k|, |l|)$, and $|\cdot|$ denotes the standard modulo. When there is no possibility of confusion about M we will abreviate it as $\|\cdot\|_{\mathcal{C}^\ell}$.

On the other hand, to understand better the behavior of the function u with respect to the variable J when it gets closer to the resonances, we will use the Fourier norm with a weight $L \leq 1$:

$$\|u\|^{[\leq M]}_{\mathcal{C}^\ell(\mathcal{I}\times\mathbb{T}^2),L} := \sum_{m=0}^{\ell}\sum_{n=0}^{m} 2^\ell \sum_{\substack{(k,l)\in\mathbb{Z}^2,\\ |k|+|l|\leq M}} |u_{k,l}|_{\mathcal{C}^n(\mathcal{I}),L}|(k,l)|^{m-n} \tag{30}$$

where $|(k,l)|$ is as before and

$$|u_{k,l}|_{\mathcal{C}^n(\mathcal{I}),L} := \sum_{i=0}^{n} L^i \frac{|D^i u_{k,l}|_{\mathcal{C}^0(\mathcal{I})}}{i!}.$$

As before, when there is no confusion about M we will abbreviate these norms as $\|\cdot\|_{\mathcal{C}^\ell,L}$ and $|\cdot|_{\mathcal{C}^n,L}$, respectively.

Note that when $L=1$, we recover the Fourier norm (29).

The basic properties of these norms are collected in Appendix B. In particular they are related by

$$L^\ell |u|_{\mathcal{C}^\ell} \leq \|u\|_{\mathcal{C}^\ell,L} \leq CM^2 |u|_{\mathcal{C}^\ell}, \tag{31}$$

where C is a constant that depends on ℓ but it is independent of M and $0 < L \leq 1$.

For the seminorm $|\cdot|_{j,\ell-j}$ defined in (3) one has that for all $0 \leq j \leq \ell$,

$$L^j \, |u|_{j,\ell-j} \leq \|u\|_{\mathcal{C}^\ell,L}\,. \tag{32}$$

Note that in the case that the function $u \in \mathcal{T}_M(\mathcal{I}\times\mathbb{T}^2)$ does not depend on the action variable J, we have that

$$|u|_{\mathcal{C}^\ell} = |u|_{0,\ell},$$

therefore by equation (32),

$$|u|_{\mathcal{C}^\ell} \leq \|u\|_{\mathcal{C}^\ell,L}\,. \tag{33}$$

Moreover, given $u \in \mathcal{T}_M(\mathcal{I}\times\mathbb{T}^2)$ and $v \in \mathcal{T}_N(\mathcal{I}\times\mathbb{T}^2)$, we have that $uv \in \mathcal{T}_{M+N}(\mathcal{I}\times\mathbb{T}^2)$ and for $0 < L \leq 1$ and $0 \leq \ell \leq r$,

$$\|uv\|^{[\leq M+N]}_{\mathcal{C}^\ell,L} \leq \|u\|^{[\leq M]}_{\mathcal{C}^\ell,L} \, \|v\|^{[\leq N]}_{\mathcal{C}^\ell,L}\,. \tag{34}$$

We will say that a function f is $\mathcal{O}_{\mathcal{C}^r,L}(\eta)$ when $\|f\|_{\mathcal{C}^r,L} \preceq \eta$.

3.2.2. The homological equation In this section, we will use the standard formalism of Lie series to perform a resonant averaging procedure. We first start discussing the infinitesimal equations for averaging, which will serve as a motivation for the phenomenon of resonances and therefore for the resonant averaging.

We begin with a Hamiltonian $K(J,A,\varphi,s) = K_0(J,A) + \varepsilon K_1(J,A,\varphi,s)$, where $(J,A,\varphi,s) \in \mathbb{R}^2\times\mathbb{T}^2$ and $K_0(J,A) = A + J^2/2$. We start looking for a canonical transformation g, given by the time-one map of the flow of a Hamiltonian $\varepsilon G(J,A,\varphi,s)$ (generating function), that eliminates, when it is possible, the dependence on the angle variables (φ,s) up to order ε. Therefore,

$$\begin{aligned} K\circ g &= K + \{K,\varepsilon G\} + \frac{1}{2}\{\{K,\varepsilon G\},\varepsilon G\} + \dots \\ &= K_0 + \varepsilon(K_1 + \{K_0,G\}) + \mathcal{O}(\varepsilon^2) \end{aligned}$$

where $\{,\}$ denotes the Poisson bracket in the canonical coordinates (J,A,φ,s):

$$\{f,g\} = \frac{\partial f}{\partial\varphi}\frac{\partial g}{\partial J} + \frac{\partial f}{\partial s}\frac{\partial g}{\partial A} - \frac{\partial f}{\partial J}\frac{\partial g}{\partial\varphi} - \frac{\partial f}{\partial A}\frac{\partial g}{\partial s}.$$

We seek for a solution G of the infinitesimal equation

$$K_1 + \{K_0, G\} = \bar{K},$$

which produces a $\bar{K}$ as simple as possible. In Fourier coefficients this equation has the form

$$K_{k,l}(J) - \mathrm{i}(\omega(J) \cdot (k,l))G_{k,l}(J) = \bar{K}_{k,l}(J) \tag{35}$$

where $K_{k,l}(J)$, $G_{k,l}(J)$ and $\bar{K}_{k,l}(J)$ are the Fourier coefficients of K_1, G and $\bar{K}$, respectively, for $(k,l) \in \mathbb{Z}^2$, and $\omega(J) \in \mathbb{R}^2$ is of the form

$$\omega(J) = \left(\frac{\partial K_0}{\partial J}, \frac{\partial K_0}{\partial A}\right) = (J, 1).$$

This vector $\omega(J)$ is called *resonant* when $(J,1) \cdot (k,l) = Jk + l = 0$, for $(k,l) \neq (0,0)$; and the values $J = -l/k$, with $k \neq 0$, for which this equation vanishes and $K_{k,l}(-l/k) \neq 0$ are called *resonances*. Looking at equation (35) it is clear that these are the places where we can not choose $G_{k,l}(J)$ in order to have $\bar{K}_{k,l}(J) \equiv 0$. So, for these values of J and, in order to keep smoothness, the ones in a neighborhood around them, we will choose $\bar{K}_{k,l}(J)$ to be the Fourier term $K_{k,l}(-l/k)$. Note that we cannot have $\bar{K}_{0,0}(J) \equiv 0$ for any J either, so we will also keep the Fourier coefficient $K_{0,0}(J)$.

The precise result with the estimates for the functions is formulated in the following Lemma:

Lemma 3.3. *Let $K(J,\varphi,s)$ be a Hamiltonian defined on $\mathcal{I} \times \mathbb{T}^2$, $\mathcal{I} \subset \mathbb{R}$ compact set, which is a $\mathcal{C}^{r+1}$ function with respect to J and a trigonometric polynomial in (φ,s) of degree M, so it can be expressed in the following way*

$$K(J,\varphi,s) = \sum_{(k,l)\in\mathcal{N}} K_{k,l}(J) e^{\mathrm{i}(k\varphi+ls)},$$

with $\mathcal{N} = \{(k,l) \in \mathbb{Z}^2, |k| + |l| \leq M\}$. We refer to resonances as the elements of the finite set of rational numbers

$$\mathcal{R} = \{-l/k \in \mathbb{Q} : (k,l) \in \mathcal{N}, k \neq 0, K_{k,l}(-l/k) \neq 0\}. \tag{36}$$

For any $(k,l) \in \mathcal{N}$, we consider $(\widetilde{k},\widetilde{l}) \in \mathbb{Z}^2$ such that $-l/k = -\widetilde{l}/\widetilde{k}$ and $\gcd(\widetilde{k},\widetilde{l}) = 1$ *and we define $L_k = L_{\widetilde{k}} = L/|\widetilde{k}|$, being $L \leq 1$ some constant small enough such that for all $-l/k \in \mathcal{R}$, the real intervals $[-l/k - 2L_k, -l/k + 2L_k]$ are all disjoint.*

Then, there exist a function $G = G^{[\leq M]}$ of class $\mathcal{C}^r$ with respect to J and $\bar{K} = \bar{K}^{[\leq M]}$ of class $\mathcal{C}^{r+1}$, which are both trigonometric polynomials in (φ,s), such that they solve the homological equation

$$K + \{K_0, G\} = \bar{K}, \tag{37}$$

and verify:

1. *If $|J + l/k| \geq 2L_k$ for any $(k,l) \in \mathcal{N}$, then*

$$\bar{K}(J,\varphi,s) = K_{0,0}(J). \tag{38}$$

2. *If* $|J + l_0/k_0| \leq L_{k_0}$ *for some* $(k_0, l_0) \in \mathcal{N}$*, then*

$$\begin{aligned} \bar{K}(J, \varphi, s) &= K_{0,0}(J) + \sum_{\substack{t \in \mathbb{Z}-\{0\} \\ |t|(|k_0|+|l_0|) \leq M}} K_{tk_0, tl_0}(-l_0/k_0) e^{\mathrm{i}t(k_0\varphi + l_0 s)} \\ &=: K_{0,0}(J) + U_{k_0,l_0}(k_0\varphi + l_0 s). \end{aligned} \tag{39}$$

3. *The function* $\bar{K}$ *verifies*

$$\left\|\bar{K}\right\|_{\mathcal{C}^{\ell},L} \leq C_{\ell} \left\|K\right\|_{\mathcal{C}^{\ell},L}, \tag{40}$$

for $\ell = 0, \ldots, r+1$*, where* C_{ℓ} *is a constant independent of* L, M*.*

4. *The function* G *verifies*

$$\left\|G\right\|_{\mathcal{C}^{\ell},L} \leq \frac{C_{\ell}}{L} \left\|K\right\|_{\mathcal{C}^{\ell+1},L} \tag{41}$$

for $\ell = 0, \ldots, r$*, where* C_{ℓ} *is a constant independent of* L, M*.*

Proof. We want to solve for each $(k, l) \in \mathcal{N}$ the equation (35)

$$K_{k,l}(J) - \mathrm{i}(Jk + l) G_{k,l}(J) = \bar{K}_{k,l}(J), \tag{42}$$

where the unknowns are the Fourier coefficients of the generating function G and the averaged Hamiltonian $\bar{K}$.

So, we first choose:

1. $\bar{K}_{0,0}(J) = K_{0,0}(J)$,
2. if $(0, l) \in \mathcal{N}$, $l \neq 0$, $\bar{K}_{0,l}(J) = 0$,
3. if $(k, l) \in \mathcal{N}$, $k \neq 0$, we choose $\bar{K}_{k,l}(J)$ as

$$\bar{K}_{k,l}(J) = K_{k,l}(-l/k)\psi\left(\frac{1}{L_k}(J + l/k)\right), \tag{43}$$

where $\psi(x)$ is a fixed $\mathcal{C}^{\infty}$ function such that: $\psi(x) = 1$, if $x \in [-1, 1]$, and $\psi(x) = 0$, if $x \notin [-2, 2]$. With this choice we have that $\bar{K}_{k,l}$ verifies:

(a) If $|J + l/k| \leq L_k$ then $\bar{K}_{k,l}(J) = K_{k,l}(-l/k)$,

(b) if $|J + l/k| \geq 2L_k$ then $\bar{K}_{k,l}(J) = 0$.

Once we have defined $\bar{K}$ as above, it is clear that it is a $\mathcal{C}^{r+1}$ function, and its Fourier coefficients satisfy:

$$\begin{aligned} \left|\bar{K}_{k,l}\right|_{\mathcal{C}^n,L} &= \sum_{i=0}^{n} L^i \frac{\left|D^i \bar{K}_{k,l}\right|_{\mathcal{C}^0}}{i!} \\ &= \sum_{i=0}^{n} \frac{L^i}{i!} \frac{\left|K_{k,l}(-l/k)\right| \left|D^i\psi\right|_{\mathcal{C}^0}}{L_k^i} \\ &\leq \left|K_{k,l}\right|_{\mathcal{C}^0} |k|^n \sum_{i=0}^{n} \frac{\left|D^i\psi\right|_{\mathcal{C}^0}}{i!} \\ &= \left|K_{k,l}\right|_{\mathcal{C}^0} |k|^n \left|\psi\right|_{\mathcal{C}^n}. \end{aligned} \tag{44}$$

Using this inequality for the Fourier coefficients it is easy to see that $\bar{K}$ verifies the desired bound (40). More precisely,

$$\begin{aligned}
\left\|\bar{K}\right\|_{\mathcal{C}^\ell,L} &= \sum_{m=0}^{\ell}\sum_{n=0}^{m} 2^\ell \sum_{\substack{(k,l)\in\mathbb{Z}^2,\\ |k|+|l|\le M}} \left|\bar{K}_{k,l}\right|_{\mathcal{C}^n,L} |(k,l)|^{m-n} \\
&\le \sum_{m=0}^{\ell}\sum_{n=0}^{m} 2^\ell \sum_{\substack{(k,l)\in\mathbb{Z}^2,\\ |k|+|l|\le M}} |\psi|_{\mathcal{C}^n} \left|K_{k,l}\right|_{\mathcal{C}^0} |k|^n |(k,l)|^{m-n} \\
&\le |\psi|_{\mathcal{C}^\ell} \sum_{m=0}^{\ell}\sum_{n=0}^{m} 2^\ell \sum_{\substack{(k,l)\in\mathbb{Z}^2,\\ |k|+|l|\le M}} \left|K_{k,l}\right|_{\mathcal{C}^0} |(k,l)|^{m} \\
&\le (\ell+1)\, |\psi|_{\mathcal{C}^\ell} \left\|K\right\|_{\mathcal{C}^\ell,L}
\end{aligned}$$

for $\ell = 0,\ldots,r+1$, so choosing $C_\ell = (\ell+1)\,|\psi|_{\mathcal{C}^\ell}$, which is independent of L, we get the desired bound.

Now, we choose G to verify equation (42) so we get:

1. $G_{0,0} = 0$ and $G_{k,l}(J) = 0$ if $(k,l) \notin \mathcal{N}$,
2. if $(0,l) \in \mathcal{N}$, $l \neq 0$, $G_{0,l}(J) = K_{0,l}(J)/\mathrm{i}l$,
3. if $(k,l) \in \mathcal{N}$, $k \neq 0$, we choose $G_{k,l}(J)$ as:
   (a) If $J \neq -l/k$ then $G_{k,l}(J) = \mathrm{i}\dfrac{\bar{K}_{k,l}(J) - K_{k,l}(J)}{Jk+l}$,
   (b) $G_{k,l}(-l/k) = \lim\limits_{J\to -l/k} \dfrac{K_{k,l}(J) - \bar{K}_{k,l}(J)}{\mathrm{i}(Jk+l)} = \dfrac{K'_{k,l}(-l/k)}{\mathrm{i}k}$.

Then $G(J,\varphi,s)$ is a trigonometric polynomial in (φ,s) of degree M, and of class $\mathcal{C}^r$ with respect to J. To bound the function G, we first need to bound its Fourier coefficients in terms of $|\cdot|_{\mathcal{C}^\ell,L}$ norm for $0 \le \ell \le r$. Given a fixed $(k_0,l_0) \in \mathcal{N}$, by the definition of $\bar{K}$ and G, we have:

1. $\forall J$, $|G_{0,l}|_{\mathcal{C}^n,L} \le |K_{0,l}|_{\mathcal{C}^n,L}/|l|$, for $\ell = 0,\ldots,r$.
2. If $|J + l_0/k_0| \le L_{k_0}$, then $|G_{k_0,l_0}|_{\mathcal{C}^n,L} \le (n+1)\dfrac{|K_{k_0,l_0}|_{\mathcal{C}^{n+1},L}}{L|k_0|}$, for $n = 0,\ldots,r$.

   This estimate comes from

$$\begin{aligned}
|G_{k_0,l_0}|_{\mathcal{C}^n,L} &= \sum_{i=0}^{n} L^i \frac{|D^i G_{k_0,l_0}|_{\mathcal{C}^0}}{i!} \\
&\le \sum_{i=0}^{n} \frac{L^i}{i!} \frac{|D^{i+1} K_{k_0,l_0}|_{\mathcal{C}^0}}{|k_0|} \\
&\le \frac{(n+1)}{L|k_0|} \sum_{i=0}^{n} \frac{L^{i+1}}{(i+1)!} \left|D^{i+1} K_{k_0,l_0}\right|_{\mathcal{C}^0} \\
&\le (n+1)\frac{|K_{k_0,l_0}|_{\mathcal{C}^{n+1},L}}{L|k_0|}.
\end{aligned}$$

3. If $|J+l_0/k_0| \geq 2L_{k_0}$ then $|G_{k_0,l_0}|_{\mathcal{C}^n} \leq \frac{n+1}{L}\sum_{i=0}^{\ell}|K_{k_0,l_0}|_{\mathcal{C}^i,L}|k_0|^{n-i}$, for $n = 0,\dots,r+1$.
   This estimate is obtained using Leibniz rule for derivatives in the following way:
$$\begin{aligned}
|G_{k_0,l_0}|_{\mathcal{C}^n,L} &= \sum_{i=0}^{n} L^i \frac{|D^iG_{k_0,l_0}|_{\mathcal{C}^0}}{i!}\\
&= \sum_{i=0}^{n}\frac{L^i}{i!}\left|D^i\left(-\mathrm{i}\frac{K_{k_0,l_0}}{Jk_0+l_0}\right)\right|_{\mathcal{C}^0}\\
&\leq \sum_{i=0}^{n}\frac{L^i}{i!}\sum_{j=0}^{i}\left(\begin{array}{c} i\\ j\end{array}\right)\frac{|D^jK_{k_0,l_0}|_{\mathcal{C}^0}}{(2L_{k_0})^{i-j+1}|k_0|}\\
&\leq \sum_{i=0}^{n}\frac{1}{L}\sum_{j=0}^{i}L^j\frac{|D^jK_{k_0,l_0}|_{\mathcal{C}^0}}{j!}|k_0|^{i-j}\\
&\leq \frac{n+1}{L}\sum_{i=0}^{n}L^i\frac{|D^iK_{k_0,l_0}|_{\mathcal{C}^0}}{i!}|k_0|^{n-i}\\
&\leq \frac{n+1}{L}\sum_{i=0}^{n}|K_{k_0,l_0}|_{\mathcal{C}^i,L}|k_0|^{n-i}.
\end{aligned}$$
4. If $L_{k_0} \leq |J+l_0/k_0| \leq 2L_{k_0}$, then
$$|G_{k_0,l_0}|_{\mathcal{C}^n,L} \leq \frac{n+1}{L}\sum_{i=0}^{n}|K_{k_0,l_0}|_{\mathcal{C}^i,L}|k_0|^{n-i} + (n+1)|K_{k,l}|_{\mathcal{C}^0}|k|^n|\psi|_{\mathcal{C}^n},$$
   for $n = 0,\dots,r$ and C is a constant independent of L.
   This estimate can be obtained in the same way as the previous one using the estimate obtained for $\bar{K}_{k,l}$ in (44), in the following way
$$\begin{aligned}
|G_{k_0,l_0}|_{\mathcal{C}^n,L} &= \sum_{i=0}^{n} L^i\frac{|D^iG_{k_0,l_0}|_{\mathcal{C}^0}}{i!}\\
&\leq \sum_{i=0}^{n}\frac{L^i}{i!}\left|D^i\left(\mathrm{i}\frac{\bar{K}_{k_0,l_0}}{Jk_0+l_0}\right)\right|_{\mathcal{C}^0} + \sum_{i=0}^{n}\frac{L^i}{i!}\left|D^i\left(-\mathrm{i}\frac{K_{k_0,l_0}}{Jk_0+l_0}\right)\right|_{\mathcal{C}^0}\\
&\leq \frac{(n+1)^2}{L}|K_{k,l}|_{\mathcal{C}^0}|k|^n|\psi|_{\mathcal{C}^n} + \frac{n+1}{L}\sum_{i=0}^{n}|K_{k_0,l_0}|_{\mathcal{C}^i,L}|k_0|^{n-i}.
\end{aligned}$$

In order to finish the proof, we will use these estimates for the Fourier coefficients of G to bound the function G.

First we concentrate on the set $\mathcal{I}' \subset \mathcal{I}$ formed by $J \in \mathbb{R}$, such that $|J+l_0/k_0| \leq L_{k_0}$ for some $-l_0/k_0 \in \mathcal{R}$. Notice that if $J \in \mathcal{I}'$, for any other $(k,l) \in \mathcal{N}$, such that $(k,l) \neq (tk_0, tl_0)$ for $t \in \mathbb{Z}$, J satisfies that $|J+l/k| \geq 2L_k$. Therefore, we will distinguish three types of Fourier coefficients $G_{k,l}$ of G, which are the ones described in points 1,2 and 3 in this proof. Using their corresponding bounds we have

$$
\begin{aligned}
\|G\|_{\mathcal{C}^\ell(\mathcal{I}'\times\mathbb{T}^2),L} &= \sum_{m=0}^{\ell}\sum_{n=0}^{m} 2^\ell \left(\sum_{l=-M}^{M} |G_{0,l}|_{\mathcal{C}^n,L}\, |l|^{m-n} + \sum_{\substack{(k,l)\in\mathcal{N}\\(k,l)\neq t(k_0,l_0)}} |G_{k,l}|_{\mathcal{C}^n,L}\, |(k,l)|^{m-n} \right.\\
&\quad \left. + \sum_{\substack{t\in\mathbb{Z}\setminus\{0\}\\|t|(|k_0|+|l_0|)\leq M}} |G_{tk_0,tl_0}|_{\mathcal{C}^n,L}\, |t(k_0,l_0)|^{m-n} \right)\\
&\leq \sum_{m=0}^{\ell}\sum_{n=0}^{m} 2^\ell \left(\sum_{l=-M}^{M} \frac{|K_{0,l}|_{\mathcal{C}^n,L}}{|l|} |l|^{m-n} \right.\\
&\quad + \sum_{\substack{(k,l)\in\mathcal{N}\\(k,l)\neq t(k_0,l_0)}} \left(\frac{(n+1)}{L} \sum_{i=0}^{n} |K_{k,l}|_{\mathcal{C}^i,L}\, |k|^{n-i} \right) |(k,l)|^{m-n}\\
&\quad \left. + \sum_{\substack{t\in\mathbb{Z}\setminus\{0\}\\|t|(|k_0|+|l_0|)\leq M}} \frac{(n+1)}{L|k_0|}\, |K_{tk_0,tl_0}|_{\mathcal{C}^{n+1},L}\, |t(k_0,l_0)|^{m-n} \right)\\
&\leq \sum_{m=0}^{\ell}\sum_{n=0}^{m} 2^\ell \left(\sum_{l=-M}^{M} |K_{0,l}|_{\mathcal{C}^n,L}\, |l|^{m-n} \right.\\
&\quad + \sum_{\substack{(k,l)\in\mathcal{N}\\(k,l)\neq t(k_0,l_0)}} \frac{(n+1)}{L} \sum_{i=0}^{n} |K_{k,l}|_{\mathcal{C}^i,L}\, |(k,l)|^{m-i}\\
&\quad \left. + \sum_{\substack{t\in\mathbb{Z}\setminus\{0\}\\|t|(|k_0|+|l_0|)\leq M}} \frac{(n+1)}{L}\, |K_{tk_0,tl_0}|_{\mathcal{C}^{n+1},L}\, |t(k_0,l_0)|^{m-n-1} \right)\\
&\leq \sum_{m=0}^{\ell}\sum_{n=0}^{m} 2^\ell \left(\sum_{l=-M}^{M} |K_{0,l}|_{\mathcal{C}^n,L}\, |l|^{m-n} \right.\\
&\quad + \frac{(n+1)(m+1)}{L} \sum_{\substack{(k,l)\in\mathcal{N}\\(k,l)\neq t(k_0,l_0)}} |K_{k,l}|_{\mathcal{C}^n,L}\, |(k,l)|^{m-n}\\
&\quad \left. + \frac{n+1}{L} \sum_{\substack{t\in\mathbb{Z}\setminus\{0\}\\|t|(|k_0|+|l_0|)\leq M}} |K_{tk_0,tl_0}|_{\mathcal{C}^{n+1},L}\, |t(k_0,l_0)|^{m-n-1} \right)\\
&\leq \|K\|_{\mathcal{C}^\ell,L} + \frac{(\ell+1)}{L}\Big((\ell+1)\, \|K\|_{\mathcal{C}^\ell,L} + \|K\|_{\mathcal{C}^{\ell+1},L} \Big)\\
&\leq \frac{C_\ell}{L}\, \|K\|_{\mathcal{C}^{\ell+1},L}
\end{aligned}
$$

for $\ell = 0, \ldots, r$, where $C_\ell = 3(\ell+1)^2$ is a constant independent of L.

Analogously, for the set $\mathcal{I}'' \subset \mathcal{I}$ formed by $J \in \mathbb{R}$ such that $L_{k_0} \le |J + l_0/k_0| \le 2L_{k_0}$ for some $-l_0/k_0 \in \mathcal{R}$, we notice that if $J \in \mathcal{I}''$ then for any other $(k,l) \in \mathcal{N}$ such that $(k,l) \neq (tk_0, tl_0)$, $t \in \mathbb{Z}$, J satisfies that $|J + l/k| \ge 2L_k$. In this case, we will distinguish three types of Fourier coefficients $G_{k,l}$ of G, which are the ones described in points 1,3 and 4 in this proof. Using the same argument as in the previous case, jointly with the bounds for the Fourier coefficients, we have that

$$\begin{aligned}\|G\|_{\mathcal{C}^\ell(\mathcal{I}''\times\mathbb{T}^2),L} &\le \|K\|_{\mathcal{C}^\ell,L} + \frac{(\ell+1)^2}{L}\Big(\|K\|_{\mathcal{C}^\ell,L} + |\psi|_{\mathcal{C}^\ell}\|K\|_{\mathcal{C}^\ell,L} + \|K\|_{\mathcal{C}^\ell,L}\Big)\\ &\le \frac{C_\ell}{L}\|K\|_{\mathcal{C}^\ell,L}\end{aligned}$$

for $\ell = 0, \ldots, r$, where $C_\ell = 4(\ell+1)^2$ is a constant independent of L.

And finally, for the remaining set $\mathcal{I}''' \subset \mathcal{I}$ formed by $J \in \mathbb{R}$, such that $|J + l/k| \ge 2L_k$ for any $(k,l) \in \mathcal{N}$, the Fourier coefficients $G_{k,l}$ of G are just the ones described in points 1 and 3. Arguing as before we have

$$\|G\|_{\mathcal{C}^\ell(\mathcal{I}'''\times\mathbb{T}^2),L} \le \|K\|_{\mathcal{C}^\ell,L} + \frac{(\ell+1)^2}{L}\|K\|_{\mathcal{C}^\ell,L} \le \frac{C_\ell}{L}\|K\|_{\mathcal{C}^\ell,L},$$

for $\ell = 0, \ldots, r$, where $C_\ell = 2(\ell+1)^2$ is a constant independent of L.

So putting all these estimates together we get the desired bound (41) for the whole domain. □

3.2.3. The main averaging result In this section we apply repeatedly the procedure stated in the previous section to the truncated Fourier series of the perturbation $R^{[\le M]}$ in (25), to get a suitable normal form.

We start the averaging procedure with the Hamiltonian (23) truncated up to order M,

$$k_0(J,\varphi,s;\varepsilon) = Z^0(J,\varphi,s;\varepsilon) + \varepsilon R^0(J,\varphi,s;\varepsilon),$$

where $Z^0(J,\varphi,s;\varepsilon) = J^2/2$ and $R^0(J,\varphi,s;\varepsilon) = R^{[\le M]}(J,\varphi,s;\varepsilon)$, which is a trigonometric polynomial of degree M in the angle variables (φ, s).

We will search for a canonical transformation g_0, given by the time-one map of the flow of Hamiltonian εG_0 provided by Lemma 3.3 that eliminates, when it is possible, the dependence on the angle variables (φ, s) at order ε.

According to expression (36), we will refer to *resonances of order 1* as the elements of

$$\mathcal{R}_1 = \{-l/k \in \mathbb{Q} \cap (I_-, I_+), |k| + |l| \le M, k \neq 0, R^0_{k,l}(-l/k; 0) \neq 0\},$$

where $R^0_{k,l}$ are the Fourier coefficients of R^0. For each resonance $-l/k$ in $\mathcal{R}_1$ we will define a strip of size $2L/|k|$, for $L \sim \varepsilon^\alpha$ and $\alpha > 0$, centered on the resonance. We will call *resonant region of order 1* the union of these strips, where the averaging transformation g_0 can not eliminate the dependence on the angle variables, and *non resonant region up* *to order 1* the complementary region in $\widetilde{\Lambda}_\varepsilon$, where $k_0 \circ g_0$ reduces to contain only the harmonic $R^0_{0,0}(J;0)$ at order ε.

Hence, the Hamiltonian $k_1 = k_0 \circ g_0$ is now of the form

$$k_1(J,\varphi,s;\varepsilon) = Z^1(J,\varphi,s;\varepsilon) + \varepsilon^2 R^1(J,\varphi,s;\varepsilon),$$

where the normal form Z^1 is a $\mathcal{C}^r$ function, which has different expressions in the resonant and non resonant regions, and the remainder $\varepsilon^2 R^1$ is a $\mathcal{C}^{r-2}$ function.

Proceeding by induction, we obtain a sequence of Hamiltonians k_{q-1}, for $q \geq 1$, which are normalized up to order ε^{q-1}, that is, in adequate symplectic coordinates Hamiltonian k_{q-1} takes the form

$$k_{q-1}(J,\varphi,s;\varepsilon) = Z^{q-1}(J,\varphi,s;\varepsilon) + \varepsilon^q R^{q-1}(J,\varphi,s;\varepsilon), \tag{45}$$

where, as before, the normal form Z^{q-1} is a $\mathcal{C}^{r-2(q-2)}$ function, which has different expressions in the resonant and non resonant regions up to order $q-1$, and the remainder $\varepsilon^q R^{q-1}$ is a $\mathcal{C}^{r-2(q-1)}$ function.

The set of resonances of order q and its associated resonant and non resonant regions up to order q, are defined recursively in the following way:

Resonances. Resonant and non resonant regions.

Definition 3.4. *The* set of resonances of order $q \geq 1$ *is the set of rational numbers* $r \in \mathcal{R}_q \backslash (\mathcal{R}_1 \cup \cdots \cup \mathcal{R}_{q-1})$*, where* $\mathcal{R}_q$ *is the set of rational numbers* $r \in \mathbb{Q} \cap (I_-, I_+)$ *which admit a representation* $r = -l/k$ *for some integers* k, l *satisfying* $|l| + |k| \leq qM$*, such that* $R^{q-1}_{k,l}(-l/k;0) \neq 0$*; in symbols,*

$$\mathcal{R}_q = \mathcal{R}_q(M) = \left\{ -\frac{l}{k} \in \mathbb{Q} \cap (I_-, I_+) : |k| + |l| \leq qM, k \neq 0, R^{q-1}_{k,l}(-l/k;0) \neq 0 \right\}, \tag{46}$$

where $R^{q-1}_{k,l}$ *are the Fourier coefficients of the remainder* R^{q-1} *in (45).*

Roughly speaking, we call resonances of order q the places in J where the q-th order averaging can not eliminate the dependence on the angles at order q.

Remark 3.5. Notice that, by hypotheses **H3'** in Theorem 2.1, for all $-l_0/k_0 \in \mathbb{Q} \cap (I_-, I_+)$ such that $|(k_0, l_0)| < M_{\mathrm{BG}}$ there exists $t^* \in \mathbb{Z}^2$ such that $h_{t^*k_0, t^*l_0}(0,0,-l_0/k_0;0) \neq 0$ and therefore, by equation (24), $R_{t^*k_0,t^*l_0}(-l_0/k_0;0) \neq 0$. Hence, by definition 3.4 for resonances of order 1, as long as $M_{\mathrm{BG}} \leq M$, all the rational numbers $-l/k$ with $|(k,l)| < M_{\mathrm{BG}}$ are resonant of order 1.

Definition 3.6. *The set* $\mathcal{R}_{[\leq q]}(M)$ *of resonances up to order* q *is the union of sets of resonances of order* i*, for* $i = 1, \ldots, q$*; in symbols,*

$$\mathcal{R}_{[\leq q]} = \mathcal{R}_{[\leq q]}(M) = \bigcup_{i=1,\ldots,q} \mathcal{R}_i(M) \subset \mathbb{Q}. \tag{47}$$

For this set of resonances we define different strips in $\widetilde{\Lambda}_\varepsilon$ of a width depending on a parameter L, which is $L \sim \varepsilon^\alpha$, with $\alpha > 0$. This divides the phase space in two types of regions:

Definition 3.7. *The non-resonant region up to order* q $\mathcal{D}^q_{\rm nr}$ *is the set of points* $(J,\varphi,s)\in\widetilde{\Lambda}_\varepsilon$ *which are at a distance greater than* $2L_k$ *in terms of the* J *variable of any resonance* $-l/k\in\mathcal{R}_{[\le q]}$*, where* $L_k=L/|k|$*; in symbols,*

$$\mathcal{D}^q_{\rm nr}=\mathcal{D}^q_{\rm nr}(M,L)=\left\{(J,\varphi,s)\in(I_-,I_+)\times\mathbb{T}^2:\left|J+\frac{l}{k}\right|\ge 2L_k,\ \text{for}\ -\frac{l}{k}\in\mathcal{R}_{[\le q]}\right\}. \quad (48)$$

Definition 3.8. *The resonant region of order* q $\mathcal{D}_{r,q}$ *is the set of points* $(J,\varphi,s)\in\widetilde{\Lambda}_\varepsilon$ *which are at a distance smaller than* $L_k=L/|k|$ *in terms of the* J *variable from any resonance* $-l/k\in\mathcal{R}_q\backslash(\mathcal{R}_1\cup\ldots\cup\mathcal{R}_{q-1})$*; in symbols,*

$$\begin{aligned}\mathcal{D}_{r,q}=\mathcal{D}_{r,q}(M,L) \;=\; &\left\{(J,\varphi,s)\in(I_-,I_+)\times\mathbb{T}^2:\left|J+\tfrac{l}{k}\right|\le L_k,\right.\\ &\left.\text{for some}\ -\tfrac{l}{k}\in\mathcal{R}_q\backslash(\mathcal{R}_1\cup\ldots\cup\mathcal{R}_{q-1})\right\}.\end{aligned} \quad (49)$$

The union of resonant regions of order i, for $i=1,\ldots,q$ gives us the resonant region up to order q, which can be defined in the following way:

Definition 3.9. *The resonant region up to order* q $\mathcal{D}^q_{\rm r}$ *is the set of points* $(J,\varphi,s)\in\widetilde{\Lambda}_\varepsilon$ *which are at a distance smaller than* $L_k=L/|k|$ *in terms of the* J *variable from any resonance* $-l/k\in\mathcal{R}_{[\le q]}$*; in symbols,*

$$\mathcal{D}^q_{\rm r}=\mathcal{D}^q_{\rm r}(M,L)=\left\{(J,\varphi,s)\in(I_-,I_+)\times\mathbb{T}^2:\left|J+\frac{l}{k}\right|\le L_k,\text{for some}-\frac{l}{k}\in\mathcal{R}_{[\le q]}\right\} \quad (50)$$

The dependence of these domains on M and L: $\mathcal{D}^q_{\rm nr}=\mathcal{D}^q_{\rm nr}(M,L)$, $\mathcal{D}_{r,q}=\mathcal{D}_{r,q}(M,L)$ and $\mathcal{D}^q_{\rm r}=\mathcal{D}^q_{\rm r}(M,L)$, will be suppressed to simplify notation.

Remark 3.10. Note that, by Remark 3.5, the big gaps region $\mathcal{D}_{BG}$ introduced in (21) is contained in the resonant region of order 1 $\mathcal{D}_{r,1}$.

The precise result to obtain a global normal form for the reduced Hamiltonian by applying repeatedly the averaging procedure, jointly with the estimates for the bounds of the normal form terms and the expression of the order of truncation M and the constant L as functions of ε are stated in the following Theorem 3.11:

Theorem 3.11. *Let* n,m *be any given integers satisfying* $1\le 2m\le n$*. Given* ρ *a real number satisfying*

$$\rho<\frac{1}{n}, \quad (51)$$

and r *an integer verifying*

$$r\ge(1/\rho-2)m+2, \quad (52)$$

consider a $\mathcal{C}^r$ *Hamiltonian of the form (23):*

$$k(J,\varphi,s;\varepsilon)=\frac{J^2}{2}+\varepsilon R(J,\varphi,s;\varepsilon), \quad (53)$$

satisfying $\varepsilon R(J,\varphi,s;\varepsilon)=\mathcal{O}_{\mathcal{C}^r}(\varepsilon)$*.*

Introduce $M\sim\varepsilon^{-\rho}$*, for any* $-l/k\in\mathcal{R}_{[\le m]}(M)$*, introduced in (47), consider* $L_k=L/|k|$*, where*

$$L=C\varepsilon^{\alpha} \quad (54)$$

with

$$\rho \le \alpha < 1/n \tag{55}$$

and C *a constant independent of* ε, *such that for* $-l/k \in \mathcal{R}_{[\le m]}$, *the real intervals* $\mathcal{I}_{-l/k} \equiv [-l/k - 2L_k, l/k + 2L_k]$ *are disjoint. Then, there exists a symplectic change of variables, depending on time,* $(J, \varphi, s) = g(\mathcal{B}, \phi, s)$, *periodic in* ϕ *and* s, *and of class* $\mathcal{C}^{r-2m}$, *which is* ε*-close to the identity in the* $\mathcal{C}^{n-2m-1}$ *sense, such that transforms the Hamiltonian system associated to* $k(J, \varphi, s; \varepsilon)$ *into a Hamiltonian system of Hamiltonian*

$$\bar{k}^m(\mathcal{B}, \phi, s; \varepsilon) = \bar{Z}^m(\mathcal{B}, \phi, s; \varepsilon) + \varepsilon^{m+1}\bar{R}^m(\mathcal{B}, \phi, s; \varepsilon), \tag{56}$$

where the function $\bar{Z}^m$ *is of class* $\mathcal{C}^{r-2m+2}$ *and* $\bar{R}^m$ *is of class* $\mathcal{C}^{r-2m}$ *and they verify:*

1. *If* $\mathcal{B} \notin \bigcup_{-l/k \in \mathcal{R}_{[\le m]}} \mathcal{I}_{-l/k}$, *then*
$$\bar{Z}^m(\mathcal{B}, \phi, s; \varepsilon) = \frac{1}{2}\mathcal{B}^2 + \varepsilon \widetilde{Z}^m(\mathcal{B}; \varepsilon), \tag{57}$$
for any $(\mathcal{B}, \phi, s) \in \mathcal{D}^m_{\mathrm{nr}}$ *($\mathcal{D}^m_{\mathrm{nr}}$ was introduced in (48)).*
2. *If* $\mathcal{B} \in \mathcal{I}_{-l_0/k_0}$ *for some* $-l_0/k_0 \in \mathcal{R}_i \setminus (\mathcal{R}_1 \cup \ldots \mathcal{R}_{i-1})$, *for some* $1 \le i \le m$, *then*
$$\bar{Z}^m(\mathcal{B}, \phi, s; \varepsilon) = \frac{1}{2}\mathcal{B}^2 + \varepsilon \widetilde{Z}^m(\mathcal{B}; \varepsilon) + \varepsilon^i U_m^{k_0,l_0}(k_0\phi + l_0 s; \varepsilon), \tag{58}$$
for any $(\mathcal{B}, \phi, s) \in \mathcal{D}_{\mathrm{r},i}$, *($\mathcal{D}_{\mathrm{r},i}$ was introduced in (49)).*
In a particular case of a resonance $-l_0/k_0$ *of order 1,* $U_m^{k_0,l_0}(k_0\phi + l_0 s; 0)$ *does not depend on* m *and is given by*
$$U_m^{k_0,l_0}(\theta; 0) = U_1^{k_0,l_0}(\theta) = \sum_{\substack{t \in \mathbb{Z}-\{0\} \\ |t|(|k_0|+|l_0|) \le M}} R_{tk_0,tl_0}(-l_0/k_0; 0)e^{\mathrm{i}t\theta} \tag{59}$$
where $\theta = k_0\phi + l_0 s$ *and* $R_{k,l}(J; \varepsilon)$ *are the Fourier coefficients of the perturbation* $R(J, \varphi, s; 0)$ *with respect to* (φ, s).
3. *The function* $\varepsilon\widetilde{Z}^m(\mathcal{B}; \varepsilon)$ *in (57) and (58) is a polynomial of degree* m *in* ε, *whose term of order* $q + 1$ *is of class* $\mathcal{C}^{r-2q}$ *and of size* $\mathcal{O}_{\mathcal{C}^\ell}(\varepsilon^{q+1-\alpha(\ell+2q)})$, *for* $\ell = 0, \ldots, n - 2q$ *and* $q = 0, \ldots, m - 1$. *The function* $\varepsilon^i U_m^{k_0,l_0}(k_0\phi + l_0 s; \varepsilon)$ *in (58) is a polynomial of degree* m *in* ε *and a trigonometric polynomial in* $\theta = k_0\phi + l_0 s$, *which is* $\mathcal{O}_{\mathcal{C}^\ell,\theta}(\varepsilon^{i-2\alpha(i-1)}|(k_0, l_0)|^{-r+2(i-1)})$, *for* $\ell = 0, \ldots, n - 2(i - 1)$. *The function* $\varepsilon^{m+1}\bar{R}(\mathcal{B}, \phi, s; \varepsilon)$ *in (56) is* $\mathcal{O}_{\mathcal{C}^\ell}(\varepsilon^{m+1-\alpha(\ell+2m)})$, *for* $\ell = 0, \ldots, n - 2m$. *Finally, the change of variables* $(J, \varphi, s) = g(\mathcal{B}, \phi, s)$ *satisfies* $g - \mathrm{Id} = \mathcal{O}_{\mathcal{C}^\ell}(\varepsilon^{1-\alpha(\ell+2)})$, *for* $\ell = 0, \ldots, n - 2m$.

Remark 3.12. We always will consider that Hamiltonian (53) is Hamiltonian (23) and therefore, by equation (24) and Remark 2.9, the function $U_m^{k_0,l_0}$ given in (59) for a resonance $-l_0/k_0 \in \mathcal{R}_1$ is equal to the function U^{k_0,l_0} in hypothesis **H3′**:

$$U_m^{k_0,l_0}(\theta; 0) = U^{k_0,l_0}(\theta) = \sum_{\substack{t \in \mathbb{Z}-\{0\}, \\ |t||(k_0,l_0)| < M}} h_{tk_0,tl_0}(0, 0, -l_0/k_0; 0)e^{\mathrm{i}t\theta}. \tag{60}$$

By the same reason $\widetilde{Z}^m(\mathcal{B}; 0)$ in formulae (57) and (58) is equal to $h(0, 0, \mathcal{B}; 0)$.

Remark 3.13. Note that the bound on the trigonometric polynomial $\varepsilon^i U_m^{k_0,l_0}(\theta;\varepsilon)$, where $\theta = k_0\phi + l_0 s$, is more precise because it incorporates the size of its Fourier coefficients. We use the notation $\mathcal{O}_{\mathcal{C}^\ell,\theta}$ to emphasize that we are bounding the derivatives with respect to the variable θ.

Remark 3.14. Notice that although the remainder term $\varepsilon^{m+1}\bar{R}^m$ is $\mathcal{C}^{r-2m}$, it is bounded in the supremum norm $|\cdot|_{\mathcal{C}^\ell}$ for ℓ only up to $n-2m$, for $n<r$, which is enough for the future application of KAM Theorem.

3.2.4. Proof of Theorem 3.11 The proof of this theorem will follow by the repeated application of the inductive Lemma 3.18 m times. Before stating it, we need two previous Lemmas that we will use to prove Lemma 3.18 and finally Theorem 3.11.

Lemma 3.15. *Let $G(J,\varphi,s)$ a Hamiltonian and assume that G is $\mathcal{C}^r$ trigonometric polynomial of order M defined in a compact domain $\mathcal{I}\times\mathbb{T}^2$, with $\mathcal{I}\subset\mathbb{R}$, such that* $\sup_{x\in\mathcal{I}\times\mathbb{T}^2}|x|\le D$. *Consider the $\mathcal{C}^{r-1}$ change of variables defined on $\mathcal{I}\times\mathbb{T}^2$,*

$$(J,\varphi,s) = g_t(\mathcal{B},\phi,s),$$

given by the time-t map of the flow of Hamiltonian $\varepsilon^p G(J,\varphi,s)$, for some $p\in\mathbb{N}$. Assume that G is $\mathcal{O}_{\mathcal{C}^\ell}(\varepsilon^{\eta_\ell})$, η_ℓ being some positive number. Then,

$$\max_{0\le t\le 1}|g_t|_{\mathcal{C}^\ell}\le D_\ell,\qquad \max_{0\le t\le 1}|g_t-\mathrm{Id}|_{\mathcal{C}^\ell}\le D'_\ell\,\varepsilon^{\eta_{\ell+1}} \tag{61}$$

for $\ell=0,\dots,r-1$, D_ℓ and D'_ℓ being some constants, which depend on the domain and ℓ, but not on ε. In terms of the notation introduced in Section 2.1, the above inequalities read $g_t=\mathcal{O}_{\mathcal{C}^\ell}(1)$ and $g_t-\mathrm{Id}=\mathcal{O}_{\mathcal{C}^\ell}(\varepsilon^{\eta_{\ell+1}})$, for $\ell=0,\dots,r-1$ and $0\le t\le 1$.

Proof. By the fundamental theorem of calculus we can write

$$g_t(x) = x + \int_0^t \frac{\partial g_\tau}{\partial\tau}(x)\mathrm{d}\tau = x+\int_0^t \mathrm{J}\nabla G\circ g_\tau(x)\mathrm{d}\tau,$$

where $x=(\mathcal{B},\phi,s)\in\mathcal{I}\times\mathbb{T}^2$ and J is the canonical matrix of the symplectic form $\omega=\mathrm{d}J\wedge\mathrm{d}\varphi+\mathrm{d}A\wedge\mathrm{d}s$. The extra variable A, conjugated to the angle s, was introduced to make apparent the symplectic character of the change of variables.

Using formula (C.5) in Appendix C we obtain

$$\begin{aligned}|g_t|_{\mathcal{C}^\ell} &\le |\mathrm{Id}|_{\mathcal{C}^\ell}+\int_0^1|\mathrm{J}\nabla G\circ g_\tau|_{\mathcal{C}^\ell}\,\mathrm{d}\tau\\ &\le |\mathrm{Id}|_{\mathcal{C}^\ell}+C_\ell\int_0^1\left(|\mathrm{J}\nabla G|_{\mathcal{C}^1}|g_\tau|_{\mathcal{C}^\ell}+|\mathrm{J}\nabla G|_{\mathcal{C}^\ell}|g_\tau|^\ell_{\mathcal{C}^{\ell-1}}\right)\mathrm{d}\tau\end{aligned} \tag{62}$$

for $\ell=2,\dots,r-1$, where C_ℓ is a constant depending on ℓ; and

$$|g_t|_{\mathcal{C}^1}\le|\mathrm{Id}|_{\mathcal{C}^1}+\int_0^1|\mathrm{J}\nabla G|_{\mathcal{C}^1}|g_\tau|_{\mathcal{C}^1}\,\mathrm{d}\tau.$$

Let us define $a_\ell=\max\limits_{0\le t\le 1}|g_t|_{\mathcal{C}^\ell}$. Then,

$$a_1\le D+\delta_1 a_1,$$

and

$$a_\ell \le D + \delta_1 a_\ell + C_\ell \delta_\ell a_{\ell-1}^\ell, \qquad \text{for } \ell \ge 2,$$

with $\delta_\ell = |G|_{\mathcal{C}^{\ell+1}}$. Hence,

$$a_\ell \le \frac{D + \delta_\ell a_{\ell-1}^\ell}{1-\delta_1} \qquad \text{for } \ell \ge 2.$$

Since $\delta_1 \sim \varepsilon^{\eta_2} \ll 1$ and $\delta_\ell \sim \varepsilon^{\eta_{\ell+1}} \ll 1$, it is easy to check by induction that $a_\ell \le D_\ell$, for $\ell \ge 1$, D_ℓ being some constant independent of ε.

Denoting by $b_\ell = \max\limits_{0\le t\le 1} |g_t - \mathrm{Id}|_{\mathcal{C}^\ell}$, one has

$$b_1 \le \delta_1 a_1,$$

and

$$b_\ell \le \delta_1 a_\ell + C_\ell \delta_\ell a_{\ell-1}^\ell, \qquad \text{for } \ell \ge 2.$$

So that,

$$b_\ell \le D_\ell \delta_1 + C_\ell D_{\ell-1}^\ell \delta_\ell \le D'_\ell \delta^\ell = D'_\ell \varepsilon^{\eta_{\ell+1}},$$

for $\ell \ge 1$, D'_ℓ being some constant independent of ε. □

Since the averaging procedure is based on the method of Lie transforms, the transformed Hamiltonian will be expressed in terms of Poisson brackets. In the following Lemma 3.16 we give an estimate for the bound of the Poisson bracket of two functions, where the second one is a generating function, in terms of the bounds on the norm (30) of each one.

Lemma 3.16. *Let ρ, α be two positive real numbers, such that $\rho \le \alpha$ and $M \sim \varepsilon^{-\rho}$ and $L = C\varepsilon^\alpha$. Given $F^p(J,\varphi,s)$ and $G^q(J,\varphi,s)$ two trigonometric polynomials in (φ, s), assume that $F^p(J,\varphi,s)$ is a $\mathcal{C}^n$, $n > 0$, function in J and a trigonometric polynomial of degree $M_p = (p+1)M$ and $G^q(J,\varphi,s)$ is a $\mathcal{C}^m$, $m > 0$, function in J and a trigonometric polynomial of degree $M_q = (q+1)M$, that satisfy $\|\varepsilon^{p+1}F^p\|_{\mathcal{C}^\ell,L} \preceq \varepsilon^{p+1-\alpha(2p)}$ and $\|\varepsilon^{q+1}G^q\|_{\mathcal{C}^\ell,L} \preceq \varepsilon^{q+1-\alpha(2q+1)}$, for $\ell = 0,\dots,n$, with $\varepsilon > 0$. Then $\{F^p, G^q\}$ is a $\mathcal{C}^r$ function in J, for $r = \min(n,m) - 1$ and a trigonometric polynomial of degree $M_{\widetilde{p}} = (\widetilde{p}+1)M$ in (φ, s), where $\widetilde{p} = p + q + 1$, and $\varepsilon^{\widetilde{p}+1}F^{\widetilde{p}} = \{\varepsilon^{p+1}F^p, \varepsilon^{q+1}G^q\}$ satisfies*

$$\left\|\varepsilon^{\widetilde{p}+1}F^{\widetilde{p}}\right\|_{\mathcal{C}^\ell,L} \preceq \varepsilon^{\widetilde{p}+1-\alpha(2\widetilde{p})},$$

for $\ell = 0,\dots,r$.

Proof. From

$$\{F^p, G^q\} = \frac{\partial F^p}{\partial\varphi}\frac{\partial G^q}{\partial J} - \frac{\partial F^p}{\partial J}\frac{\partial G^q}{\partial\varphi},$$

we have

$$\{F^p, G^q\} = \sum_{\substack{(k,l)\in\mathbb{Z}^2,\\ |k|+|l|\le M_p}} \mathrm{i}kF^p_{k,l}(J)e^{\mathrm{i}(k\varphi+ls)} \sum_{\substack{(k,l)\in\mathbb{Z}^2,\\ |k|+|l|\le M_q}} \frac{\partial G^q_{k,l}(J)}{\partial J}e^{\mathrm{i}(k\varphi+ls)} - \sum_{\substack{(k,l)\in\mathbb{Z}^2,\\ |k|+|l|\le M_p}} \frac{\partial F^p_{k,l}(J)}{\partial J}e^{\mathrm{i}(k\varphi+ls)} \sum_{\substack{(k,l)\in\mathbb{Z}^2,\\ |k|+|l|\le M_q}} \mathrm{i}kG^p_{k,l}(J)e^{\mathrm{i}(k\varphi+ls)}$$

It is clear from this expression that $\{F^p, G^q\}$ is a trigonometric polynomial of degree $M_p + M_q = (p+q+2)M$.

On the other hand, using equation (34), it follows that

$$\begin{aligned}
&\left\|\{\varepsilon^{p+1}F^p, \varepsilon^{q+1}G^q\}\right\|_{\mathcal{C}^\ell,L} \le \\
&\left\|\varepsilon^{p+1}\sum_{\substack{(k,l)\in\mathbb{Z}^2,\\ |k|+|l|\le M_p}} \mathrm{i}kF^p_{k,l}(J)e^{\mathrm{i}(k\varphi+ls)}\right\|_{\mathcal{C}^\ell,L} \left\|\varepsilon^{q+1}\sum_{\substack{(k,l)\in\mathbb{Z}^2,\\ |k|+|l|\le M_q}} \frac{\partial G^q_{k,l}(J)}{\partial J}e^{\mathrm{i}(k\varphi+ls)}\right\|_{\mathcal{C}^\ell,L} \\
&+ \left\|\varepsilon^{p+1}\sum_{\substack{(k,l)\in\mathbb{Z}^2,\\ |k|+|l|\le M_p}} \frac{\partial F^p_{k,l}(J)}{\partial J}e^{\mathrm{i}(k\varphi+ls)}\right\|_{\mathcal{C}^\ell,L} \left\|\varepsilon^{q+1}\sum_{\substack{(k,l)\in\mathbb{Z}^2,\\ |k|+|l|\le M_q}} \mathrm{i}kG^p_{k,l}(J)e^{\mathrm{i}(k\varphi+ls)}\right\|_{\mathcal{C}^\ell,L} \\
&\le \left\|\varepsilon^{p+1}\sum_{\substack{(k,l)\in\mathbb{Z}^2,\\ |k|+|l|\le M_p}} F^p_{k,l}(J)e^{\mathrm{i}(k\varphi+ls)}\right\|_{\mathcal{C}^{\ell+1},L} \frac{1}{L}\left\|\varepsilon^{q+1}\sum_{\substack{(k,l)\in\mathbb{Z}^2,\\ |k|+|l|\le M_q}} G^q_{k,l}(J)e^{\mathrm{i}(k\varphi+ls)}\right\|_{\mathcal{C}^{\ell+1},L} \\
&+ \frac{1}{L}\left\|\varepsilon^{p+1}\sum_{\substack{(k,l)\in\mathbb{Z}^2,\\ |k|+|l|\le M_p}} F^p_{k,l}(J)e^{\mathrm{i}(k\varphi+ls)}\right\|_{\mathcal{C}^{\ell+1},L} \left\|\varepsilon^{q+1}\sum_{\substack{(k,l)\in\mathbb{Z}^2,\\ |k|+|l|\le M_q}} G^q_{k,l}(J)e^{\mathrm{i}(k\varphi+ls)}\right\|_{\mathcal{C}^{\ell+1},L} \\
&\le \frac{2}{L}\left\|\varepsilon^{p+1}F^p\right\|_{\mathcal{C}^{\ell+1},L}\left\|\varepsilon^{q+1}G^q\right\|_{\mathcal{C}^{\ell+1},L}.
\end{aligned}$$

Using now the hypotheses on $\varepsilon^{q+1}F^p$ and $\varepsilon^{p+1}G^p$ in this Lemma and the fact that $L = C\varepsilon^\alpha$, where C is a constant independent of ε, we have

$$\begin{aligned}
\left\|\{\varepsilon^{p+1}F, \varepsilon^{q+1}G\}\right\|_{\mathcal{C}^\ell,L} &\preceq \varepsilon^{-\alpha}\varepsilon^{p+1-\alpha(2p)}\varepsilon^{q+1-\alpha(2q+1)} \\
&= \varepsilon^{p+q+2-\alpha(2(p+q+1))} \\
&= \varepsilon^{\tilde{p}+1-\alpha(2\tilde{p})}.
\end{aligned}$$

□

Remark 3.17. This Lemma will be applied a certain number of times and expresses the fact that given two functions $\varepsilon^{p+1}F^p$ and $\varepsilon^{q+1}G^q$, which are trigonometric polynomials in (φ, s) of degree $M_p = (p+1)M$ and $M_q = (q+1)M$, respectively, with bounds $\|\varepsilon^{p+1}F^p\|_{\mathcal{C}^\ell,L} \preceq \varepsilon^{p+1-\alpha(2p)}$ and $\|\varepsilon^{q+1}G^q\|_{\mathcal{C}^\ell,L} \preceq \varepsilon^{q+1-\alpha(2q+1)}$, its Poisson bracket is a function $\varepsilon^{\widetilde{p}+1}F^{\widetilde{p}}$, with $\widetilde{p}=p+q+1$, that is, $\varepsilon^{\widetilde{p}+1}F^{\widetilde{p}}=\{\varepsilon^{p+1}F^{p},\varepsilon^{q+1}G^{q}\}$ is a trigonometric polynomial in (φ,s) of degree $M_{\widetilde{p}}=(\widetilde{p}+1)M$ with a bound $\left\|\varepsilon^{\widetilde{p}+1}F^{\widetilde{p}}\right\|_{\mathcal{C}^{\ell},L}\preceq\varepsilon^{\widetilde{p}+1-\alpha(2\widetilde{p})}$.

Notice that this process of "$\varepsilon^{q+1}G^{q}$ Poisson-bracketing" can be iterated: $\varepsilon^{\widehat{p}+1}F^{\widehat{p}}=\{\varepsilon^{\widetilde{p}+1}F^{\widetilde{p}},\varepsilon^{q+1}G^{q}\}$, with $\widehat{p}=\widetilde{p}+q+1$, is a trigonometric polynomial in (φ,s) of degree $M_{\widehat{p}}=(\widehat{p}+1)M$ with a bound $\left\|\varepsilon^{\widehat{p}+1}F^{\widehat{p}}\right\|_{\mathcal{C}^{\ell},L}\preceq\varepsilon^{\widehat{p}+1-\alpha(2\widehat{p})}$.

We state and prove now the iterative Lemma 3.18 for averaging, which we will apply a finite number of times $q=1,\dots,m$ in the proof of Theorem 3.11 and m will be chosen $m\le 10$ in Theorem 3.28. It basically tells us that given a Hamiltonian already in normal form up to some order ε^{q}, we can produce another Hamiltonian which is normalized up to order ε^{q+1}. The averaged Hamiltonian is given rather explicitly both in the resonant regions and in the non-resonant ones, which are redefined at every step according to the new resonances that will come up.

Lemma 3.18. *Let $r>n>1$ and $0\le 2q<n$ be any given integers. Consider a Hamiltonian of the form*

$$k_q(J,\varphi,s;\varepsilon)=Z^{q}(J,\varphi,s;\varepsilon)+\varepsilon^{q+1}R^{q}(J,\varphi,s;\varepsilon),$$

satisfying the following hypotheses:

1. *$Z^{0}(J,\varphi,s;\varepsilon)=\frac{J^2}{2}$ and, for $q\ge 1$, $Z^{q}(J,\varphi,s;\varepsilon)$ is a $\mathcal{C}^{r-2q+2}$ function that verifies: There exist finite sets $\mathcal{R}_i\subset\mathbb{Q}$, $i=1,\dots,q$, depending on $M\sim\varepsilon^{-\rho}$, where ρ is a positive number satisfying $\rho<\frac{1}{n}$, and a number $L=C\varepsilon^{\alpha}>0$, which satisfy hypothesis (55), that is, $\rho\le\alpha<\frac{1}{n}$ and C a constant independent of ε, such that:*
   - *1a For a resonance $-l/k$ up to order q, that is $-l/k\in\mathcal{R}_{[\le q]}\equiv\bigcup_{i=1\dots q}\mathcal{R}_i$ (see (47)), the intervals $\mathcal{I}_{-l/k}\equiv[-l/k-2L_k,-l/k+2L_k]$, with $L_k=L/|k|$, are disjoint.*
   - *1b If $J\notin\bigcup_{-l/k\in\mathcal{R}_{[\le q]}}\mathcal{I}_{-l/k}$, then*
     $$Z^{q}(J,\varphi,s;\varepsilon)=\frac{J^2}{2}+\varepsilon\widetilde{Z}^{q}(J;\varepsilon),$$
     for any $(J,\varphi,s)\in\mathcal{D}^{q}_{\mathrm{nr}}$ ($\mathcal{D}^{q}_{\mathrm{nr}}$ was introduced in (48)), where $\varepsilon\widetilde{Z}^{q}(J;\varepsilon)$ is a polynomial of degree q in ε whose term of order $p+1$ is $\mathcal{O}_{\mathcal{C}^{\ell},L}(\varepsilon^{p+1-\alpha(2p)})$, for $\ell=0,\dots,r-2p$ and $p=0,\dots,q-1$.
   - *1c If $J\in\mathcal{I}_{-l_0/k_0}$, for some resonance $-l_0/k_0$ of order q, that is $-l_0/k_0\in\mathcal{R}_i\setminus\mathcal{R}_1\cup\dots\cup\mathcal{R}_{i-1}$ for some $1\le i\le q$, then*
     $$Z^{q}(J,\varphi,s;\varepsilon)=\frac{J^2}{2}+\varepsilon\widetilde{Z}^{q}(J;\varepsilon)+\varepsilon^{i}U_q^{k_0,l_0}(k_0\varphi+l_0 s;\varepsilon)$$
     for any $(J,\varphi,s)\in\mathcal{D}_{\mathrm{r},i}$ ($\mathcal{D}_{\mathrm{r},i}$ was introduced in (49)), where $\varepsilon\widetilde{Z}^{q}(J;\varepsilon)$ is a polynomial of degree q in ε and $U_q^{k_0,l_0}(\theta;\varepsilon)$ is a polynomial of degree $q-i$ in ε and a trigonometric polynomial in $\theta=k_0\varphi+l_0 s$. The term of order $p+1$ in ε of Z^{q} is $\mathcal{O}_{\mathcal{C}^{\ell},L}(\varepsilon^{p+1-\alpha(2p)})$, for $\ell=0,\dots,r-2p$ and $p=0,\dots,q-1$.
2. *$\varepsilon^{q+1}R^{q}(J,\varphi,s;\varepsilon)$ is a $\mathcal{C}^{r-2q}$ function and is $\mathcal{O}_{\mathcal{C}^{\ell}}(\varepsilon^{q+1-\alpha(\ell+2q)})$, for $\ell=0,\dots,n-2q$. For the particular case of the first iteration ($q=0$), εR^{0} is $\mathcal{O}_{\ell}(\varepsilon)$, for $\ell=0,\dots,n$.*

The term of order $i+1$ of the Taylor expansion with respect to ε of $\varepsilon^{q+1}R^q(J,\varphi,s;\varepsilon)$ is a trigonometric polynomial in (φ,s) of degree $M_i = (i+1)M$ and is $\mathcal{O}_{\mathcal{C}^\ell,L}(\varepsilon^{i+1-\alpha(2i)})$, for $\ell = 0,\dots,r-q-i$ and for $i = q,\dots,r-q$.

Denote $K = R^q(J,\varphi,s;0)$, which is the term of the perturbation of order exactly $q+1$ in ε. Following Definition 3.4, introduce the set

$$\mathcal{R}_{q+1} = \{-l/k \in \mathbb{Q}\cap(I_-,I_+), |k|+|l| \le M_q, k \neq 0, R^q_{k,l}(-l/k;0) \neq 0\}, \quad (63)$$

where $M_q = (q+1)M$ and $R^q_{k,l}$ are the Fourier coefficients of R^q.

Choose a new value of C, independent of ε, in $L = C\varepsilon^\alpha$, such that the intervals $\mathcal{I}_{-l/k} \equiv [-l/k - 2L_k, -l/k + 2L_k]$, with $L_k = L/|k|$, are disjoint for $-l/k \in \mathcal{R}_{[\le q+1]}$.

Let $G(J,\varphi,s) = G_q(J,\varphi,s)$ be the $\mathcal{C}^{r-2q-1}$ trigonometric polynomial of order M_q given by Lemma 3.3, verifying (37) with $K = R^q(J,\varphi,s;0)$.

Then, the $\mathcal{C}^{r-2q-2}$ change of variables

$$(J,\varphi,s) = g_q(\mathcal{B},\phi,s),$$

given by the time-one map of the flow of Hamiltonian $\varepsilon^{q+1}G_q(\mathcal{B},\phi,s)$, transforms the Hamiltonian $k_q(J,\varphi,s;\varepsilon)$ into a Hamiltonian $k_{q+1} = k_q \circ g_q$ of the form

$$k_{q+1}(\mathcal{B},\phi,s;\varepsilon) = Z^{q+1}(\mathcal{B},\phi,s;\varepsilon) + \varepsilon^{q+2}R^{q+1}(\mathcal{B},\phi,s;\varepsilon),$$

with

$$Z^{q+1}(\mathcal{B},\phi,s;\varepsilon) = Z^q(\mathcal{B},\phi,s;\varepsilon) + \varepsilon^{q+1}\bar{R}^q(\mathcal{B},\phi,s;0)$$

where $\bar{R}^q(\mathcal{B},\phi,s;0) = \bar{K}(\mathcal{B},\phi,s)$ given in Lemma 3.3, is a $\mathcal{C}^{r-2q}$ function, such that

i. If $\mathcal{B} \notin \bigcup_{-l/k\in\mathcal{R}_{[\le q+1]}} \mathcal{I}_{-l/k}$, then

$$\bar{R}^q(\mathcal{B},\phi,s;0) = R^q_{0,0}(\mathcal{B};0),$$

for any $(\mathcal{B},\phi,s) \in \mathcal{D}^{q+1}_{\mathrm{nr}}$ and $\varepsilon^{q+1}\bar{R}^q$ is $\mathcal{O}_{\mathcal{C}^\ell,L}(\varepsilon^{q+1-\alpha(2q)})$, for $\ell = 0,\dots,r-2q$.

ii. If $\mathcal{B} \in \mathcal{I}_{-l_0/k_0}$, for some $-l_0/k_0 \in \mathcal{R}_i \setminus \mathcal{R}_1 \cup \dots \cup \mathcal{R}_{i-1}$ for some $1 \le i \le q+1$, then

$$\bar{R}^q(\mathcal{B},\phi,s) = R^q_{0,0}(\mathcal{B};0) + \sum_{\substack{t\in\mathbb{Z}-\{0\},\\ |t|(|k_0|+|l_0|)\le M_q}} R^q_{tk_0,tl_0}(-l_0/k_0;0)e^{\mathrm{i}t\theta}, \quad (64)$$

for any $(\mathcal{B},\phi,s) \in \mathcal{D}_{\mathrm{r},i}$, where $R^q_{k,l}(J;\varepsilon)$ are the Fourier coefficients of the function $R^q(J,\varphi,s;\varepsilon)$ with respect to (φ,s). Moreover, $\varepsilon^{q+1}\bar{R}^q$ is $\mathcal{O}_{\mathcal{C}^\ell,L}(\varepsilon^{q+1-\alpha(2q)})$, for $\ell = 0,\dots,r-2q$.

Moreover, the Hamiltonian $Z^{q+1}(\mathcal{B},\phi,s;\varepsilon)$ verifies properties 1b and 1c up to order $q+1$, and $R^{q+1}(\mathcal{B},\phi,s;\varepsilon)$ verifies property 2 replacing q by $q+1$.

Remark 3.19. Note that all the terms of order $p+1$, for $p \ge 0$, in the Taylor expansion in ε that appear in Lemma 3.18 are $\mathcal{C}^{r-2p}$ functions in J and trigonometric polynomials in the variables (φ,s) and they are bounded independently of ε in the Fourier weighted norm $\|\cdot\|_{\mathcal{C}^\ell,L}$ defined in (30) for ℓ up to $r-2p$. However, the whole remainder term $\varepsilon^{q+2}R^{q+1}$ is not a trigonometric polynomial in the variables (φ,s), so we can not use the Fourier weighted norm. In this case we estimate their supremum norm $|\cdot|_{\mathcal{C}^\ell}$ defined in (2), but only for ℓ up to $n-2q$, as in Theorem 3.11 (see Remark 3.14).

Proof. We will apply Lemma 3.3 with $K = R^q(J, \varphi, s; 0)$, which is a $\mathcal{C}^{r-2q}$ function, as well as a trigonometric polynomial in (φ, s) of degree $M_q = (q + 1)M$. Accordingly, by Definition 3.4, resonances of order $q + 1$ correspond to the set of rational numbers $r \in \mathcal{R}_{q+1} \setminus (\mathcal{R}_1 \cup \ldots \cup \mathcal{R}_q)$.

Let us see first that taking $L = C\varepsilon^\alpha$, with α satisfying $\alpha < 1/n$ and $C = C_q$ chosen adequately, the real intervals $\mathcal{I}_{-l/k} = [-l/k - 2L_k, -l/k + 2L_k]$, with $L_k = L/|k|$, for $-l/k \in \mathcal{R}_{[\leq q+1]}$ are disjoint. Indeed, the distance d_{k,k_0} between any two resonances $-l_0/k_0, -l/k \in \mathcal{R}_{[\leq q+1]}$ is greater or equal than $1/(|k_0||k|)$. In order to avoid overlapping between all these intervals, the distance d_{k,k_0} must be greater than $2L_{k_0} + 2L_k$. Taking into account that we only consider resonances with denominators $|k|, |k_0| \leq (q + 1)M$, the condition that ensures that these intervals are separated is $1/((q + 1)M) > 4L$, which requires $\rho \leq \alpha$ in terms of exponents of ε. This is guaranteed by the hypothesis on α and ρ in this Lemma.

Hence, we can apply Lemma 3.3, obtaining a $\mathcal{C}^{r-2q-1}$ function $G_q(J, \varphi, s)$ and a $\mathcal{C}^{r-2q}$ function $\bar{K} = R^q(J, \varphi, s)$, which are also trigonometric polynomials in (φ, s) of degree M_q.

Under the canonical change of variables $(J, \varphi, s) = g_q(\mathcal{B}, \phi, s)$, where g_q is the time-one map of the flow of Hamiltonian $\varepsilon^{q+1}G_q$, the extended autonomous Hamiltonian $A + k_q$ becomes

$$\begin{aligned} A + k_{q+1} &= (A + k_q) \circ g_q \\ &= (A + Z^q + \varepsilon^{q+1} R^q) \circ g_q \\ &= A + Z^q + \varepsilon^{q+1}(\{A + Z^0, G_q\} + R^q(\cdot, 0)) \\ &\quad + (Z^q - Z^0) \circ g_q - (Z^q - Z^0) \\ &\quad + (A + Z^0) \circ g_q - (A + Z^0) - \{A + Z^0, \varepsilon^{q+1} G_q\} \\ &\quad + \varepsilon^{q+1}(R^q \circ g_q - R^q) + \varepsilon^{q+1}(R^q - R^q(\cdot, 0)) \\ &:= A + Z^q + \varepsilon^{q+1}\bar{R}^q + \varepsilon^{q+2} R^{q+1}, \end{aligned}$$

where

$$\bar{R}^q = \{A + Z^0, G_q\} + R^q(\cdot, 0), \tag{65}$$

and

$$\begin{aligned} \varepsilon^{q+2} R^{q+1} &= (Z^q - Z^0) \circ g_q - (Z^q - Z^0) \\ &\quad + (A + Z^0) \circ g_q - (A + Z^0) - \{A + Z^0, \varepsilon^{q+1} G_q\} \\ &\quad + \varepsilon^{q+1}(R^q \circ g_q - R^q) + \varepsilon^{q+1}(R^q - R^q(\cdot, 0)). \end{aligned} \tag{66}$$

We first see that the the normal form term $\varepsilon^{q+1}\bar{R}^q$ is bounded in the $\|\cdot\|_{\mathcal{C}^\ell, L}$ norm by $\varepsilon^{q+1-\alpha(2q)}$, for $\ell = 0, \ldots, n - 2q$.

Indeed, using (38) and (39) from Lemma 3.3 we have:

i. If $\mathcal{B} \notin \bigcup_{-l/k \in \mathcal{R}_{[\leq q+1]}} \mathcal{I}_{-l/k}$, then

$$\bar{R}^q(\mathcal{B}, \phi, s) = R^q_{0,0}(\mathcal{B}; 0) \tag{67}$$

for any $(\mathcal{B},\phi,s)\in\mathcal{D}_{\rm nr}^{q+1}$ and, by formula (40) and the second part of hypothesis 2 for $i=q$ of Lemma 3.18, we have

$$\left\|\varepsilon^{q+1}\bar{R}^q\right\|_{\mathcal{C}^\ell,L}\leq\left\|\varepsilon^{q+1}R^q(\cdot;0)\right\|_{\mathcal{C}^\ell,L}\preceq\varepsilon^{q+1-\alpha(2q)},\tag{68}$$

for $\ell=0,\dots,r-2q$.

ii. If $\mathcal{B}\in\mathcal{I}_{-l_0/k_0}$, for some $-l_0/k_0\in\mathcal{R}_i\setminus\mathcal{R}_1\cup\cdots\cup\mathcal{R}_{i-1}$ for some $1\leq i\leq q+1$, then, by equation (39) in Lemma 3.3,

$$\bar{R}^q(\mathcal{B},\phi,s)=R^q_{0,0}(\mathcal{B};0)+\sum_{\substack{t\in\mathbb{Z}^2-\{0\},\\|t|(|k_0|+|l_0|)\leq M_q}}R^q_{tk_0,tl_0}(-l_0/k_0;0)e^{\mathrm{i}t\theta}\tag{69}$$

for any $(\mathcal{B},\phi,s)\in\mathcal{D}_{\mathrm{r},i}$, where $R^q_{k,l}(J;0)$ are the Fourier coefficients of the function $R^q(J,\varphi,s;0)$ with respect to (φ,s).

As before, by formula (40) from Lemma 3.3 and the second part of hypothesis 2 of this Lemma for $i=q$, we have

$$\left\|\varepsilon^{q+1}\bar{R}^q\right\|_{\mathcal{C}^\ell,L}\preceq\left\|\varepsilon^{q+1}R^q(\cdot;0)\right\|_{\mathcal{C}^\ell,L}\preceq\varepsilon^{q+1-\alpha(2q)}\tag{70}$$

for $\ell=0,\dots,r-2q$.

Note that, since $\alpha<1/n$ and $2q<n$, the power of ε in the bounds obtained in (68) and (70), is a positive number greater than q.

To finish the proof, we only need to estimate the remainder term $\varepsilon^{q+2}R^{q+1}$ in (66) and its Taylor expansion coefficients with respect to ε.

We will first estimate the the remainder term $\varepsilon^{q+2}R^{q+1}$ in (66). Since it is not a trigonometric polynomial we will estimate it in terms of the supremum norm $|\cdot|_{\mathcal{C}^\ell}$. Using the integral bound for the Taylor remainder and definitions (65) and (66) of $\bar{R}^q$ and $\varepsilon^{q+2}R^{q+1}$, respectively, we have

$$\begin{aligned}\left|\varepsilon^{q+2}R^{q+1}\right|_{\mathcal{C}^\ell}\leq&\int_0^1\left|\{Z^q-Z^0,\varepsilon^{q+1}G_q\}\circ g_{q,t}\right|_{\mathcal{C}^\ell}\mathrm{d}t\\&+\int_0^1\left|(1-t)(\{\{A+Z^0,\varepsilon^{q+1}G_q\},\varepsilon^{q+1}G_q\}\circ g_{q,t})\right|_{\mathcal{C}^\ell}\mathrm{d}t\\&+\int_0^1\left|\{\varepsilon^{q+1}R^q,\varepsilon^{q+1}G_q\}\circ g_{q,t}\right|_{\mathcal{C}^\ell}\mathrm{d}t\\&+\left|\varepsilon^{q+1}(R^q-R^q(\cdot;0))\right|_{\mathcal{C}^\ell}\\=&\int_0^1\left|\{Z^q-Z^0,\varepsilon^{q+1}G_q\}\circ g_{q,t}\right|_{\mathcal{C}^\ell}\mathrm{d}t\\&+\int_0^1\left|(1-t)\{\varepsilon^{q+1}(\bar{R}^q-R^q(\cdot;0)),\varepsilon^{q+1}G_q\}\circ g_{q,t}\right|_{\mathcal{C}^\ell}\mathrm{d}t\\&+\int_0^1\left|\{\varepsilon^{q+1}R^q,\varepsilon^{q+1}G_q\}\circ g_{q,t}\right|_{\mathcal{C}^\ell}\mathrm{d}t\\&+\left|\varepsilon^{q+1}(R^q-R^q(\cdot;0))\right|_{\mathcal{C}^\ell},\end{aligned}$$

for $\ell=0,\dots,n-2(q+1)$.

Using Faa-di Bruno formulae (C.4) we obtain

$$\begin{aligned}\left|\varepsilon^{q+2}R^{q+1}\right|_{\mathcal{C}^\ell} \preceq\ & \left|\{Z^q - Z^0, \varepsilon^{q+1}G_q\}\right|_{\mathcal{C}^\ell}\int_0^1 |g_{q,t}|^\ell_{\mathcal{C}^\ell}\,\mathrm{d}t \\ & + \left|\{\varepsilon^{q+1}(\bar{R}^q - R^q(\cdot;0)), \varepsilon^{q+1}G_q\}\right|_{\mathcal{C}^\ell}\int_0^1 (1-t)\,|g_{q,t}|^\ell_{\mathcal{C}^\ell}\,\mathrm{d}t \\ & + \left|\{\varepsilon^{q+1}R^q, \varepsilon^{q+1}G_q\}\right|_{\mathcal{C}^\ell}\int_0^1 |g_{q,t}|^\ell_{\mathcal{C}^\ell}\,\mathrm{d}t \\ & + \left|\varepsilon^{q+1}(R^q - R^q(\cdot;0))\right|_{\mathcal{C}^\ell},\end{aligned} \tag{71}$$

for $\ell = 0, \dots, n-2(q+1)$.

By formula (41) from Lemma 3.3, the second part of hypothesis 2 for $i = q$ of this Lemma, and using that $L \sim \varepsilon^\alpha$, we get that

$$\left\|\varepsilon^{q+1}G_q\right\|_{\mathcal{C}^\ell,L} \leq \frac{C}{L}\left\|\varepsilon^{q+1}R^q(\cdot;0)\right\|_{\mathcal{C}^{\ell+1},L} \preceq \varepsilon^{q+1-\alpha(2q+1)},$$

for $\ell = 0, \dots, r-2q-1$. Hence, using the equivalence relation (31) between $\|\cdot\|_{\mathcal{C}^\ell,L}$ and $|\cdot|_{\mathcal{C}^\ell}$ norms, $\varepsilon^{q+1}G_q$ satisfies

$$\left|\varepsilon^{q+1}G_q\right|_{\mathcal{C}^\ell} \preceq \varepsilon^{q+1-\alpha(\ell+2q+1)}, \tag{72}$$

for $\ell = 0, \dots, n-2q-1$, and the power of ε, $\eta_\ell = q+1-\alpha(\ell+2q+1) > q+1-\alpha n$ in equation (72) is positive. So, we can apply Lemma 3.15 with $G = \varepsilon^{q+1}G_q$ in $\mathcal{D} = (I_-, I_+) \times \mathbb{T}^2$, and we have that $g_{q,t} = \mathcal{O}_{\mathcal{C}^\ell}(1)$ and $g_{q,t} - \mathrm{Id} = \mathcal{O}_{\mathcal{C}^\ell}(\varepsilon^{q+1-\alpha(\ell+2(q+1))})$ for $t \in [0,1)$ and $\ell = 0, \dots, n-2(q+1)$.

In the expression (71), the terms $Z^q - Z^0$, G_q, $\bar{R}^q$ and $R^q(\cdot;0)$ are trigonometric polynomials in the variables (φ, s). In order to bound their corresponding Poisson brackets in the $|\cdot|_{\mathcal{C}^\ell}$ norm, we will first estimate their $\|\cdot\|_{\mathcal{C}^\ell,L}$ norm and apply Lemma 3.16. Finally, using the equivalence relation (31) between $|\cdot|_{\mathcal{C}^\ell}$ and $\|\cdot\|_{\mathcal{C}^\ell,L}$ norms, we will bound their corresponding Poisson bracket in the $|\cdot|_{\mathcal{C}^\ell}$ norm. On the other hand, the terms R^q and therefore $R^q - R^q(\cdot;0)$ are not trigonometric polynomials, so we can not use the $\|\cdot\|_{\mathcal{C}^\ell,L}$ norm. For this reason we will bound the $|\cdot|_{\mathcal{C}^\ell}$ norm for the Poisson brackets directly.

The terms $\varepsilon^{q+1}R^q(\cdot;0)$ and $\varepsilon^{q+1}\bar{R}^q$ in (71) are both bounded in the $\|\cdot\|_{\mathcal{C}^\ell,L}$ norm by $\varepsilon^{q+1-\alpha(2q)}$, for $\ell = 0, \dots, r-2q$, because of the second part of hypothesis 2 for $i = q$ and points (i) and (ii) already proved, respectively. Note that both terms are of type $\varepsilon^{q+1}F^q$, according to Remark 3.17.

The term $Z^q - Z^0 = \varepsilon\bar{R}^0 + \varepsilon^2\bar{R}^1 + \dots$, is a polynomial with respect to ε, so it can be bounded by its main term $\varepsilon\bar{R}^0$. Hence, using the bound for the term of order 1 ($p = 0$) of Z^q given in hypotheses 1b and 1c, we have

$$\left\|Z^q - Z^0\right\|_{\mathcal{C}^\ell,L} \preceq \left\|\varepsilon\bar{R}^0\right\|_{\mathcal{C}^\ell,L} \preceq \varepsilon, \tag{73}$$

for $\ell = 0, \dots, r-2(q-1)$. Note that $\varepsilon\bar{R}^0$ is of type εF^0, according to Remark 3.17.

The estimate for the $|\cdot|_{\mathcal{C}^\ell}$ norm of the term $(R^q - R^q(\cdot;0))$ can be obtained from the bound for the Taylor remainder and the first part of hypothesis 2. More precisely,

$$\left|\varepsilon^{q+1}(R^q - R^q(\cdot;0))\right|_{\mathcal{C}^\ell} \leq \varepsilon^{q+2}\left|R^q\right|_{\mathcal{C}^{\ell+1}} \preceq \varepsilon^{q+2-\alpha(\ell+1+2q)}, \tag{74}$$

for $\ell = 0, \ldots, n-2q-1$.

Moreover using the bounds for $\varepsilon^{q+1}R^q$ and $\varepsilon^{q+1}G_q$ in the $|\cdot|_{\mathcal{C}^\ell}$ norm, and Leibniz rule for derivatives we have

$$\begin{aligned}
&\left|\{\varepsilon^{q+1}R^q, \varepsilon^{q+1}G_q\}\right|_{\mathcal{C}^\ell} \\
&\preceq \sum_{i=0}^{\ell} \begin{pmatrix} \ell \\ i \end{pmatrix} \left(\left|\varepsilon^{q+1}\frac{\partial R^q}{\partial \varphi}\right|_{\mathcal{C}^i} \left|\varepsilon^{q+1}\frac{\partial G_q}{\partial J}\right|_{\mathcal{C}^{\ell-i}} + \left|\varepsilon^{q+1}\frac{\partial R^q}{\partial J}\right|_{\mathcal{C}^i} \left|\varepsilon^{q+1}\frac{\partial G_q}{\partial \varphi}\right|_{\mathcal{C}^{\ell-i}} \right) \\
&\preceq \sum_{i=0}^{\ell} \begin{pmatrix} \ell \\ i \end{pmatrix} \left|\varepsilon^{q+1}R^q\right|_{\mathcal{C}^{i+1}} \left|\varepsilon^{q+1}G_q\right|_{\mathcal{C}^{\ell-i+1}}.
\end{aligned}$$

Hence, using that $|R^0|_{\mathcal{C}^{\ell+1}} \preceq 1$ and $|G_0|_{\mathcal{C}^{\ell+1}} \preceq \varepsilon^{-\alpha(\ell+2)}$ from (72), we have

$$\left|\{\varepsilon R^0, \varepsilon G_0\}\right|_{\mathcal{C}^\ell} \preceq \varepsilon\varepsilon^{1-\alpha(\ell+2)} \preceq \varepsilon^{2-\alpha(\ell+2)},$$

otherwise,

$$\begin{aligned}
\left|\{\varepsilon^{q+1}R^q, \varepsilon^{q+1}G_q\}\right|_{\mathcal{C}^\ell} &\preceq \sum_{i=0}^{\ell} \begin{pmatrix} \ell \\ i \end{pmatrix} \varepsilon^{q+1-\alpha(i+1+2q)}\varepsilon^{q+1-\alpha(\ell-i+1+2q+1)} \\
&\preceq \varepsilon^{2(q+1)-\alpha(\ell+2(2q+1)+1)},
\end{aligned}$$

for $\ell = 0, \ldots, n-2(q-1)$.

Putting together in (71) the estimates in (72), (73) and (74), as well as the estimate for $\{\varepsilon^{q+1}R^q, \varepsilon^{q+1}G_q\}$ and $\varepsilon^{q+1}\bar{R}^q$ (these last two not relevant for $q \neq 0$), and using Lemma 3.16 and the equivalence relation (31) one gets the following bound for (66):

$$\left|\varepsilon^{q+2}R^{q+1}\right|_{\mathcal{C}^\ell} \preceq \varepsilon^{q+2-\alpha(\ell+2(q+1))},$$

for $\ell = 0, \ldots, n-2(q+1)$.

Finally, all the terms in the Taylor expansion of $\varepsilon^{q+2}R^{q+1}(\mathcal{B}, \phi, s, \varepsilon)$ with respect to ε, are obtained from a finite number of algebraic operations and a process of "$\varepsilon^{q+1}G_q$ Poisson bracketing", as stated in Remark 3.17, to the Taylor coefficients in ε of Z^q and of $\varepsilon^{q+1}R^q$, which are all of them of the form $\varepsilon^{p+1}F^p$. Applying Lemma 3.16 iteratively, we conclude that the Taylor expansion coefficient of order $i+1$ of $\varepsilon^{q+2}R^{q+1}(\mathcal{B}, \phi, s, \varepsilon)$ with respect to ε is of the type $\varepsilon^{i+1}F^i$ according to Remark 3.17, that is a trigonometric polynomial of order $M_i = (i+1)M$ in the angle variables, satisfying $\mathcal{O}_{\mathcal{C}^\ell,L}(\varepsilon^{i+1-\alpha(2i)})$ for $\ell = 0, \ldots, r-q-i$ and for $i = q, \ldots, r-q$. Again, by condition $\alpha < 1/n$, the power of ε is a positive number greater than i. □

Proof of Theorem 3.11

The proof is by induction in q. To begin induction process, we consider $R^{[\leq M]}$, which is the truncated Fourier series of the perturbation R up to some order $M_0 = M$ as in (25). The order of truncation M is $M \sim \varepsilon^{-\rho}$, with ρ satisfying hypothesis (51). We want to apply Lemma 3.18 for $q = 0$ to the Hamiltonian

$$k_0(J, \varphi, s; \varepsilon) = Z^0(J, \varphi, s; \varepsilon) + \varepsilon R^0(J, \varphi, s; \varepsilon),$$

where $Z^0(J, \varphi, s; \varepsilon) = J^2/2$ and $R^0(J, \varphi, s; \varepsilon) = R^{[\leq M]}(J, \varphi, s; \varepsilon)$.

We introduce the finite set

$$\mathcal{R}_1 = \{-l/k \in \mathbb{Q} \cap (I_-, I_+), |k| + |l| \leq M, k \neq 0, R^0_{k,l}(-l/k; 0) \neq 0\},$$

where $R^0_{k,l}$ are the Fourier coefficients of R^0. According to Definition 3.4 we will refer to resonances of order 1 the elements of the set $\mathcal{R}_1$.

Since $Z^0 = J^2/2$ and R^0 satisfy trivially hypothesis 1 and 2 of Lemma 3.18 and hypothesis (55) holds, we can apply Lemma 3.18 for $q = 0$, which provides a symplectic change of variables $(\mathcal{B}, \phi, s) \mapsto (J, \varphi, s) = g_0(\mathcal{B}, \phi, s)$ of class $\mathcal{C}^{r-2}$ and we get a Hamiltonian of the form

$$k_1(J, \varphi, s; \varepsilon) = Z^1(J, \varphi, s; \varepsilon) + \varepsilon^2 R^1(J, \varphi, s; \varepsilon),$$

where Z^1 is a $\mathcal{C}^r$ function and $\varepsilon^2 R^1$ is a $\mathcal{C}^{r-2}$ function, verifying properties 1b,1c and 2 of Lemma 3.18 with $q = 1$.

In particular, in the resonant regions of order 1 $\mathcal{D}_{r,1}$ defined in (49), expression (64) in Lemma 3.18 for $q = 0$ provides that Z^1 has the form (58) for $i = m = 1$, that is

$$Z^1(\mathcal{B}, \phi, s; \varepsilon) = \frac{1}{2}\mathcal{B}^2 + \varepsilon \widetilde{Z}^1(\mathcal{B}) + \varepsilon U_1^{k_0,l_0}(k_0\phi + l_0 s; \varepsilon),$$

where $U_1^{k_0,l_0}$ is given by expression (59).

Proceeding by induction, we assume that we have applied Lemma 3.18 up to order q, for $0 < q < m$, so that in adequate symplectic coordinates, the Hamiltonian k_q of this Theorem takes the form

$$k_q(J, \varphi, s; \varepsilon) = Z^q(J, \varphi, s; \varepsilon) + \varepsilon^{q+1} R^q(J, \varphi, s; \varepsilon),$$

and satisfies hypotheses 1 and 2 of Lemma 3.18, so that it can be applied again to the Hamiltonian k_q, providing a Hamiltonian

$$k_{q+1}(J, \varphi, s; \varepsilon) = Z^{q+1}(J, \varphi, s; \varepsilon) + \varepsilon^{q+2} R^q(J, \varphi, s; \varepsilon)$$

satisfying properties 1 and 2 of Lemma 3.18 replacing q by $q + 1$ and a new constant $C = C_q$ in $L = C\varepsilon^\alpha$ to accommodate new resonances.

Applying the inductive Lemma m times, we get a Hamiltonian k_m

$$k_m(J, \varphi, s; \varepsilon) = Z^m(J, \varphi, s; \varepsilon) + \varepsilon^{m+1} R^m(J, \varphi, s; \varepsilon),$$

that consists of an integrable Hamiltonian Z^m, which already satisfies thesis 1 and 2 of Theorem 3.11 for $\bar{Z}^m = Z^m$, plus a perturbation $\varepsilon^{m+1} R^m$ of order $\mathcal{O}_{\mathcal{C}^\ell}(\varepsilon^{m+1-\alpha(\ell+2m)})$, $0 \leq \ell \leq n - 2m$.

Moreover, Lemma 3.18 gives us estimates for the terms of the integrable part $\bar{Z}^m$ of the Hamiltonian k_m in the Fourier weighted norm $\|\cdot\|_{\mathcal{C}^\ell, L}$ defined in (30). More precisely, we know that $\bar{Z}^m$ is a polynomial of degree m in ε, whose term of order $q+1$ is $\mathcal{O}_{\mathcal{C}^\ell, L}(\varepsilon^{q+1-\alpha(2q)})$, for $\ell = 0, \ldots, r - 2q$ and $q = 0, \ldots, m - 1$. By the equivalence relation (31) we immediately also have that this term of order $q+1$ is $\mathcal{O}_{\mathcal{C}^\ell}(\varepsilon^{q+1-\alpha(\ell+2q)})$, for $\ell = 0, \ldots, n - 2q$ and $q = 0, \ldots, m - 1$.

It remains to prove the estimates of thesis 3 of Theorem 3.11 on $\bar{Z}^m$ and $\bar{R}^m$ in the supremum norm $|\cdot|_{\mathcal{C}^\ell}$.

The estimation for $\widetilde{Z}^m$ follows from the ones obtained for $\bar{Z}^m$ and we will concentrate on the ones for $U_m^{k_0,l_0}$.

In particular, in $\mathcal{D}_{r,i}$, we can obtain a better estimate for the $|\cdot|_{\mathcal{C}^\ell}$ norm of the term $\varepsilon^i U_m^{k_0,l_0}(\theta;\varepsilon)$ in expression (64), which is the one claimed in point 3 of the Theorem. In order to check this, we first notice that the function $U_m^{k_0,l_0}(\theta;\varepsilon)$ in expression (58) is a polynomial of degree $m-i$ in ε and a trigonometric polynomial in $\theta = k_0\phi + l_0 s$. So, $\varepsilon^i U_m^{k_0,l_0}(\theta;\varepsilon)$ can be bounded by its main term $\varepsilon^i U_m^{k_0,l_0}(\theta;0)$, which is a trigonometric polynomial in the variable $\theta = k_0\phi + l_0 s$ and independent of the action variable $\mathcal{B}$. Using that

$$\left\|\varepsilon^i U_m^{k_0,l_0}(\cdot;0)\right\|_{\mathcal{C}^{r-2(i-1)},L} = \preceq \varepsilon^{i-\alpha(2(i-1))},$$

and the definition of the Fourier weighted norm in (30), we have

$$\left\|\varepsilon^i U_m^{k_0,l_0}(\cdot;0)\right\|_{\mathcal{C}^{r-2(i-1)},L} = \varepsilon^i \sum_{\substack{t\in\mathbb{Z}-\{0\}\\ |t|(|k_0|+|l_0|)\le M_q}} |U_{tk_0,tl_0}|_{\mathcal{C}^0} |t(k_0,l_0)|^{r-2(i-1)} \preceq \varepsilon^{i-\alpha(2(i-1))},$$

where $U_{k,l}$ are the Fourier coefficients of the function $U_m^{k_0,l_0}(\theta;0)$, $M_q = (q+1)M$ and $|(k,l)| = \max(|k|,|l|)$. From this expression it is clear that

$$|U_{tk_0,tl_0}|_{\mathcal{C}^0} \le C\varepsilon^{i-\alpha(2(i-1))}/|t(k_0,l_0)|^{r-2(i-1)},$$

for some constant C independent of ε. Hence, bounding derivatives with respect to the variable θ we have

$$\begin{aligned}
|U_m^{k_0,l_0}(\cdot;\varepsilon)|_{\mathcal{C}^\ell,\theta} &\preceq \sum_{\substack{t\in\mathbb{Z}-\{0\}\\ |t|(|k_0|+|l_0|)\le M_q}} |U_{tk_0,tl_0}|_{\mathcal{C}^0}\, |t|^\ell \\
&\preceq \sum_{\substack{t\in\mathbb{Z}-\{0\}\\ |t|(|k_0|+|l_0|)\le M_q}} \frac{\varepsilon^{i-\alpha(2(i-1))}}{|t(k_0,l_0)|^{r-2(i-1)}} |t|^\ell \\
&\preceq \frac{\varepsilon^{i-\alpha(2(i-1))}}{|(k_0,l_0)|^{r-2(i-1)}} \sum_{\substack{t\in\mathbb{Z}-\{0\}\\ |t|(|k_0|+|l_0|)\le M_q}} \frac{1}{|t|^{r-2(i-1)-\ell}} \\
&\preceq \frac{\varepsilon^{i-\alpha(2(i-1))}}{|(k_0,l_0)|^{r-2(i-1)}},
\end{aligned}$$

for $\ell = 0,\ldots,n-2(i-1)$, as claimed in point 3 of Theorem 3.11.

Finally, it remains to prove that the tail $\varepsilon R^{[>M]}$ of the Fourier series of the perturbation εR that we have truncated at order $M \sim \varepsilon^{-\rho}$ at the beginning of this proof is $\mathcal{O}_{\mathcal{C}^\ell}(\varepsilon^{m+1-\alpha(\ell+2m)})$, for $0 \le \ell \le n-2m$. Since the perturbation R in Hamiltonian (53) of Theorem 3.11 is a $\mathcal{O}_{\mathcal{C}^r}(1)$ function, the Fourier coefficients $R_{k,l}(J,\varepsilon)$ of $R(J,\varphi,s,\varepsilon)$ decrease at a rate of order $1/|(k,l)|^r$, for $(k,l) \longrightarrow \infty$. So, by equation (A.2) in Proposition A.2 we have the following bound for $\varepsilon R^{[>M]}$,

$$|\varepsilon R^{[>M]}|_{\mathcal{C}^\ell} \preceq \frac{\varepsilon}{M^{r-\ell-2}} \preceq \varepsilon^{1+\rho(r-\ell-2)}, \tag{75}$$

for $\ell = 0,\ldots,n-2m$.

From Lemma 3.18 and equation (61), we know that the changes of coordinates g_q satisfy, for $q = 0, \dots, m-1$, $g_q = \mathcal{O}_\ell(1)$ and $g_q - \mathrm{Id} = \mathcal{O}_\ell(\varepsilon^{q+1-\alpha(\ell+2(q+1))})$, for $\ell = 0, \dots, n-2(q+1)$. Therefore, the total change of coordinates of Theorem 3.11 $(J, \varphi, s) = g(\mathcal{B}, \phi, s)$ where $g = g_{m-1} \circ \cdots \circ g_0$, satisfies $g = \mathcal{O}_\ell(1)$ and $g - \mathrm{Id} = \mathcal{O}_\ell(\varepsilon^{1-\alpha(\ell+2)})$, for $\ell = 0, \dots, n-2m$. Then, using this fact and formula (75), by Faadi-Bruno formula (C.4) we have

$$|R^{[>M]} \circ g|_{\mathcal{C}^\ell} \preceq \varepsilon^{1+\rho(r-\ell-2)}.$$

To get $|\varepsilon R^{[>M]} \circ g|_{\mathcal{C}^\ell} \preceq \varepsilon^{m+1-\alpha(\ell+2m)}$, we need $\varepsilon^{1+\rho(r-\ell-2)} \leq \varepsilon^{m+1-\alpha(\ell+2m)}$, that is

$$\rho \geq \frac{m - \alpha(\ell+2m)}{(r-\ell-2)}, \tag{76}$$

for $\ell = 0, \dots, n-2m$. In order that bounds (55) and (76) were compatible, we need to choose $r \geq \left(\frac{1}{\rho} - 2\right) m + 2$, which is condition (52) in the hypotheses of this Theorem.

Finally the choice $\bar{Z} = Z^m$ and $\bar{R} = R^m + R^{[>M]} \circ g$, with $g = g_m \circ \cdots \circ g_0$, gives the desired averaged Hamiltonian (56) which satisfies theses 1,2 and 3. □

Remark 3.20. Choosing $\rho = 1/(n+\delta)$, with $0 < \delta < 1/m$, so that condition (55) is fulfilled for any α between ρ and $1/n$, we have that r must satisfy

$$r \geq (n - 2 + \delta)m + 2,$$

where m is the number of steps of averaging performed. So, as long as the regularity r of the Hamiltonian satisfies

$$r > r_{\min} := (n-2)m + 2, \tag{77}$$

there exist ρ, α satisfying condition (55) and therefore (51) of Theorem 3.11 and henceforth, m steps of averaging can be performed to provide estimates of class $\mathcal{C}^{n-2m}$, contained in the theses of Theorem 3.11.

Remark 3.21. It is important to note that the averaging procedure is valid in the full domain $(I_-, I_+) \times \mathbb{T}^2 \subset \widetilde{\Lambda}_\varepsilon$. Indeed, we have performed an averaging procedure to the Hamiltonian $k(J, \varphi, s; \varepsilon)$ in all $(I_-, I_+) \times \mathbb{T}^2$, except at the subsets $\mathcal{D}_{\mathrm{t}}(L)$, where

$$\mathcal{D}_{\mathrm{t}}(L) = \{(J, \varphi, s) \in (I_-, I_+) \times \mathbb{T}^2; L_k \leq |J + l/k| \leq 2L_k, \text{ for } -l/k \in \mathcal{R}_{[\leq m]}\}.$$

To provide an averaging procedure in the full domain $(I_-, I_+) \times \mathbb{T}^2$, we apply again Theorem 3.11 with $\widetilde{L}_k = \widetilde{L}/|k|$, where $\widetilde{L} = L/2$. The region $\mathcal{D}_{\mathrm{t}}(L)$ is now contained in the non resonant region corresponding to $\widetilde{L}_k$, $\mathcal{D}^m_{\mathrm{nr}}(M, \widetilde{L})$ defined in Definition 3.7. So the averaged Hamiltonian in $\mathcal{D}_{\mathrm{t}}$ is also given by Theorem 3.11, with slightly different constants.

3.3. KAM theorem

Up to this point, once we choose m, by Theorem 3.11 we can perform m steps of averaging and we obtain a $\mathcal{C}^{r-2m}$ Hamiltonian (56) which consists of an integrable Hamiltonian $\bar{Z}^m$ plus a perturbation $\varepsilon^{m+1}\bar{R}^m$ which is $\mathcal{C}^{n-2m}$ small, more precisely it is $\mathcal{O}_{\mathcal{C}^\ell}(\varepsilon^{m+1-\alpha(\ell+2m)})$, for $\ell = 0, \ldots, n-2m$. Notice that $n \geq 2m$ is required as well as a large r and that the integrable Hamiltonian has different expressions in resonant regions and non-resonant regions as specified in Theorem 3.11.

The integrable part of the Hamiltonian gives us an approximate equation $\bar{Z}^m = \text{cte}$ for the invariant tori in $\widetilde{\Lambda}_\varepsilon$. To finish the proof of Theorem 3.1 it remains to determine which tori survive and what is the distance between them when we add the perturbation term $\varepsilon^{m+1}\bar{R}^m$. By choosing an adequate m large enough the goal is to show that we can cover the whole region $(I_-, I_+) \times \mathbb{T}^2 \subset \widetilde{\Lambda}_\varepsilon$ with invariant tori which are $\mathcal{O}(\varepsilon^{1+\eta})$-closely spaced, for some $\eta > 0$, and obtain an approximate expression for them.

To that end, we will use KAM Theorem 3.22 stated in Section 3.3.1, which is a result about the existence of invariant tori for a periodic perturbation of a Hamiltonian expressed in action-angle variables. It is a direct adaptation of Theorem 8.12 in [DLS06a].

Since the integrable Hamiltonian (56) has different expressions in the resonant and non-resonant regions, we perform this study separately.

Non-resonant regions are studied in Section 3.3.2. In Proposition 3.24, we apply Theorem 3.22 directly to Hamiltonian (56) for $m \geq 2$ and we conclude that for these regions there exist primary KAM tori which are $\mathcal{O}(\varepsilon^{1+\eta})$-closely spaced, for some $\eta > 0$.

Resonant regions are studied in Section 3.3.3. As it has been described in Section 2.3.3, we will see that for these regions, gaps of different sizes are created in the foliation of primary KAM tori. According to the size of the gaps, we will distinguish two types of resonant regions: the resonant regions with big gaps, where gaps are of size greater or equal than ε, which is the size of the heteroclinic jumps provided by the scattering map, and the resonant regions with small gaps, where gaps are of size smaller than ε.

In the referred Section 3.3.3, we will see that the resonant regions with big gaps introduced in (21) correspond to the resonances $J = -l/k$ of order 1 such that $|(k,l)| < M_{\mathrm{BG}} = \varepsilon^{-(1+\nu)/r}$, for $1/(r/6-1) < \nu \leq 1/16$, whereas resonant regions with small gaps correspond to the rest of the resonances.

The case of resonant regions with small gaps is studied in Section 3.3.4. It will not be different from the non-resonant case and it will be enough to apply KAM Theorem 3.22 to Hamiltonian (56) for $m \geq 2$ to obtain primary tori $\mathcal{O}(\varepsilon^{1+\eta})$-closely spaced, for some $\eta > 0$. This is done in Proposition 3.26. Resonant regions with small gaps constitute, jointly with the non resonant regions, what we call the flat tori region introduced in Section 2.3.3.

The case of resonant regions with big gaps is significantly different and it will be studied in Section 3.3.5. In this case the integrable Hamiltonian $\bar{Z}^m$ is like a pendulum, and we will need to write it first in action-angle variables before applying KAM Theorem 3.22 to Hamiltonian (56) for $m \geq 10$. We will see that in these regions we can find other invariant objects, the secondary tori, which fill the region inside the gaps and they get rather close to the frontier of the gaps among the primary KAM tori. The precise result, jointly with the approximate equations for the invariant tori is given in Proposition 3.28.

Finally, Theorem 3.1 follows directly from Propositions 3.24, 3.26 and Theorem

3.28.

3.3.1. The KAM Theorem The following result is about the existence of invariant tori for a periodic perturbation $2\pi k_0$-periodic in the variable φ and 2π-periodic in the variable s, of a Hamiltonian system expressed in action-angle variables and it is standard in KAM theory (see [Lla01] for a tutorial on this theory). We skip its proof since it is simply an adaptation of Theorem 8.12 in [DLS06a], where the explicit dependence of the constants on k_0 is given, since k_0 will be chosen depending on ε. It relies on a quantitative KAM Theorem of Herman [Her83, Theorem 5.4, p. 198] for exact symplectic mappings of the annulus.

Theorem 3.22. *Let $K(I,\varphi,s;\varepsilon)$ be Hamiltonian of the form*

$$K(I,\varphi,s;\varepsilon) = K_0(I;\varepsilon) + K_1(I,\varphi,s;\varepsilon), \tag{78}$$

for $(I,\varphi,s) \in \mathcal{I} \times (\mathbb{R}/2\pi k_0\mathbb{Z}) \times \mathbb{T}$, for some $k_0 \in \mathbb{N}$. Assume that

i. K is a $\mathcal{C}^{n_0+\beta}$ function of the variables (I,φ,s), with $n_0 \geq 5$ and $0 < \beta < 1$,

ii. For any $s \in \mathbb{T}$, $|K_1(\cdot,s;\varepsilon)|_{\mathcal{C}^{n_0+\beta}} \leq \delta$ and $\left|K_0''(\cdot;\varepsilon)\right|_{\mathcal{C}^0} \geq M > 0$, where $\delta = \delta(\varepsilon)$ and $M = M(\varepsilon)$ depend on ε.

Then, for ε sufficiently small and fixed, there exists a constant $C(k_0) =$ cte $|k_0|^{(n_0+\beta)/2}$ and a finite set of values $I_i \in \mathcal{I}$, such that the Hamiltonian $K(I,\varphi,s;\varepsilon)$ has invariant tori $\mathcal{T}_i$, such that:

a. The torus $\mathcal{T}_i$ can be written as a graph of the variable I over the angle variables (φ,s):

$$\mathcal{T}_i = \{(I,\varphi,s) \in \mathcal{I} \times \mathbb{T}^2 : I = I_i + \Psi_i(\varphi,s;\varepsilon)\},$$

where $\Psi_i(\varphi,s;\varepsilon)$ is a $\mathcal{C}^{n_0-2+\beta}$ function and $|\Psi_i(\cdot;\varepsilon)|_{\mathcal{C}^{n_0-2+\beta}} \leq C(k_0)M^{-1}\delta^{1/2}$.

b. The motion on the torus is $\mathcal{C}^{n_0-4+\beta}$ conjugate to a rigid translation of frequency $(\omega(I_i),1)$, where $\omega(I_i)$ is a Diophantine number of constant type and Markov constant $\kappa = C(k_0)\delta^{1/2}$, *that is*

$$|\omega(I_i)k - l|^{-1} \leq C\kappa^{-1}|(k,l)| \quad \forall (k,l) \in \mathbb{Z}^2 \setminus \{(0,0)\}.$$

c. The union of neighborhoods of size $C(k_0)M^{-1}\delta^{1/2}$ of these tori cover all the region $\mathcal{I} \times (\mathbb{R}/2\pi k\mathbb{Z}) \times \mathbb{T}$.

Remark 3.23. This version of KAM Theorem requires to have the system written in action angle variables. We would like to mention that recently there have appeared some quantitative results on KAM theory without action angle variables (see [LGJV05] and [FLS07]) for analytic maps, which could be adapted but some extra work is required.

3.3.2. Non-resonant region In this section we apply directly Theorem 3.22 to the averaged Hamiltonian (56) in the non-resonant region up to order m $\mathcal{D}_{\rm nr}^m$ introduced in (48). According to Remark 3.21, we use $L/2$ instead of L, so that

$$\mathcal{D}_{\rm nr}^m = \mathcal{D}_{\rm nr}^m(M, L/2) = \{(J, \varphi, s) \in (I_-, I_+) \times \mathbb{T}^2 : |J + l/k| \geq L_k, \text{for} - l/k \in \mathcal{R}_{[\leq m]}\},$$

where $L_k = L/|k|$, with $L = C\varepsilon^\alpha$ and $\alpha < 1/n$, as required in Theorem 3.11.

Going back to the original variables (I, φ, s), using the changes given by Theorem 3.11 and equation (22), we obtain the following result about the existence of invariant tori of Hamiltonian (1):

Proposition 3.24 (Invariant tori in the non-resonant region)**.** *Assume that $m \geq 2$, $n \geq 2m+6$ and $r > (n-2)m+2$. Choose any $0 < \eta \leq (m-1-\alpha n)/2$, where $\alpha < 1/n$ as required in Theorem 3.11. Then, for ε small enough, in any connected component of the non resonant region up to order m $\mathcal{D}_{\rm nr}^m$, there exists a finite set of values E_i such that:*

i. For any E_i there exists a torus $\mathcal{T}_i$ invariant by the flow of Hamiltonian (1) contained in $\widetilde{\Lambda}_\varepsilon$, which is given in $\widetilde{\Lambda}_\varepsilon$ by the equation $F(I, \varphi, s; \varepsilon) \equiv E_i$, where F is a $\mathcal{C}^{n-2m-2-\varrho}$ function F, for any $\varrho > 0$, of the form

$$F(I, \varphi, s; \varepsilon) = I + \mathcal{O}_{\mathcal{C}^2}(\varepsilon^{1+\eta}). \tag{79}$$

ii. The torus $\mathcal{T}_i$ contained in $\widetilde{\Lambda}_\varepsilon$ can also be written as a graph of the variable I over the angle variables (φ, s):

$$\mathcal{T}_i = \{(I, \varphi, s) \in \mathcal{D}_{\rm nr}^m, I = \lambda_{E_i}(\varphi, s; \varepsilon)\},$$

with

$$\lambda_E(\varphi, s; \varepsilon) = E + U_E(\varphi, s; \varepsilon); \tag{80}$$

where $U_E(\varphi, s; \varepsilon)$ is a $\mathcal{C}^{n-2m-2-\varrho}$ function, for any $\varrho > 0$, and $U_E = \mathcal{O}_{\mathcal{C}^2}(\varepsilon^{1+\eta})$.

iii. These tori are $\mathcal{O}(\varepsilon^{1+\eta})$-closely spaced in terms of the variable I.

Proof: By equations (56) and (57) in Theorem 3.11, in one connected component of the non-resonant region $\mathcal{D}_{\rm nr}^m$, the Hamiltonian (23) expressed in the averaged variables $(\mathcal{B}, \phi, s)$ has the following expression

$$k_m(\mathcal{B}, \phi, s; \varepsilon) = \frac{\mathcal{B}^2}{2} + \varepsilon \widetilde{Z}^m(\mathcal{B}, \varepsilon) + \varepsilon^{m+1} \bar{R}^m(\mathcal{B}, \phi, s; \varepsilon), \tag{81}$$

where $\varepsilon \widetilde{Z}^m(\mathcal{B}; \varepsilon)$ is a polynomial of degree m in ε, whose coefficient in terms of ε of order $q+1$, for $q = 0, \dots, m-1$, is a $\mathcal{C}^{r-2q}$ function and is $\mathcal{O}_{\mathcal{C}^\ell}(\varepsilon^{q+1-\alpha(\ell+2q)})$, for $\ell = 0, \dots, n-2q$. Moreover, $\varepsilon^{m+1}\bar{R}^m(\mathcal{B}, \phi, s; \varepsilon)$ is a $\mathcal{C}^{r-2m}$ function, which is $\mathcal{O}_{\mathcal{C}^\ell}(\varepsilon^{m+1-\alpha(\ell+2m)})$ for $\ell = 0, \dots, n-2m$.

Our next step is to apply KAM Theorem 3.22 to the Hamiltonian (81), which is of the form (78), for $K_0 = \mathcal{B}^2/2 + \varepsilon \widetilde{Z}^m(\mathcal{B}, \varepsilon)$ and $K_1 = \varepsilon^{m+1}\bar{R}^m(\mathcal{B}, \phi, s; \varepsilon)$ and 2π-periodic in φ and s, so that $k_0 = 1$. Assuming that $n \geq 2m+6$, it satisfies properties (i) and (ii) with $n_0 = n-2m-1$, $\beta = 1-\varrho$, for any $\varrho > 0$, $\delta = \varepsilon^{m+1-\alpha n}$ and M independent of

ε. Therefore we can apply KAM Theorem 3.22 and we conclude that the non-resonant region $\mathcal{D}^m_{\mathrm{nr}}$ contains KAM tori given by

$$\mathcal{B} = \mathcal{B}_i + \Psi_i(\phi, s; \varepsilon),$$

where Ψ_i is a $\mathcal{C}^{n-2m-2-\varrho}$ function, for any $\varrho > 0$, and $|\Psi_i|_{\mathcal{C}^2} \preceq \varepsilon^{(m+1-\alpha n)/2}$. These tori are $\mathcal{O}(\varepsilon^{(m+1-\alpha n)/2})$-closely spaced in terms of the averaged variable $\mathcal{B}$.

For a fixed value of $\varepsilon \ll 1$, we have that $\varepsilon^{(m+1-\alpha n)/2} \leq \varepsilon^{1+\eta}$, where $\eta = 1/2(m-1-\alpha n)$ is positive by hypotheses $m \geq 2$ and $\alpha < 1/n$ for $n \geq 2m+6$.

After applying KAM Theorem to Hamiltonian (81), we can go back to the original variables (I, φ, s). Using that the change $(J, \varphi, s) \mapsto (\mathcal{B}, \phi, s)$ is $\varepsilon^{1-\alpha(\ell+2)}$-close to the identity in the $\mathcal{C}^\ell$ sense for $\ell = 0, \ldots, n-2m$ by Theorem 3.11 and $(I, \varphi, s) \mapsto (J, \varphi, s)$ is ε-close to the identity in the $\mathcal{C}^{r-1}$ sense by equation (22), the invariant tori obtained in the region $\mathcal{D}^m_{\mathrm{nr}}$ are given by

$$I = I_i + U_i(\varphi, s; \varepsilon)$$

where the function U_i verifies the same properties as Ψ_i, and they are $\mathcal{O}(\varepsilon^{1+\eta})$-closely spaced in terms of the variable I. We get the results claimed for $E_i = I_i$. $\square$

3.3.3. Resonant region In this section, we analyze Hamiltonian (23) in the resonant region up to order m $\mathcal{D}^m_{\mathrm{r}}$ defined in (50).

We will perform an accurate study in this resonant region $\mathcal{D}^m_{\mathrm{r}}$ and we will estimate the size of the gaps created in the foliation of primary KAM tori. We will see that this size depends on the order j of the resonance, for $1 \leq j \leq m$, and on the size of the harmonic associated to the corresponding resonance. According to this, we define two types of regions: the small gaps regions $\mathcal{D}_{\mathrm{SG}}$ where the size of the gap is smaller than ε and the big gaps regions $\mathcal{D}_{\mathrm{BG}}$ where the size of the gap is bigger or equal than ε.

We will work in one connected component of the resonant domain $\mathcal{D}^m_{\mathrm{r}}$ which, according to (50), is of the form

$$\{(J, \varphi, s) \in [-l_0/k_0 - L_{k_0}, -l_0/k_0 + L_{k_0}] \times \mathbb{T}^2\}, \tag{82}$$

for some $-l_0/k_0 \in \mathcal{R}_j \setminus (\mathcal{R}_1 \cup \cdots \cup \mathcal{R}_{j-1})$, for $1 \leq j \leq m$, where $L_{k_0} = L/|k_0|$, with $L = C\varepsilon^\alpha$ and $\alpha < 1/n$, as required in Theorem 3.11.

By formulas (56) and (58) of Theorem 3.11, in component (82), Hamiltonian (23) expressed in the averaged variables $(\mathcal{B}, \phi, s)$, can be written as

$$\begin{aligned} k_m(\mathcal{B}, \phi, s; \varepsilon) &= \frac{1}{2}\mathcal{B}^2 + \varepsilon \widetilde{Z}^m(\mathcal{B}; \varepsilon) + \varepsilon^j U^{k_0,l_0}_m(k_0\phi + l_0 s; \varepsilon) + \varepsilon^{m+1}\bar{R}^m(\mathcal{B}, \phi, s; \varepsilon), \\ &:= \bar{Z}^m(\mathcal{B}, \phi, s; \varepsilon) + \varepsilon^{m+1}\bar{R}^m(\mathcal{B}, \phi, s; \varepsilon), \end{aligned} \tag{83}$$

where $\widetilde{Z}^m(\mathcal{B}; \varepsilon)$ and $U^{k_0,l_0}_m(k_0\phi + l_0 s; \varepsilon)$ are polynomials of degree $m-1$ and $m-j$ in ε, respectively, and $U^{k_0,l_0}_m(k_0\phi + l_0 s; \varepsilon)$ is a trigonometric polynomial in $\theta = k_0\phi + l_0 s$.

For $q = 0, \ldots, m-1$, the coefficient of order $q+1$ in ε of $\varepsilon\widetilde{Z}^m$ is a $\mathcal{C}^{r-2q}$ function which is $\mathcal{O}_{\mathcal{C}^\ell}(\varepsilon^{q+1-\alpha(\ell+2q)})$ for $\ell = 0, \ldots, n-2q$. The function $\varepsilon^j U^{k_0,l_0}_m(\theta; \varepsilon)$, for $\theta = k_0\phi + l_0 s$, satisfies

$$\left|\varepsilon^j U^{k_0,l_0}_m(\cdot; \varepsilon)\right|_{\mathcal{C}^\ell} \preceq \varepsilon^{j-2\alpha(j-1)}|(k_0, l_0)|^{-r+2(j-1)}, \tag{84}$$

for $\ell = 0, \ldots, n-2(j-1)$ and $|(k_0, l_0)| = \max(|k_0|, |l_0|)$.

Moreover, $\varepsilon^{m+1}\bar{R}^m$ is a $\mathcal{C}^{r-2m}$ function which is $\mathcal{O}_{\mathcal{C}^\ell}(\varepsilon^{m+1-\alpha(\ell+2m)})$, for $\ell = 0 \ldots n - 2m$.

From expression (83) it is clear that the integrable part $\bar{Z}^m$ is like a pendulum. The integrable pendulum has rotational and librational orbits as well as separatrices, which separate these two types of motion. It is straightforward to see that the size of the gap, created by the separatrix loop, associated to the resonance $-l_0/k_0 \in \mathcal{R}_j \setminus \mathcal{R}_1 \cup \cdots \cup \mathcal{R}_{j-1}$, in terms of the J variables, can be bounded from above by $\sqrt{2}\varepsilon^{j/2}\left|U_m^{k_0,l_0}(\cdot;\varepsilon)\right|_{\mathcal{C}^0}^{1/2}$.

From expression (84), we have that the size of the gap for a resonance $-l_0/k_0$ of order j is

$$\mathcal{O}(\varepsilon^{(j-2\alpha(j-1))/2}|(k_0, l_0)|^{(-r+2(j-1))/2}). \tag{85}$$

Expression (85) shows that the gaps form a heterogeneous sea since their size depends on the order $j \geq 1$ of the resonance and the size of the harmonic $|(k_0, l_0)|$. Among them, the biggest gaps are those of order $j = 1$ and harmonic $|(k_0, l_0)| \leq M_{\mathrm{BG}}$, where $M_{\mathrm{BG}} = \varepsilon^{-(1+\nu)/r}$ was introduced in Theorem 2.1 and satisfies $M_{\mathrm{BG}} > M$, where M is the order of truncation. Indeed, in the particular case of a resonance $-l_0/k_0$ of order 1 ($j = 1$), the size of the gap is

$$\mathcal{O}(\varepsilon^{1/2}|(k_0, l_0)|^{-r/2}), \tag{86}$$

so that for any $\nu > 0$, the resonances of order 1 such that $|(k_0, l_0)| \geq M_{\mathrm{BG}} = \varepsilon^{-(1+\nu)/r}$, create gaps of size $\mathcal{O}(\varepsilon^{1+\nu/2})$, that is, smaller than ε.

On the other hand, we know that resonances $-l_0/k_0$ of order greater than 1 satisfy $M_{\mathrm{BG}} \leq |(k_0, l_0)| \leq mM$ (see Remark 3.5). Hence, according to (85) the size of the gap created by a resonance $-l_0/k_0$ or order j, for $j = 2, \ldots, m$ is

$$\mathcal{O}(\varepsilon^{(j+1+\nu-(\alpha+(1+\nu)/r)2(j-1))/2})).$$

Using the condition $\alpha < 1/n$, with $r > n \geq 2m$ and $m \geq 2$, the size of the gap is

$$\mathcal{O}(\varepsilon^{(j+1-4\alpha(j-1))/2}).$$

For $j \geq 2$ the size of the gaps is smaller than $\varepsilon^{1+\eta}$, for $\eta = (1-4\alpha)/2$. Notice that $\eta > 0$ thanks to the condition on α.

As we already said, we will distinguish between two types of resonant regions depending whether the size of the gaps created in the foliation of primary KAM tori are bigger or smaller than the size ε of the heteroclinic jumps provided by the scattering map (17).

- **Resonant regions with big gaps $\mathcal{D}_{\mathrm{BG}}$.** Gaps of size of order equal or greater than ε are created in the foliation of primary invariant tori. According to (86) they correspond to resonances $-l_0/k_0$ of order 1 with $\gcd(k_0, l_0) = 1$, satisfying $|(k_0, l_0)| < M_{\mathrm{BG}}$, where $M_{\mathrm{BG}} = \varepsilon^{-(1+\nu)/r}$, for $1/(r/6-1) < \nu \leq 1/16$. See definition (21).

- **Resonant regions with small gaps $\mathcal{D}_{\mathrm{SG}}$.** Gaps between primary tori are smaller than ε. They correspond to the resonant regions of resonances $-l_0/k_0$ of order 1 such that $|(k_0,l_0)| \geq M_{\mathrm{BG}}$, and resonances of order greater or equal than 2 (which also satisfy $|(k_0,l_0)| \geq M_{\mathrm{BG}}$, see Remark 3.5).

Remark 3.25. We would like to emphasize that our result about resonances is remarkably different from the one obtained in [DLS06a], where it was considered the case of a perturbation h with a finite number of harmonics. In that case there was a uniform size for the gaps created by the resonances. For instance, the gaps created by the resonances of order 1 and 2 were $C\varepsilon^{1/2}$ and $C\varepsilon$, respectively. In our case we have a heterogeneous sea of gaps of different sizes. Moreover, in our case the resonances that create big gaps are just the resonances of order 1 up to some order M_{BG}, whereas in [DLS06a], both resonances of order 1 and 2 created big gaps.

3.3.4. Resonant regions with small gaps In this section, we will study the resonant regions with small gaps $\mathcal{D}_{\mathrm{SG}}$, which correspond to resonances $-l_0/k_0$ such that $|(k_0,l_0)| \geq M_{\mathrm{BG}}$, where M_{BG} was introduced in Theorem 2.1, of order j greater or equal than 1.

We will work in one connected component, and we will apply directly Theorem 3.22 to Hamiltonian (83) in order to prove that this component is covered by primary tori which are $\mathcal{O}(\varepsilon^{1+\eta})$-closely spaced, for some $\eta > 0$.

Going back to the original variables (I,φ,s) using the changes given by Theorem 3.11 and equation (22), we obtain the following result about the existence of invariant primary KAM tori of Hamiltonian (1):

Proposition 3.26 (Invariant tori in the small gaps region)**.** *Assume that* $m \geq 2$, $n \geq 2m+6$ *and* $r > (n-2)m+2$. *Choose any* $0 < \eta \leq 1/2\min(\nu - 6(1+\nu)/r, m-1-\alpha(6+2m))$, *for* $\nu > 1/(r/6-1)$. *Then, for* ε *small enough, in any connected component of* $\mathcal{D}_{\mathrm{SG}}$, *which is of the form (82) for some* $-l_0/k_0 \in \mathcal{R}^{[\leq m]}$ *with* $|(k_0,l_0)| \geq M_{\mathrm{BG}}$ *and* $L_{k_0} = L/|k_0|$ *with* $L = C\varepsilon^{\alpha}$ *and* $\alpha < 1/n$, *as required in Theorem 3.11, there exists a finite set of values* E_i *such that:*

i. For any E_i *there exists a torus* $\mathcal{T}_i$ *invariant by the flow of Hamiltonian (1) contained in* $\widetilde{\Lambda}_\varepsilon$, *which is given in* $\widetilde{\Lambda}_\varepsilon$ *by the equation* $F(I,\varphi,s;\varepsilon) \equiv E_i$, *where* F *is a* $\mathcal{C}^{n-2m-2-\varrho}$ *function, for any* $\varrho > 0$, *of the form*

$$F(I,\varphi,s;\varepsilon) = I + \mathcal{O}_{\mathcal{C}^2}(\varepsilon^{1+\eta}). \tag{87}$$

ii. The torus $\mathcal{T}_i$ *can be written as a graph of the variable* I *over the angle variables* (φ,s)*:*

$$\mathcal{T}_i = \{(I,\varphi,s) \in \mathcal{D}_{\mathrm{SG}}; I = \lambda_{E_i}(\varphi,s;\varepsilon)\},$$

with

$$\lambda_E(\varphi,s;\varepsilon) = E + U_E(\varphi,s;\varepsilon) \tag{88}$$

where U_E *is a* $\mathcal{C}^{n-2m-2-\varrho}$ *function, for any* $\varrho > 0$, *and* $U_E = \mathcal{O}_{\mathcal{C}^2}(\varepsilon^{1+\eta})$.

iii. These tori are $\mathcal{O}(\varepsilon^{1+\eta})$*-closely spaced in terms of the variable* I.

Proof: By Theorem 3.11, in any connected component of $\mathcal{D}_{\mathrm{SG}}$, Hamiltonian (23) expressed in the averaged variables $(\mathcal{B},\phi,s)$ has the expression (83).

Hamiltonian (83) is of the form (78), with $K_0(\mathcal{B};\varepsilon)=\frac{1}{2}\mathcal{B}^2+\varepsilon\widetilde{Z}^m(\mathcal{B};\varepsilon)$, which is a $\mathcal{C}^{r-2m+2}$ function and

$$K_1(\mathcal{B},\phi,s;\varepsilon)=\varepsilon^j(U_m^{k_0,l_0}(k_0\phi+l_0s;\varepsilon)+\varepsilon^{m+1-j}\bar{R}^m(\mathcal{B},\phi,s;\varepsilon)), \qquad (89)$$

which is a $\mathcal{C}^{r-2m}$ function and 2π-periodic in both angle variables ϕ and s.

Our aim is to apply KAM Theorem 3.22. It is clear that $|K_0''(\cdot;\varepsilon)|\geq M>0$, for M independent of ε. We now will see that K_1 in (89) satisfies $|K_1(\cdot,\cdot,s;\varepsilon)|_{\mathcal{C}^6}\leq\delta$, for $\delta=\varepsilon^{2+2\eta}$, for $\eta>0$.

Recall from Theorem 3.11 that $U_m^{k_0,l_0}(k_0\phi+l_0s;\varepsilon)$ is a polynomial in ε of degree $m-j$ and a trigonometric polynomial in $\theta=k_0\phi+l_0s$, which has the following bound with respect to θ

$$|\varepsilon^jU_m^{k_0,l_0}(\cdot;\varepsilon)|_{\mathcal{C}^{\ell},\theta}\preceq\varepsilon^{j-2\alpha(j-1)}|(k_0,l_0)|^{-r+2(j-1)}, \qquad (90)$$

and therefore

$$|\varepsilon^jU_m^{k_0,l_0}(\cdot;\varepsilon)|_{\mathcal{C}^{\ell},(\phi,s)}\preceq\varepsilon^{j-2\alpha(j-1)}|(k_0,l_0)|^{-r+2(j-1)+\ell} \qquad (91)$$

for $\ell=0,\dots,n-2m$. Moreover, $\varepsilon^{m+1}\bar{R}^m$ is a $\mathcal{C}^{r-2m}$ function with a bounded $\mathcal{C}^\ell$ norm up to $\ell=n-2m$ given by

$$\left|\varepsilon^{m+1}\bar{R}^m(\cdot;\varepsilon)\right|_{\mathcal{C}^\ell}\preceq\varepsilon^{m+1-\alpha(\ell+2m)}. \qquad (92)$$

Hence, from the estimates for the $\mathcal{C}^\ell$ norm of functions $\varepsilon^jU_m^{k_0,l_0}$ in (91) and $\varepsilon^{m+1}\bar{R}^m$ in (92) with $\ell=6$, one gets

$$|K_1(\cdot,\cdot,s;\varepsilon)|_{\mathcal{C}^6}\preceq\varepsilon^{j-2\alpha(j-1)}|(k_0,l_0)|^{-r+2(j-1)+6}+\varepsilon^{m+1-\alpha(6+2m)},$$

for any $1\leq j\leq m$. Taking into account that $|(k_0,l_0)|\geq M_{\mathrm{BG}}=\varepsilon^{-(1+\nu)/r}$ and that the worse estimate comes from $j=1$, one gets

$$|K_1(\cdot,\cdot,s;\varepsilon)|_{\mathcal{C}^6}\preceq\varepsilon\varepsilon^{\frac{1+\nu}{r}(r-6)}+\varepsilon^{m+1-\alpha(6+2m)}=\varepsilon^{2+\eta_1}+\varepsilon^{2+\eta_2},$$

where $\eta_1=\nu-6(1+\nu)/r$ and $\eta_2=m-1-\alpha(6+2m)$ are both positive. Indeed, by hypotheses $m\geq2$ and $\alpha<1/n\leq1/(2m+6)$, we have $\eta_2>0$ and $\eta_1>0$ is equivalent to $\nu>1/(r/6-1)$.

So, for any $\eta\leq1/2\min(\eta_1,\eta_2)$ we have $|K_1(\cdot,\cdot,s;\varepsilon)|_{\mathcal{C}^6}\preceq\varepsilon^{2+2\eta}$ and we can apply KAM Theorem 3.22 with $n_0=5$, $\beta=1-\varrho$, for any $\varrho>0$, $\delta=\varepsilon^{2+2\eta}$ and M independent of ε. Therefore, we conclude that for a constant $C(k_0)=\text{cte}$ because $k_0=1$, regions $\mathcal{D}_{\mathrm{BG}}$ contain KAM tori given by

$$\mathcal{B}=\mathcal{B}_i+\Psi_i(\phi,s;\varepsilon),$$

where $\Psi_i(\phi,s;\varepsilon)$ is a $\mathcal{C}^{4-\varrho}$ function, for any $\varrho>0$, and

$$|\Psi_i|_{\mathcal{C}^2}\preceq\varepsilon^{1+\eta}.$$

These tori are $\mathcal{O}(\varepsilon^{1+\eta})$-closely spaced in terms of the variable $\mathcal{B}$.

As in the non-resonant regions we can go back to the original variables (I,φ,s). Using that the change $(J,\varphi,s)\mapsto(\mathcal{B},\phi,s)$ is $\varepsilon^{1-\alpha(\ell+2)}$-close to the identity in the $\mathcal{C}^\ell$

sense for $\ell = 0, \ldots, n-2m$ by Theorem 3.11 and $(I, \varphi, s) \mapsto (J, \varphi, s)$ is ε-close to the identity in the $\mathcal{C}^{r-1}$ sense by equation (22), the invariant tori obtained in the region $\mathcal{D}_{\mathrm{nr}}^m$ are given by

$$I = I_i + U_i(\varphi, s; \varepsilon)$$

where the function U_i verifies the same properties as Ψ_i, and they are $\mathcal{O}(\varepsilon^{1+\eta})$-closely spaced in terms of the variable I. We get the results claimed for $E_i = I_i$. □

Remark 3.27. Notice that invariant tori in the small gaps region $\mathcal{D}_{\mathrm{SG}}$ are given by a certain function F in (87) that, as in the case of non-resonant regions (see (79)), is of the form

$$F(I, \varphi, s; \varepsilon) = I + \mathcal{O}_{\mathcal{C}^2}(\varepsilon^{1+\eta}), \tag{93}$$

for some $\eta > 0$.

3.3.5. Resonant regions with big gaps In this section, we will see that the resonant regions with big gaps $\mathcal{D}_{\mathrm{BG}}$ which correspond to resonances of order 1 such that $|(k_0, l_0)| < M_{\mathrm{BG}}$ are covered with invariant objects (either primary tori or secondary tori) which are $\mathcal{O}(\varepsilon^{1+\eta})$-closely spaced in terms of the action variable I, for some $\eta > 0$.

To that end, we will apply Theorem 3.22 to Hamiltonian (81) as we did in the previous cases. The main difference is that in this case the integrable Hamiltonian is not written down into action angle variables, so we will need to perform a change of coordinates before applying KAM theorem. Furthermore, we will perform two useful changes of coordinates, which are not symplectic but conformally symplectic.

Finally, going back to the original variables (I, φ, s) using the changes given by Theorem 3.11 and equation (22), we obtain the following result about the existence of invariant tori of Hamiltonian (1):

Theorem 3.28 (Invariant tori in the big gaps region)**.** *Assume that $m \geq 10$, $n \geq 2m+6$ and $r > (n-2)m + 2$. Assume that the function $U_m^{k_0,l_0}(k_0\phi + l_0 s; 0)$ in Hamiltonian (83) has a global maximum which is non degenerate (this assumption corresponds to the hypothesis* **H3'** *on (k_0, l_0) in Theorem 2.1). Choose any $0 < \eta \leq \nu/2$ and assume $\nu \leq 1/16$.*

Then, for ε small enough, in any connected component of $\mathcal{D}_{\mathrm{BG}}$, which is of the form (82), for some $-l_0/k_0$ of order 1 such that $|(k_0, l_0)| < M_{\mathrm{BG}}$, $L_{k_0} = L/|k_0|$ with $L = C\varepsilon^{\alpha}$ and $\alpha < 1/n$, as required in Theorem 3.11, there exists a finite set of values E_i in some range of energies $-\varepsilon|(k_0, l_0)|^{-r+2} \leq E \leq L^2$ such that:

i. For any E_i there exist invariant objects by the flow of Hamiltonian (1) contained in $\widetilde{\Lambda}_\varepsilon$, which are given in $\widetilde{\Lambda}_\varepsilon$ by the equation $F(I, \varphi, s; \varepsilon) \equiv E_i$, where F is a $\mathcal{C}^{4-\varrho}$ function, for any $\varrho > 0$, of the form

$$\begin{aligned} F(I, \varphi, s; \varepsilon) = {} & \frac{(k_0 I + l_0)^2}{2}(1 + \varepsilon k_0^2 \tilde{h}(k_0 I + l_0; \varepsilon)) + \varepsilon k_0^2 U_m^{k_0,l_0}(k_0\varphi + l_0 s; \varepsilon) \\ & + \mathcal{O}_{\mathcal{C}^2}(k_0^4 |(k_0, l_0)|^{-r/2} \varepsilon^{3/2+\eta}), \end{aligned} \tag{94}$$

where $\tilde{h}$ satisfies (106). For values of $E_i > 0$, equation $F \equiv E_i$ consists of two invariant objects that are primary KAM tori $\mathcal{T}^{\pm}_{E_i}$, whereas for $E_i < 0$ consists of an invariant object which is a secondary KAM torus $\mathcal{T}_{E_i}$. In this case we denote $\mathcal{T}^{\pm}_{E_i}$ each of the components of

$$\mathcal{T}_{E_i} \cap \{(I, \varphi, s) \in \mathcal{D}_{\mathrm{BG}}; \rho \leq k_0\varphi + l_0 s \leq 2\pi - \rho\},$$

for some $0 < \rho < 2\pi$.

ii. *There exists $\rho \geq 0$, such that the two primary KAM tori (components of the secondary tori) $\mathcal{T}^{\pm}_{E_i}$ contained in $\widetilde{\Lambda}_\varepsilon$ can be written as graphs of the variable I over the angle variables (φ, s):*

$$\mathcal{T}^{\pm}_{E_i} = \{(I, \varphi, s) \in [-l_0/k_0 - L_{k_0}, -l_0/k_0 + L_{k_0}] \times [\rho, 2\pi - \rho] \times \mathbb{T};\ I = \lambda^{\pm}_{E_i}(\varphi, s; \varepsilon)\},$$

where

$$\lambda^{\pm}_{E}(\varphi, s; \varepsilon) = -\frac{l_0}{k_0} + \frac{1}{k_0}\mathcal{Y}_{\pm}(\theta, E) + \mathcal{O}_{\mathcal{C}^2}(\varepsilon^{1+\eta}), \tag{95}$$

for $\rho \leq \theta = k_0\varphi + l_0 s \leq 2\pi - \rho$, where

$$\mathcal{Y}_{\pm}(x, E) = \pm(1 + \varepsilon b)\ell(\theta, E) + \varepsilon\widetilde{\mathcal{Y}}_{\pm}(\ell(\theta, E)), \tag{96}$$

$\ell(\theta, E) = \sqrt{2(E - \varepsilon k_0^2 U_m^{k_0,l_0}(\theta; \varepsilon))}$ *and $\widetilde{\mathcal{Y}}_{\pm}$ satisfies (118).*

iii. *These invariant tori are $\mathcal{O}(\varepsilon^{1+\eta})$-closely spaced in terms of the variable I and $\mathcal{O}(\varepsilon^{3/2+\eta}|(k_0, l_0)|^{-r/2+1})$ in terms of energies E_i.*

Remark 3.29. In Remark 3.12 we already pointed out that the function $U_m^{k_0,l_0}(k_0\varphi + l_0 s; 0)$ given explicitly in (59) is the function $U^{k_0,l_0}(\theta)$ for $\theta = k_0\varphi + l_0 s$ in hypothesis **H3** on (k_0, l_0) in Theorem 2.1.

3.3.6. Proof of Theorem 3.28 The proof of this theorem is organized in three parts.

Invariant tori given by the averaged Hamiltonian

By Theorem 3.11, in any connected component of the resonant domain $\mathcal{D}_{\mathrm{BG}}$, which is of the form (82), Hamiltonian (23) expressed in the averaged variables $(\mathcal{B}, \phi, s)$ is of the form (83) with $j = 1$, so it can be written as

$$\begin{aligned} k_m(\mathcal{B}, \phi, s; \varepsilon) &= \frac{1}{2}\mathcal{B}^2 + \varepsilon\widetilde{Z}^m(\mathcal{B}; \varepsilon) + \varepsilon U_m^{k_0,l_0}(k_0\phi + l_0 s; \varepsilon) + \varepsilon^{m+1}\bar{R}^m(\mathcal{B}, \phi, s; \varepsilon) \\ &:= \bar{Z}^m(\mathcal{B}, \phi, s; \varepsilon) + \varepsilon^{m+1}\bar{R}^m(\mathcal{B}, \phi, s; \varepsilon), \end{aligned} \tag{97}$$

on the domain

$$\{(\mathcal{B}, \phi, s) \in \mathbb{R} \times \mathbb{T}^2;\ |\mathcal{B} + l_0/k_0| \leq \bar{L}_{k_0}\}, \tag{98}$$

where $|L_{k_0} - \bar{L}_{k_0}| \leq \mathrm{cte}\,\varepsilon$.

In this domain, $\varepsilon\widetilde{Z}^m(\mathcal{B}; \varepsilon)$ is a $\mathcal{C}^{r-2m+2}$ function in the variable $\mathcal{B}$ and it is a polynomial of degree m in ε, whose coefficient of order $q + 1$, for $q = 1, \ldots, m - 1$ is a $\mathcal{C}^{r-2q}$ function and $\mathcal{O}_{\mathcal{C}^\ell}(\varepsilon^{q+1-\alpha(\ell+2q)})$, for $\ell = 0, \ldots, n - 2q$, so that $\bar{Z}^m$ has a bounded norm up to $\ell = n-2m+2$. Its main term $\widetilde{Z}^m(\mathcal{B};0)$ is equal to $h_{0,0}(0,0,\mathcal{B};0)$ by Remark 3.12.

Moreover $U_m^{k_0,l_0}(k_0\phi+l_0 s;\varepsilon)$ is a polynomial of degree $m-1$ in ε and a trigonometric polynomial in $\theta = k_0\phi + l_0 s$, satisfying $\varepsilon U_m^{k_0,l_0}(\theta;\varepsilon) = \mathcal{O}_{\mathcal{C}^\ell}(\varepsilon|(k_0,l_0)|^{-r})$, for $\ell = 0,\dots,n$. Its main term $U_m^{k_0,l_0}(\theta;0)$ is given in expression (60) in Remark 3.12.

Finally, $\varepsilon^{m+1}\bar{R}^m$ is a $\mathcal{C}^{r-2m}$ function in the variables $(\mathcal{B},\phi,s)$ with a bounded $\mathcal{C}^\ell$ norm up to $\ell = n-2m$, which is

$$\left|\varepsilon^{m+1}\bar{R}^m(\cdot;\varepsilon)\right|_{\mathcal{C}^\ell} \preceq \varepsilon^{m+1-\alpha(\ell+2m)}. \tag{99}$$

By the hypothesis in Theorem 3.28, the function $U_m^{k_0,l_0}(\theta,0)$ (the first order term in ε of the function $U_m^{k_0,l_0}(\theta,\varepsilon)$) has a global maximum which is non-degenerate and this implies that the integrable part $\bar{Z}^m$ of the Hamiltonian (97) is like an integrable pendulum.

As it has been done in Section 8.5.2 in [DLS06a], we perform two useful changes of coordinates which are not symplectic but conformally symplectic. The first one depends on the time s and the resonance (k_0,l_0) and is given by:

$$b = k_0(\mathcal{B} + l_0/k_0), \quad \theta = k_0\phi + l_0 s, \quad s = s, \tag{100}$$

hence the system of equations verified by (b,θ,s) is also Hamiltonian of Hamiltonian:

$$\bar{K}(b,\theta,s;\varepsilon) = \bar{K}^0(b;\varepsilon) + \varepsilon\bar{V}(\theta;\varepsilon) + \varepsilon^{m+1}\bar{K}^1(b,\theta,s;\varepsilon), \tag{101}$$

with

$$\begin{aligned} \bar{K}^0(b,\varepsilon) &= b^2/2 + \varepsilon k_0^2 \widetilde{Z}^m(-l_0/k_0 + b/k_0;\varepsilon),\\ \bar{V}(\theta;\varepsilon) &= k_0^2 U_m^{k_0,l_0}(\theta;\varepsilon),\\ \bar{K}^1(b,\theta,s;\varepsilon) &= k_0^2 \bar{R}^m(-l_0/k_0 + b/k_0, \frac{\theta - l_0 s}{k_0}, s;\varepsilon). \end{aligned} \tag{102}$$

Note that $\bar{K}^0$ is of class $\mathcal{C}^{r-2m+2}$ with a bounded $\mathcal{C}^\ell$ norm up to $\ell = n-2m+2$ and $\bar{V}$ is analytic because it is a trigonometric polynomial in θ and a polynomial of degree $m-1$ in ε. $\bar{K}^1$ is a function of class $\mathcal{C}^{r-2m}$ with a bounded $\mathcal{C}^\ell$ norm up to $\ell = n-2m$, which is $2\pi k_0$-periodic in θ and 2π-periodic in s. Notice that $\bar{V}$ is 2π-periodic in θ, whereas $\bar{K}^1$ is $2\pi k_0$-periodic in θ.

The integrable part $\bar{K}^0(b;\varepsilon) + \varepsilon\bar{V}(\theta;\varepsilon)$ of the Hamiltonian (101) is a one degree of freedom Hamiltonian close to a pendulum-like Hamiltonian

$$\frac{b^2}{2} + \varepsilon\bar{V}(\theta;0) = \frac{b^2}{2} + \varepsilon k_0^2 U_m^{k_0,l_0}(\theta;0),$$

where $U_m^{k_0,l_0}(\theta;0)$ is given in (60). By hypothesis **H3'** on (k_0,l_0) this pendulum-like Hamiltonian has a hyperbolic saddle at $(0,\theta_1)$ and by the implicit function theorem the whole integrable Hamiltonian $\bar{K}^0(b;\varepsilon) + \varepsilon\bar{V}(\theta;\varepsilon)$ has also a saddle at $(b(\varepsilon),\theta_1(\varepsilon))$. Since $\widetilde{Z}^m(\mathcal{B};0) = h(0,0,\mathcal{B};0)$ does not depend on ε, the function $b(\varepsilon)$ is of class $\mathcal{C}^{r-2m+1}$ in ε and of the form $b(\varepsilon) = \mathcal{O}(|k_0|\varepsilon)$ whereas $\theta_1(\varepsilon)$ is analytic in ε and of the form $\theta_1(\varepsilon) = \theta_1 + \mathcal{O}(\varepsilon)$.

To make the analysis of this system easier we perform a second change of variables, which depends on ε and consists of the following translation

$$y = b - b(\varepsilon), \quad x = \theta - \theta_1(\varepsilon), \quad s = s, \tag{103}$$

in such a way that the integrable part of the Hamiltonian expressed in these new variables has a saddle point at $(0,0)$ and the energy of the saddle and the separatrices is 0. More precisely, we obtain the $\mathcal{C}^{r-2m}$ Hamiltonian with respect to (y,x,s) with a bounded $\mathcal{C}^\ell$ norm up to $\ell = n-2m$

$$K(y,x,s;\varepsilon) = h^0(y;\varepsilon) + \varepsilon U(x;\varepsilon) + \varepsilon^{m+1} S(y,x,s;\varepsilon) \tag{104}$$

which consists of an integrable part corresponding to the terms up to order ε^m, which is the following $\mathcal{C}^{r-2m+2}$ function with a bounded $\mathcal{C}^\ell$ norm up to $\ell = n-2m+2$,

$$K_0(y,x;\varepsilon) = h^0(y;\varepsilon) + \varepsilon U^{k_0,l_0}(x;\varepsilon), \tag{105}$$

and a perturbation $\varepsilon^{m+1}S(y,x,s;\varepsilon)$, which is a $\mathcal{C}^{r-2m}$ function with a bounded $\mathcal{C}^\ell$ norm up to $\ell = n-2m$.

The function $h^0(y;\varepsilon)$ in the integrable part K_0 is a $\mathcal{C}^{r-2m+2}$ function in y with a bounded $\mathcal{C}^\ell$ norm up to $\ell = n-2m+2$ of the form

$$h^0(y;\varepsilon) = \frac{y^2}{2}\widehat{h}(y;\varepsilon) = \frac{y^2}{2}(1+\varepsilon k_0^2 \widetilde{h}(y;\varepsilon)), \tag{106}$$

for some $\mathcal{C}^{r-2m}$ function in (y,ε), $\widetilde{h}(y;\varepsilon)$, with a bounded $\mathcal{C}^\ell$ norm up to $\ell = n-2m$ in y. The function U in K_0 is given by

$$U(x;\varepsilon) = k_0^2(U_m^{k_0,l_0}(x+\theta_1(\varepsilon);\varepsilon) - U_m^{k_0,l_0}(\theta_1(\varepsilon);\varepsilon)), \tag{107}$$

and it satisfies

$$|\varepsilon U(\cdot;\varepsilon)|_{\mathcal{C}^\ell} \preceq \varepsilon |k_0|^2 |(k_0,l_0)|^{-r} \tag{108}$$

for $\ell = 0,\dots n$.

We also notice that the following conditions are satisfied,

$$h^0(0;\varepsilon) = \frac{\partial h^0}{\partial y}(0;\varepsilon) = 0, U(0;\varepsilon) = \frac{\partial U}{\partial x}(0;\varepsilon) = 0, \frac{\partial^2 U}{\partial x^2}(0;\varepsilon) < 0,$$

as well as that $x=0$ is a global maximum of U.

The perturbation term $\varepsilon^{m+1}S(y,x,s;\varepsilon)$ is given by

$$S(y,x,s;\varepsilon) = k_0^2 \bar{R}^m \left(-\frac{l_0}{k_0} + \frac{y+b(\varepsilon)}{k_0}, \frac{x+\theta_1(\varepsilon) - l_0 s}{k_0}, s; \varepsilon \right)$$

and by equation (99) it can be bounded in the variables (y,x) by

$$\left|\varepsilon^{m+1} S(\cdot,s;\varepsilon)\right|_{\mathcal{C}^\ell} \preceq |k_0|^{2-\ell} \varepsilon^{m+1-\alpha(\ell+2m)} \tag{109}$$

for $\ell = 0,\dots,n-2m$.

Since we will want to apply some of the results in [DLS06a], it will be convenient for us to have K_0 written in another way adapted to the notation in [DLS06a]. Motivated by the size $\varepsilon|k_0|^2|(k_0,l_0)|^{-r}$ of εU estimated in formula (108), we introduce here the parameter $\gamma\in\mathbb{R}$, $2>\gamma\geq 1$, depending on (k_0,l_0) and ε, such that

$$\varepsilon^{\gamma}=\varepsilon|k_0|^2|(k_0,l_0)|^{-r}, \tag{110}$$

in such a way that $\varepsilon U(\cdot;\varepsilon)=\mathcal{O}_{\mathcal{C}^\ell}(\varepsilon^{\gamma})$, for $\ell=0,\dots,n$.

Notice that $\gamma=1$ for small values of (k_0,l_0), that is $|(k_0,l_0)|\sim 1$, and in general $1<\gamma<2+\nu$ for $|(k_0,l_0)|\sim\varepsilon^{-\varrho}$, for any $0<\varrho<(1+\nu)/r$, where $0<\nu\leq 1/16$.

With this choice of γ, we will denote K_0 the one degree of freedom $\mathcal{C}^{r-2m+2}$ Hamiltonian in (y,x)

$$K_0(y,x;\varepsilon)=h^0(y;\varepsilon)+\varepsilon^{\gamma}\widetilde{U}(x;\varepsilon), \tag{111}$$

where

$$\varepsilon^{\gamma}\widetilde{U}(x;\varepsilon)=\varepsilon U(x;\varepsilon), \tag{112}$$

with $2+\nu>\gamma\geq 1$ and $\widetilde{U}(\cdot;\varepsilon)=\mathcal{O}_{\mathcal{C}^\ell}(1)$, for $\ell=0,\dots,n$.

The energy level $K_0(y,x;\varepsilon)=0$ consists of the saddle $(0,0)$ and its separatrices.

The Hamiltonian $K(y,x,s;\varepsilon)$ introduced in (104) is $2\pi k_0$-periodic in x and 2π-periodic in s and is defined in the domain D_{k_0} given by

$$D_{k_0}=\{(y,x,s)\in\mathbb{R}\times\mathbb{R}/(2\pi k_0\mathbb{Z})\times\mathbb{T}, |y|\leq\bar{L}\}, \tag{113}$$

where $\bar{L}=k_0\bar{L}_{k_0}$, whereas the integrable part $K_0(y,x;\varepsilon)$ in (111) is 2π-periodic in x and independent of s, therefore the region D_{k_0} can be seen as k_0 copies of the region

$$D=\{(y,x,s)\in\mathbb{R}\times\mathbb{T}^2, |y|\leq\bar{L}\}.$$

This effect is colloquially described as saying that the resonance $-l_0/k_0$ has k_0 eyes. As k_0 increases, these eyes form long necklaces.

The region D (and also D_{k_0}) is filled by the energy surfaces of the Hamiltonian K_0,

$$\mathcal{T}_E^0=\{(y,x,s)\in[-\bar{L},\bar{L}]\times\mathbb{T}^2 : K_0(y,x;\varepsilon)=E\}$$

which are invariant under the flow of Hamiltonian K_0.

As we already said, the energy surface $\mathcal{T}_0^0$ corresponding to $E=0$ consists of the saddle $(0,0)$ and its separatrices forming a separatrix loop. Therefore, this separatrix loop $\mathcal{T}_0^0$ separates two types of topological invariant objects. The energy surfaces corresponding to the values $E>0$ are primary tori and the ones corresponding to the the values $E<0$ are called secondary tori, which are tori of different topology than the primary ones because they are contractible to points. Secondary tori cover all the region inside the separatrix loop $\mathcal{T}_0^0$. In the next section we will discuss the persistence of primary and secondary tori when we add the perturbation term.

KAM Theorem

In this section, we will show that many of the invariant tori $\mathcal{T}_E^0$ of the Hamiltonian $K_0(y,x;\varepsilon)$ in (111), inside the region D_{k_0} given in (113), both primary and secondary, survive when we add the perturbation term $\varepsilon^{m+1}S(y,x,s;\varepsilon)$ to consider the Hamiltonian

K given in equation (104). Moreover, we will estimate the number of steps of averaging m required to get invariant tori with a distance of $\mathcal{O}(\varepsilon^{1+\eta})$ between them, for some $\eta > 0$, in terms of the original variables (I, φ, s).

To establish this we will write the Hamiltonian (111) into action-angle variables and apply KAM Theorem 3.22. Since the unperturbed Hamiltonian $K_0(y, x; \varepsilon)$ is a pendulum, we can not define global action-angle variables because the change of coordinates becomes singular on the separatrix. Therefore, we will define different action-angle variables inside and outside the separatrix and we will exclude a thin neighborhood around it.

We will find convenient to consider different regions in the domain D_{k_0} in terms of the values of the energy E, in which the behavior of the tori is different.

Recall that tori $\mathcal{T}_E^0$ in D_{k_0} are given approximately by the energy surfaces of Hamiltonian K_0, that is

$$K_0(y, x; \varepsilon) = E,$$

and we will see that excluding an small interval they can be seen as a graph of the action variable y over the angle variables (x, s).

Introducing $\delta = \varepsilon^\gamma$, we consider the foliation given by the level sets

$$h^0(y; \varepsilon) + \delta \widetilde{U}(x; \varepsilon) = E, \tag{114}$$

where $h^0(y; \varepsilon)$ is of the form (106) and $\widetilde{U}(\cdot; \varepsilon) = \mathcal{O}_{\mathcal{C}^\ell}(1)$ for $\ell = 0, \dots, n$ satisfies also that on $x = 0$ there is a non-degenerate global maximum of $\widetilde{U}(x; \varepsilon)$, which verifies $-c \leq \widetilde{U}(\cdot; \varepsilon) \leq 0$ and $\widetilde{U}(\cdot; \varepsilon) \simeq -ax^2$ as $x \to 0$, with $a > 0$.

Since $h^0(y; \varepsilon) + \delta \widetilde{U}(x; \varepsilon) \simeq \frac{y^2}{2} + \delta \widetilde{U}(x; \varepsilon)$, the main term in the solution of (114) is

$$y = \pm \ell(x, E), \tag{115}$$

where

$$\ell(x, E) = \sqrt{2(E - \delta \widetilde{U}(x; \varepsilon))}. \tag{116}$$

Writing y in (114) as a function of (115), we can apply the implicit function theorem to equation (114) and we get a solution $y = \mathcal{Y}_\pm(x, E)$ given by

$$\mathcal{Y}_\pm(x, E) = \pm(1 + \varepsilon b)\ell(x, E) + \varepsilon \widetilde{\mathcal{Y}}_\pm(\ell(x, E)), \tag{117}$$

where

i. $b = \mathcal{O}(|k_0|\varepsilon)$ and independent of δ. Moreover, $\widetilde{\mathcal{Y}}_\pm(0) = \widetilde{\mathcal{Y}}'_\pm(0) = 0$.

ii. $\varepsilon \widetilde{\mathcal{Y}}_\pm$ is a $\mathcal{C}^{r-2m+2}$ function and

$$\begin{aligned} \left| \varepsilon \widetilde{\mathcal{Y}}_\pm \circ \ell \right|_{\mathcal{C}^s(\mathcal{I}_{E_0})} &\preceq |k_0|\varepsilon, \quad s = 0, 1, \\ \left| \varepsilon \widetilde{\mathcal{Y}}_\pm \circ \ell \right|_{\mathcal{C}^s(\mathcal{I}_{E_0})} &\preceq |k_0|\varepsilon E_0^{-s+1/2}, \quad 2 \leq s \leq n - 2m + 2, \end{aligned} \tag{118}$$

where $\mathcal{I}_{E_0} := \{(x, E), x \in \mathbb{T}, E \geq E_0 > 0\}$.

This result is stated explicitly in Lemma 8.34 in [DLS06a]. For more details and a rigorous proof we refer the reader to it.

From expression (116) it is clear that the size of the energy determines the dominant terms in $\ell(x, E)$. Thus, if $E \gg \delta = \varepsilon^\gamma$ the tori $\mathcal{T}_E^0$ are rather flat because the term $\varepsilon^\gamma \widetilde{U}(x; \varepsilon)$ is very small compared with E, whereas if $E \leq \varepsilon^\gamma$, the term $\sqrt{E - \varepsilon^\gamma \widetilde{U}(x; 0)}$ and therefore the size of y oscillates between E and ε^γ and it has the effect of bending the tori up to the point that they are bunched near the critical point (see Figure 1).

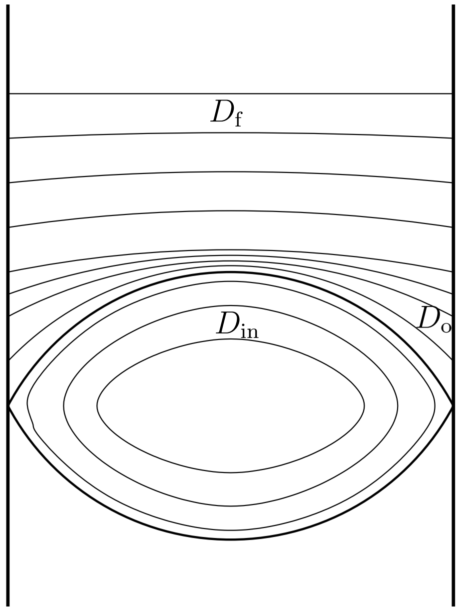

Figure 1. Schematic representation for the bending effect

Hence D_{k_0} will be divided in three regions in a similar way as in [DLS06a]: D_f is the region far from the separatrix, D_o close to the separatrix but outside the region bounded by the separatrix loop and D_in close to the separatrix but inside the separatrix loop, in the following way:

$$D_\text{f} = \{(y, x, s) \in D_{k_0} : K_0(y, x; \varepsilon) = E, \varepsilon^\gamma \leq E \leq \bar{L}^2\} \tag{119}$$

$$D_\text{o} = \{(y, x, s) \in D_{k_0} : K_0(y, x; \varepsilon) = F, \varepsilon^\beta \leq F \leq \varepsilon^\gamma\} \tag{120}$$

$$D_\text{in} = \{(y, x, s) \in D_{k_0} : K_0(y, x; \varepsilon) = G, -\varepsilon^\gamma \leq G \leq -\varepsilon^\beta\} \tag{121}$$

where $1 \leq \gamma < 2 + \nu$ as in (110) and β is arbitrary provided that $\beta > \gamma$ (see Figure 1).

Theorem 3.30 establishes the existence of primary tori in $D_\text{f} \cup D_\text{o}$ and secondary tori in D_in at a certain distance between them that depends on the number m of averaging steps and close to the level sets of the averaged Hamiltonian $K_0(y, x; \varepsilon)$.

Theorem 3.30 (KAM Theorem in the big gaps region)**.** *Consider the $\mathcal{C}^{r-2m}$ reduced Hamiltonian $K(y, x, s; \varepsilon)$ given in (104) inside the region D_{k_0} defined in (113). Consider $\beta > \gamma$, with γ as in (110) and assume that $r > (n-2)m + 2$, $n \geq 2m + 6$ and $m \geq 14(\beta - \gamma) + 3\gamma/2$. Then, for $|\varepsilon|$ small enough, one has:*

1. ***Primary tori far from resonance***. *There exists a set of values* $E_1 < \cdots < E_{l_E}$ *verifying* $\varepsilon^\gamma \le E_i \le \bar{L}^2 \sim \varepsilon^{2\alpha}$ *and* $\alpha < 1/n$*, such that*
   - (a) *The frequencies* $\omega(E_i)$ *are Diophantine numbers of constant type and Markov constant* $\mathrm{cte}\, E_i^{-1/4} \varepsilon^{\frac{m+1-\alpha(6+2m)}{2}} |k_0|$.
   - (b) *For any value* E_i*, there exist two primary invariant tori* $\mathcal{T}^\pm_{E_i}$ *of Hamiltonian (104) contained in* D_f.
   - (c) *The motion of the tori* $\mathcal{T}^\pm_{E_i}$ *is* $\mathcal{C}^1$*-conjugated to a rigid translation of frequencies* $(\omega(E_i), 1)$.
   - (d) *This tori can be written as*
   $$\mathcal{T}^+_{E_i} = \{(y,x,s) \in D_\mathrm{f}, K_{E_i}(y,x,s;\varepsilon) = E_i, y > 0\}$$
   $$\mathcal{T}^-_{E_i} = \{(y,x,s) \in D_\mathrm{f}, K_{E_i}(y,x,s;\varepsilon) = E_i, y < 0\}$$
   where $K_{E_i}(y,x,s;\varepsilon)$ *is a* $\mathcal{C}^{4-\varrho}$ *function, for any* $\varrho > 0$*, given by*
   $$K_{E_i}(y,x,s;\varepsilon) = K_0(y,x;\varepsilon) + \mathcal{O}_{\mathcal{C}^2}\left(\varepsilon^{\frac{m+1-\alpha(6+2m)}{2}} E_i^{1/4} |k_0|\right) \tag{122}$$
   - (e) $D_\mathrm{f} \subset \bigcup_i B(\mathcal{T}^\pm_{E_i}, \varepsilon^{\frac{m+1-\alpha(6+2m)}{2}} E_i^{1/4}|k_0|)$*, where*
   $$B(\mathcal{T}^\pm_E, \delta) = \{(y,x,s) \in D_{k_0}, |K_0(y,x;\varepsilon) - E| \le \delta\}$$
2. ***Primary tori close to resonance.*** *There exists a set of values* $F_1 < \cdots < F_{l_F}$ *verifying* $\varepsilon^\beta \le F_i \le \varepsilon^\gamma$*, such that*
   - (a) *The frequencies* $\omega(F_i)$ *are Diophantine numbers of constant type and Markov constant* $\mathrm{cte}\, \varepsilon^{\frac{m+1-\alpha(6+2m)-\gamma/2+6\gamma}{2}} F_i^{-3} |k_0|$.
   - (b) *For any value* F_i*, there exist two primary invariant tori* $\mathcal{T}^\pm_{F_i}$ *of Hamiltonian (104) contained in* D_o.
   - (c) *The motion of the tori* $\mathcal{T}^\pm_{F_i}$ *is* $\mathcal{C}^1$*-conjugated to a rigid translation of frequencies* $(\omega(F_i), 1)$.
   - (d) *This tori can be written as*
   $$\mathcal{T}^+_{F_i} = \{(y,x,s) \in D_\mathrm{o}, K_{F_i}(y,x,s;\varepsilon) = F_i, y > 0\}$$
   $$\mathcal{T}^-_{F_i} = \{(y,x,s) \in D_\mathrm{o}, K_{F_i}(y,x,s;\varepsilon) = F_i, y < 0\}$$
   where $K_{F_i}(y,x,s;\varepsilon)$ *is a* $\mathcal{C}^{4-\varrho}$ *function, for any* $\varrho > 0$*, given by*
   $$K_{F_i}(y,x,s;\varepsilon) = K_0(y,x;\varepsilon) + \mathcal{O}_{\mathcal{C}^2}\left(\varepsilon^{\frac{m+1-\alpha(6+2m)+\gamma/2+14\gamma}{2}} F_i^{-7} |k_0|\right) \tag{123}$$
   - (e) $D_\mathrm{o} \subset \bigcup_i B(\mathcal{T}^\pm_{F_i}, \varepsilon^{\frac{m+1-\alpha(6+2m)+\gamma/2+10\gamma}{2}} F_i^{-5}|k_0|)$*, where*
   $$B(\mathcal{T}^\pm_E, \delta) = \{(y,x,s) \in D_{k_0}, |K_0(y,x;\varepsilon) - E| \le \delta\}$$
3. ***Secondary tori close to resonance.*** *There exists a set of values* $G_1 < \cdots < G_{l_G}$ *verifying* $-\varepsilon^\gamma \le G_i \le -\varepsilon^\beta$*, such that*
   - (a) *The frequencies* $\omega(G_i)$ *are Diophantine numbers of constant type and Markov constant* $\mathrm{cte}\, \varepsilon^{\frac{m+1-\alpha(6+2m)-\gamma/2+6\gamma}{2}} |G_i|^{-3} |k_0|$.
   - (b) *For any value* G_i*, there exist a secondary invariant torus* $\mathcal{T}^\pm_{G_i}$ *of Hamiltonian (104) contained in* D_in*, contractible to the set*
   $$\{(0,a,s), a \in \mathbb{R}, s \in \mathbb{R}/(2\pi k_0 \mathbb{Z})\} \subset D_\mathrm{in}$$

(c) *The motion on the torus* $\mathcal{T}_{G_i}$ *is* $\mathcal{C}^1$*-conjugated to a rigid translation of frequencies* $(\omega(G_i), 1)$.

(d) *This torus can be written as*

$$\mathcal{T}_{G_i} = \{(y, x, s) \in D_{\rm in}, K_{G_i}(y, x, s; \varepsilon) = G_i\}$$

where $K_{G_i}(y, x, s; \varepsilon)$ *is a* $\mathcal{C}^{4-\varrho}$ *function, for any* $\varrho > 0$*, given by*

$$K_{G_i}(y, x, s; \varepsilon) = K_0(y, x; \varepsilon) + \mathcal{O}_{\mathcal{C}^2}\left(\varepsilon^{\frac{m+1-\alpha(6+2m)+\gamma/2+14\gamma}{2}} |G_i|^{-7} |k_0|\right) \tag{124}$$

(e) $D_{\rm in} \subset \bigcup_i B(\mathcal{T}^{\pm}_{G_i}, \varepsilon^{\frac{m+1-\alpha(6+2m)+\gamma/2+10\gamma}{2}} |G_i|^{-5}|k_0|)$.

The following Corollary makes more explicit the assertions about the proximity of these tori as a function of m, and it also gives properties of the KAM tori when expressed as graphs of the action y in terms of the angle variables (x, s).

Corollary 3.31. *Consider the* $\mathcal{C}^{r-2m}$ *reduced Hamiltonian* $K(y, x, s; \varepsilon)$ *given in (104) inside the region* D_{k_0} *defined in (113). Consider* $\beta = \gamma/2 + 1 + \nu/2$*, with* $1 \leq \gamma < 2 + \nu$ *as in (110) and* $\nu \leq 1/16$*. Assume that* $r > (n-2)m + 2$*,* $n \geq 2m + 6$ *and* $m \geq 10$*. Then, the tori obtained in Theorem 3.30 verify:*

1. *For any value* E_i*, the primary tori* $\mathcal{T}^{\pm}_{E_i}$ *can be written as graphs of the action* y *over the angles* (x, s)*:*

$$\mathcal{T}^{\pm}_{E_i} = \{(y, x, s) \in D_{\rm f}, y = f^{\pm}_{E_i}(x, s; \varepsilon)\}.$$

2. *For any value* F_i*, the primary tori* $\mathcal{T}^{\pm}_{F_i}$ *can be written as graphs of the action* y *over the angles* (x, s)*:*

$$\mathcal{T}^{\pm}_{F_i} = \{(y, x, s) \in D_{\rm o}, y = f^{\pm}_{F_i}(x, s; \varepsilon)\}.$$

3. *There exists* $\rho_0 > 0$ *such that for any* $0 < \rho_0 \leq \rho \leq \pi$*, and for any value* G_i*, each of the components of*

$$\mathcal{T}_{G_i} \cap \{(y, x, s) : x \in I_\rho\}, \qquad I_\rho = \bigcup_{l=0}^{k_0-1} [2\pi l + \rho, 2\pi(l+1) - \rho],$$

that we will denote by $\mathcal{T}^{\pm,\rho}_{G_i}$*, can be written as a graph of the action* y *over the angles* (x, s)*:*

$$\mathcal{T}^{\pm,\rho}_{G_i} = \{(y, x, s) \in D_{\rm i}, x \in I_\rho, y = f^{\pm}_{G_i}(x, s; \varepsilon)\}$$

4. *All these functions* $f_v = f^{\pm}_v$ *are at least of class* $\mathcal{C}^2$ *with respect to* (x, s)*, and, denoting by* D *the derivatives with respect to* x *and* s*, for* $v = E_i$*,* $i = 1, \ldots, l_E$*,* $v = F_i$*,* $i = 1, \ldots, l_F$*, and* $v = G_i$*,* $i = 1, \ldots, l_G$*, they verify:*

   (a) *There exists a function* $\mathcal{Y}(x, E)$ *given explicitly in (117) such that:*

$$|f_v - \mathcal{Y}(x, v)|_{\mathcal{C}^1} \preceq |k_0| \varepsilon^{1+\nu/2} \tag{125}$$

   (b) $|Df_v| \preceq \varepsilon^{\gamma/2}$, $|D^2 f_v| \preceq \varepsilon^{\gamma/2}$.

   (c) *For any two consecutive values* v *and* $\overline{v}$ *we have:*

$$|v - \bar{v}| \preceq |k_0| \varepsilon^{\beta},$$

and

$$|f_v - f_{\overline{v}}|_{\mathcal{C}^1} \preceq \frac{|v - \overline{v}|}{\varepsilon^{\gamma/2}} \preceq |k_0|\varepsilon^{1+\nu/2}.$$

Proof of Theorem 3.30

The proof follows the strategy established in [DLS06a], with the same scaling in the domains D_{o} and D_{in}. The main difference is that we will perform a sequence of scalings in the far domain D_{f}, whereas in [DLS06a] there was no scaling in this region. This sequence of scalings in D_{f} will reduce the number of averaging steps m needed to get tori close enough in the region D_{f}, and therefore the required differentiability r.

We will first give a detailed proof of part 1) of this Theorem. Notice that in D_{f} defined in (119), the energy E ranges from ε^γ to $\bar{L}^2 \sim \varepsilon^{2\alpha}$. Hence, we consider a value of E, let us say E_l, in the interval $[\varepsilon^\gamma, \varepsilon^{2\alpha}]$ and a small neighborhood around it of the form $[c_a E_l, c_b E_l] \subseteq [\varepsilon^\gamma, \varepsilon^{2\alpha}]$, where c_a, c_b are constants independent of ε and E_l, such that $c_a < 1$ and $c_b > 1$. Thus, we introduce the following domain contained in D_{f}:

$$D_{E_l} = \{(y, x, s) \in D_{\mathrm{f}} : K_0(y, x; \varepsilon) = E,\ c_a E_l \leq E \leq c_b E_l\}. \tag{126}$$

By the equation of K_0 in (111) and the expression of h^0 in (106), the main term in y is given in (116). Therefore, in D_{E_l} the coordinate y is of size $\mathcal{O}(\sqrt{E_i})$ and it is natural to perform the scaling

$$y = \sqrt{E_l}Y, \tag{127}$$

which transforms the Hamiltonian system of Hamiltonian $K(y, x, s; \varepsilon)$ given in (104), which is $\mathcal{C}^{r-2m}$ with respect to the variables (y, x, s) with a bounded $\mathcal{C}^\ell$ norm up to $\ell = n - 2m$, into a Hamiltonian system of $\mathcal{C}^{r-2m}$ Hamiltonian with respect to (Y, x, s) with a bounded $\mathcal{C}^\ell$ norm up to $\ell = n - 2m$,

$$\begin{aligned} \mathcal{K}(Y, x, s; \sqrt{E_l}, \varepsilon) &= \frac{1}{\sqrt{E_l}} \mathcal{K}(\sqrt{E_l}Y, x, s; \varepsilon) \\ &= \sqrt{E_l}\mathcal{K}_0(Y, x; \sqrt{E_l}, \varepsilon) + \frac{\varepsilon^{m+1}}{\sqrt{E_l}} S(\sqrt{E_l}Y, x, s; \varepsilon), \end{aligned} \tag{128}$$

with

$$\begin{aligned} \mathcal{K}_0(Y, x; \sqrt{E_l}, \varepsilon) &= \frac{1}{E_l} K_0(\sqrt{E_l}Y, x; \varepsilon) \\ &= \frac{Y^2}{2}\widehat{h}(\sqrt{E_i}Y; \varepsilon) + \frac{\varepsilon^\gamma}{E_i}\widetilde{U}(x; \varepsilon), \end{aligned} \tag{129}$$

where $\widehat{h}(y; \varepsilon) = 1 + \mathcal{O}(|k_0|^2\varepsilon)$ is given in (106) and, consequently, $\mathcal{K}_0$ is a $\mathcal{C}^{r-2m+2}$ function with respect to (Y, x) with a bounded $\mathcal{C}^\ell$ norm up to $\ell = n - 2m + 2$, because $\widehat{h}(y; \varepsilon)$ is $\mathcal{C}^{r-2m+2}$ with respect to y with a bounded $\mathcal{C}^\ell$ norm up to $\ell = n - 2m + 2$.

The scaling (127) transforms the domain D_{E_l} in (126) into

$$\begin{aligned} \widetilde{D} &= \{(Y, x, s) \in \mathbb{R} \times \mathbb{R}/2\pi k_0\mathbb{Z} \times \mathbb{T} : \mathcal{K}_0(Y, x; \sqrt{E_l}, \varepsilon) = E/E_l,\ c_a E_i \leq E \leq c_b E_i\} \\ &= \{(Y, x, s) \in \mathbb{R} \times \mathbb{R}/2\pi k_0\mathbb{Z} \times \mathbb{T} : \mathcal{K}_0(Y, x; \sqrt{E_l}, \varepsilon) = e,\ c_a \leq e \leq c_b\}. \end{aligned} \tag{130}$$

Next we will define the action-angle variables (A, ψ) associated to the Hamiltonian $\mathcal{K}_0(Y, x; \sqrt{E_l}, \varepsilon)$ in the domain $\widetilde{D}$. Note that the Hamiltonian $\mathcal{K}(Y, x, s; \sqrt{E_l}, \varepsilon)$ is $2\pi k_0$-periodic in x and 2π-periodic in s, whereas $\mathcal{K}_0(Y, x; \sqrt{E_l}, \varepsilon)$ is 2π-periodic in x and independent of s. Therefore, the domain $\widetilde{D}$ is nothing else but k_0 copies of the domain $D^* \times \mathbb{T}$, where

$$D^* = \{(Y,x) \in \mathbb{R} \times \mathbb{T} : \mathcal{K}_0(Y,x;\sqrt{E_l},\varepsilon) = e,\ c_a \le e \le c_b\}. \tag{131}$$

Notice that, by expression (129) for $\mathcal{K}_0$, the equation

$$\mathcal{K}_0(Y,x;\sqrt{E_l},\varepsilon) = e$$

has the same form as equation (114) with $\delta = \varepsilon^\gamma/E_l$ and it defines two functions $Y = \mathcal{Y}_\pm(x,e)$ on D^*, given in (117), which are of the form

$$\mathcal{Y}_\pm(x,e) = \pm\sqrt{2\left(e - \frac{\varepsilon^\gamma}{E_l}\widetilde{U}(x;\varepsilon)\right)}(1 + \mathcal{O}_{\mathcal{C}^{n-2m+2}}(|k_0|\varepsilon)).$$

Since, by construction of $\widetilde{U}(x;\varepsilon)$, on $x = 0$ there is a global maximum such that $-c \le \widetilde{U}(x;\varepsilon) \le 0$, in the domain D^* we have

$$0 \le c_a \le e \le e - \frac{\varepsilon^\gamma}{E_l}\widetilde{U}(x;\varepsilon) \le e + c\frac{\varepsilon^\gamma}{E_l} \le c_b + \text{cte}\,,$$

where we have used $E_l \ge \varepsilon^\gamma$ and therefore $\widetilde{c}_a \le \mathcal{Y}_\pm(x,e) \le \widetilde{c}_b$+cte and $\mathcal{Y}_\pm$ is $\mathcal{O}_{\mathcal{C}^{n-2m+2}}(1)$, for some constants $\widetilde{c}_a$ and $\widetilde{c}_b$.

We consider in D^* the action angle variables

$$\begin{aligned} A &= \frac{1}{2\pi}\int_0^{2\pi} \mathcal{Y}_\pm(x,e)\mathrm{d}x, \\ \psi &= \frac{2\pi}{T(e)}\tau(x,e), \end{aligned} \tag{132}$$

where $\tau(x,e)$ is the time along the orbit of the Hamiltonian $\mathcal{K}_0(Y,x;\sqrt{E_l},\varepsilon)$ with energy e given by

$$\tau(x,e) = \int_0^x \frac{\partial \mathcal{Y}_\pm}{\partial e}(u,e)\mathrm{d}u. \tag{133}$$

We have chosen the origin of time at $x = 0$ and with this choice $T(e) = \tau(2\pi, e)$ is the period of the periodic orbit.

From expression (132) it is obvious that A satisfies $\widetilde{c}_a \le A \le \widetilde{c}_b$ and that A is $\mathcal{O}_{\mathcal{C}^{n-2m+3}}(1)$.

The action-angle variables (A,ψ) introduced in (132) have already been studied in Proposition 8.35 of [DLS06a] for the case when they become singular, that is when the domain D^* depends on ε. In our case, we can adapt the result in Proposition 8.35 of [DLS06a] for the domain D^* not depending on ε. We obtain that we can express the integrable Hamiltonian $\sqrt{E_l}\mathcal{K}_0(Y,x;\sqrt{E_l},\varepsilon)$ in (129) into action-angle variables (A,ψ) in the domain D^* and the change of coordinates is away from the singularity in this domain. More precisely, there exists a $\mathcal{C}^{r-2m+2}$ change of variables in D^*

$$\begin{aligned} \mathcal{X} : D^{**} &\to D^* \\ (A,\psi) &\mapsto (Y,x) \end{aligned} \tag{134}$$

given in (132) with $D^{**} = \{(A,\psi) : \widetilde{c}_a \le A \le \widetilde{c}_b, \psi \in \mathbb{T}\} = [\widetilde{c}_a, \widetilde{c}_b] \times \mathbb{T}$ and $\widetilde{c}_a, \widetilde{c}_b$, suitable constants independent of ε and E_l, such that:

i. $\mathcal{K}_0(\mathcal{X}(A,\psi);\sqrt{E_l},\varepsilon)=\mathcal{G}(A;\sqrt{E_l},\varepsilon)$.

ii. $|\mathcal{X}|_{\mathcal{C}^{n_0}(\mathcal{D}^{**})}\preceq 1$, $\left|\mathcal{X}^{-1}\right|_{\mathcal{C}^{n_0}(D^*)}\preceq 1$, $0\le n_0\le n-2m+2$.

iii. $|\mathcal{G}|_{\mathcal{C}^3(D^{**})}\preceq 1$ and $\left|\mathcal{G}''\right|_{\mathcal{C}^0(D^{**})}\succeq 1$

where the constants in above inequalities do not depend on ε and E_l.

Now, we consider the Hamiltonian $\mathcal{K}$ in (128) expressed in action-angle variables,

$$\widetilde{\mathcal{K}}(A,\psi,s;\sqrt{E_l},\varepsilon)=\sqrt{E_l}\mathcal{G}(A;\sqrt{E_l},\varepsilon)+\frac{\varepsilon^{m+1}}{\sqrt{E_l}}\widetilde{S}(A,\psi,s;\sqrt{E_l},\varepsilon), \quad (135)$$

where $\widetilde{\mathcal{K}}=\mathcal{K}\circ\mathcal{X}$ and $\widetilde{S}=S\circ\mathcal{X}$.

The Hamiltonian (135) is of the form (78) with $K_0=\sqrt{E_l}\mathcal{G}(A;\sqrt{E_l},\varepsilon)$ and $K_1=\varepsilon^{m+1}E_l^{-1/2}\widetilde{S}(A,\psi,s;\sqrt{E_l},\varepsilon)$ and $2\pi k_0$-periodic in ψ.

The functions $\mathcal{G}$ and $\widetilde{S}$ are $\mathcal{C}^{r-2m+2}$ and $\mathcal{C}^{r-2m}$ with bounded $\mathcal{C}^\ell$ norms up to $\ell=n-2m+2$ and $\ell=n-2m$ in the variables (A,ψ), respectively. Since by hypotheses of Theorem 3.30 we have that $r>n\ge 2m+6$, $\mathcal{G}$ and $\widetilde{S}$ have a bounded $\mathcal{C}^6$ norm in the variables (A,ψ). Therefore, using Faa-di Bruno formula (C.3) and the bound for the $\mathcal{C}^6$ norm in the variables (y,x) for $\varepsilon^{m+1}S$ in expression (109) jointly with the bounds for the change of coordinates $\mathcal{X}$ in item ii) we have that, for any $s\in\mathbb{T}$,

$$\left|\frac{\varepsilon^{m+1}}{\sqrt{E_l}}\widetilde{S}(\cdot,s;\sqrt{E_l},\varepsilon)\right|_{\mathcal{C}^6(D^{**}_{k_0})}\preceq|k_0|^{-4}E_l^{-1/2}\varepsilon^{m+1-\alpha(6+2m)},$$

where $D^{**}_{k_0}=[\widetilde{c}_a,\widetilde{c}_b]\times\mathbb{R}/2\pi k_0\mathbb{Z}$. Moreover, by item iii) in this proof we have that

$$\sqrt{E_l}\left|\mathcal{G}''(\cdot;\sqrt{E_l},\varepsilon)\right|_{\mathcal{C}^0(D^{**})}\succeq\sqrt{E_l}.$$

Therefore, we can apply KAM Theorem 3.22 to Hamiltonian (135) with $n_0=5$, $\beta=1-\varrho$, for any $\varrho>0$, $\delta=\delta(\varepsilon)=|k_0|^{-4}E_l^{-1/2}\varepsilon^{m+1-\alpha(6+2m)}$ and $M=M(\varepsilon)=\text{cte}\sqrt{E_l}$ and we obtain:

1. There exist a set of values A_l, such that the Hamiltonian $\mathcal{K}\circ\mathcal{X}$ has invariant tori given by
$$\mathcal{T}_l=\{(A,\psi,s)\in D^{**}_{k_0}\times\mathbb{T}:A=A_l+\mathcal{A}_l(\psi,s;\sqrt{E_l},\varepsilon)\}$$
where $\mathcal{A}_l$ are $\mathcal{C}^{4-\varrho}$ functions in the variables (ψ,s), for any $\varrho>0$ and
$$\left|\mathcal{A}_l(\cdot;\sqrt{E_l},\varepsilon)\right|_{\mathcal{C}^2(\mathbb{R}/2\pi k_0\mathbb{Z}\times\mathbb{T})}\preceq|k_0|E_l^{-3/4}\varepsilon^{(m+1-\alpha(6+2m))/2}.$$
2. The motion of these tori is $\mathcal{C}^{2-\varrho}$-conjugate to a rigid translation of frequencies $(\omega(A_l),1)$, where $\omega(A_l)$ is a Diophantine number of constant type and Markov constant $\text{cte}\,|k_0|E_l^{-1/4}\varepsilon^{(m+1-\alpha(6+2m))/2}$.
3. The union of neighborhoods of size $|k_0|E_l^{-3/4}\varepsilon^{(m+1-\alpha(6+2m))/2}$ of these tori cover all the region $D^{**}_{k_0}\times\mathbb{T}$.

In the variables $(Y,x,s)=(\mathcal{X}(A,\psi),s)$, the torus $\mathcal{T}_l$ satisfies $\mathcal{K}_0(Y,x;\sqrt{E_l},\varepsilon)=\mathcal{G}(A_l+\mathcal{A}_l(\psi,s;\sqrt{E_l},\varepsilon);\sqrt{E_l},\varepsilon)$, so that, introducing $\mathcal{G}(A_l;\sqrt{E_l},\varepsilon)=e_l$ and using the estimates in items (ii) and (iii) in this proof as well as Faa-Di Bruno formulae, one obtains that the tori are given by

$$\begin{aligned}\mathcal{K}_0(Y,x;\sqrt{E_l},\varepsilon) &= \mathcal{G}(A_l;\sqrt{E_l},\varepsilon) + +\mathcal{O}_{\mathcal{C}^2}\left(|\mathcal{G}|_{\mathcal{C}^3}\,|\mathcal{A}_l|_{\mathcal{C}^2}\,\left|\mathcal{X}^{-1}\right|^2_{\mathcal{C}^2}\right)\\ &= e_l + \mathcal{O}_{\mathcal{C}^2}\left(|k_0|E_l^{-3/4}\varepsilon^{(m+1-\alpha(6+2m))/2}\right) \qquad (136)\end{aligned}$$

Going back to the variables (y,x,s) performing the scaling $y=\sqrt{E_l}Y$ and using the expression for $\mathcal{K}_0$ given in (129) one obtains that the tori are given by

$$K_0(y,x;\varepsilon) = E_i + \mathcal{O}_{\mathcal{C}^2}\left(|k_0|E_i^{1/4}\varepsilon^{\frac{m+1-\alpha(6+2m)}{2}}\right),$$

where $E_i = E_l e_l$.

By compactness of $D_{\rm f}$, the covering $\{\mathrm{int}(D_{E_i})\}_{i=1}^{\infty}$ of $D_{\rm f}$ admits a finite subcovering $D_{\rm f} = \bigcup_{i=0}^{N}\mathrm{int}(D_{E_i})$, and we get the claimed results in part 1 of Theorem 3.30.

The proof of parts 2) and 3) of this Theorem follows as in [DLS08]. The only difference is that we introduce a sequence of domains as we did in this proof in the far region and we perform adequate scalings which allow us to get better estimates for the functions describing the searched tori. More precisely, consider the region $D_{\rm o}$ (the case for $D_{\rm in}$ is analogous) and introduce the domain

$$D_{F_l} = \{(y,x,s)\in D_{\rm o} : K_0(y,x;\varepsilon) = F,\ c_aF_l \le F \le c_bF_l\},$$

analogous to (126) in part 1). Since the energy $F_l \le \varepsilon^{\gamma}$ in $D_{\rm o}$ (see (120)), from the expression for the main term of y given by $\ell(x,E)$ in (116), the coordinate y ranges from $\sqrt{F_l}$ to $\varepsilon^{\gamma/2}$. Hence we perform the scaling $y=\varepsilon^{\gamma/2}Y$ and we proceed as in Lemma 8.36 in [DLS06a]. We obtain that the original system is transformed into a Hamiltonian system of $\mathcal{C}^{r-2m}$ Hamiltonian with respect to (Y,x,s) of the form

$$\mathcal{K}(Y,x,s;\varepsilon^{\gamma/2},\varepsilon) = \varepsilon^{\gamma/2}\mathcal{K}_0(Y,x;\varepsilon^{\gamma/2},\varepsilon) + \varepsilon^{m+1-\gamma/2}S(\varepsilon^{\gamma/2}Y,x,s;\varepsilon),$$

with

$$\mathcal{K}_0(Y,x;\varepsilon^{\gamma/2},\varepsilon) = \frac{Y^2}{2}\widehat{h}(\sqrt{E_i}Y;\varepsilon) + \widetilde{U}(x;\varepsilon)$$

where $\widehat{h}(y;\varepsilon) = 1+\mathcal{O}(|k_0|^2\varepsilon)$ is given in (106). The Hamiltonian is defined now on the domain

$$\begin{aligned}\widetilde{D} &= \{(Y,x,s)\in\mathbb{R}\times\mathbb{R}/2\pi k_0\mathbb{Z}\times\mathbb{T} : \mathcal{K}_0(Y,x;\varepsilon^{\gamma/2}) = F/F_l,\ c_a^0F_i \le F \le c_b^0F_i\}\\ &= \{(Y,x,s)\in\mathbb{R}\times\mathbb{R}/2\pi k_0\mathbb{Z}\times\mathbb{T} : \mathcal{K}_0(Y,x;\varepsilon^{\gamma/2}) = e,\ c_a^0F_l/\varepsilon^{\gamma} \le e \le c_b^0F_l/\varepsilon^{\gamma}\}\end{aligned}$$

Next, we define the action angle variables in the domain $\widetilde{D}$ by formulas (132). The only change is that we need to take into account that instead of expression (8.77) in [DLS06a] we have

$$c_a\frac{F_l}{\varepsilon^{\gamma}} \le e - \widetilde{U}(x;\varepsilon) \le c_b\frac{F_l}{\varepsilon^{\gamma}} + c \le \mathrm{cte}\,,$$

and by (109) the perturbation $\varepsilon^{m+1-\gamma/2}S(\varepsilon^{\gamma/2}Y,x,s;\varepsilon^{\gamma/2})$ can be bounded in the $\mathcal{C}^6$ norm in the variables (Y,x) by $\varepsilon^{-\gamma/2}\varepsilon^{m+1-\alpha(6+2m)}|k_0|^{-4}$.

Therefore we can apply Proposition 8.38 in [DLS06a] and proceed as in the proof of parts 2) and 3) of Theorem 8.30 in [DLS06a] replacing in the estimates in terms of ε in (2.1), equation (8.50) and (2.5), ε^j by ε^γ, ε^{m+1} by $\varepsilon^{m+1-\alpha(6+2m)}|k_0|^{-4}$ and $\varepsilon^{\alpha-j}$ by $F_l\varepsilon^{-\gamma}$, and multiplying by the constant $C_{k_0} = \text{cte}\,|k_0|^3$ of KAM Theorem 3.22, to obtain the estimates in 2.(a), equation (123) and 2.(e). Finally, by compactness of $D_{\rm o}$, we get the claimed results. We skip the proof of these two parts and we refer the reader to Section 8.5.4 in [DLS06a] for it. □

Proof of Corollary 3.31. It is totally analogous to the proof of corollary 8.31 in [DLS06a] and it follows from Theorem 3.30 just applying the implicit function theorem.

We apply Theorem 3.30, with $m \geq 10$ and $\beta = \gamma/2 + 1 + \nu/2$, where $1 \leq \gamma < 2 + \nu$ and $\nu \leq 1/16$. From these conditions it follows that $\beta > \gamma$ and $m \geq 14(\beta - \gamma) + 3\gamma/2$ and therefore, we obtain that the invariant tori in the domains $D_{\rm f}$, $D_{\rm o}$ and $D_{\rm in}$ are given by the implicit equations (122), (123) and (124), which are of the form

$$K_0(y, x, s; \varepsilon) = E + \delta g(y, x, s, E; \varepsilon) \tag{137}$$

with $|g|_{\mathcal{C}^2} \leq \text{cte}$ and

$$\begin{aligned} &E = E_i,\, \delta = \varepsilon^{\frac{m+1-\alpha(6+2m)}{2}} E_i^{1/4} |k_0|, \\ &E = F_i,\, \delta = \varepsilon^{\frac{m+1-\alpha(6+2m)+\gamma/2+14\gamma}{2}} F_i^{-7} |k_0|, \\ &E = G_i, \delta = \varepsilon^{\frac{m+1-\alpha(6+2m)+\gamma/2+14\gamma}{2}} |G|_i^{-7} |k_0|, \end{aligned} \tag{138}$$

respectively.

Equation (137) is equivalent to equation

$$M(y, x, s, t; \delta, \varepsilon) \equiv y - \mathcal{Y}_\pm(x, t) = 0,$$

where $t = E + \delta g(y, x, s, E; \varepsilon)$ and $\mathcal{Y}_\pm(x, t)$ is given in equation (117). The above equation has been studied in full detail in Lemma 8.39 of [DLS06a]. It is not difficult to check that one has

$$\left| \frac{\partial M}{\partial y} - 1 \right| \leq \text{cte}\, \delta \varepsilon^{-\gamma/2},$$

which is a bound analogous to (8.95) in Lemma 8.39 in [DLS06a], where the factor ε^γ comes directly from the expression (111)of K_0. So, as long as $\delta\varepsilon^{-\gamma/2} \leq \delta_0 \ll 1$, for some constant δ_0 independent of ε, we can apply the implicit function Theorem in order to get the invariant tori of items 1,2 and 3 written as graphs of the action y over the angles (x, s) as

$$y = f_v^\pm(x, s; \varepsilon)$$

where $v = E_i, F_i, G_i$, respectively and

$$f_v^\pm(x, s; \varepsilon) = \mathcal{Y}_\pm(x, v) + \mathcal{O}_{\mathcal{C}^1}(\delta\varepsilon^{-\gamma/2}).$$

Let us check first that condition $\delta\varepsilon^{-\gamma/2} \ll 1$ is fulfilled. Notice, first, that by the choice $m \geq 10$ and $\beta = \gamma/2 + 1 + \nu/2$, where $\nu \leq 1/16$, $E_i \preceq \varepsilon^{2\alpha}$ and $F_i, G_i \geq \varepsilon^\beta$, one obtains in the three cases of (138), that $|\delta| \leq |k_0|\varepsilon^\beta$, which clearly implies $\delta\varepsilon^{-\gamma/2} \leq |k_0|\varepsilon^{1+\nu/2} \leq \delta_0$, for some constant $\delta_0 \ll 1$ since, by expression (110), $|k_0| \leq \varepsilon^{-(1+\nu)/r} \leq \varepsilon^{-1}$. Thus, we obtain results in items 1), 2), 3) and

$$|f_v - \mathcal{Y}(x,v)|_{\mathcal{C}^1} \preceq \delta\varepsilon^{-\gamma/2} \preceq |k_0|\varepsilon^{\beta-\gamma/2} = |k_0|\varepsilon^{1+\nu/2},$$

as claimed in 4a). In an analogous way one gets 4b).

Finally, from results 1e), 2e) and 3e) in Theorem 3.30 and definitions of $D_{\rm f}$, $D_{\rm o}$ and $D_{\rm in}$ given in (119), (120) and (121) we have

$$\begin{aligned} |E_i - E_{i+1}| &\preceq \varepsilon^{\frac{m+1-\alpha(6+2m)}{2}}(E_i^{1/4} + E_{i+1}^{1/4})|k_0| \\ |F_i - F_{i+1}| &\preceq \varepsilon^{\frac{m+1-\alpha(6+2m)+\gamma/2+10\gamma}{2}}(F_i^{-5} + F_{i+1}^{-5})|k_0| \\ |G_i - G_{i+1}| &\preceq \varepsilon^{\frac{m+1-\alpha(6+2m)+\gamma/2+10\gamma}{2}}(|G_i|^{-5} + |G_{i+1}|^{-5})|k_0| \end{aligned}$$

and taking into account that $E_1 \sim F_{l_F} \sim \varepsilon^\gamma$ and $F_1 \sim G_{l_G} \sim \varepsilon^\beta$ we get

$$\begin{aligned} |E_1 - F_{l_F}| &\preceq \varepsilon^{\frac{m+1-\alpha(6+2m)+\gamma/2}{2}}|k_0| \\ |F_1 - G_{l_G}| &\preceq (\varepsilon^\beta + \varepsilon^{\frac{m+1-\alpha(6+2m)+\gamma/2-10(\beta-\gamma)}{2}})|k_0|. \end{aligned}$$

Since $\beta = \gamma/2 + 1 + \nu/2$, $m \geq 10$, all these exponents are bigger than β as claimed in item 4c). The last estimate in item 4c) follows from the inequalities above and the following bounds

$$\left|\frac{\partial f_E}{\partial E}\right| \preceq \varepsilon^{-\gamma/2}, \qquad \left|\frac{\partial D f_E}{\partial E}\right| \preceq \varepsilon^{-\gamma/2}.$$

analogous to (8.91) given by Lemma 8.39 in [DLS06a]. □

Remark 3.32. In the case considered in [DLS06a], where the perturbation h in (1) was assumed to be a trigonometric polynomial in the angular variables (φ, t), there exist a finite number of resonances so L can be chosen independently of ε, that is $\alpha = 0$. Moreover γ is simply replaced by the values $j = 1, 2$ in [DLS06a] corresponding to resonances of order 1 and 2, respectively. In this case, Corollary 3.31 only requires $m \geq 9$ and $r = n \geq 24$ since there is no need of truncation process, so that Hamiltonian in (1) only needs to be $\mathcal{C}^{26}$. This improves substantially the regularity required in [DLS06a], since Hamiltonian (1) was assumed to be $\mathcal{C}^{60}$ because m was chosen $m = 26$.

Invariant tori in the original variables

Theorem 3.30 gives KAM tori, both primary and secondary, in the variables (y, x, s). From equations (122), (123) and (124) in Theorem 3.30, we know that these tori are given approximately by the level sets of the Hamiltonian $K_0(y, x; \varepsilon)$ in (111).

We can write them in the original variables (I, φ, s) using the change of coordinates given by Theorem 3.11 and changes (22), (100) and (103). More precisely, we have that the relation with the original variables is given by

$$y = k_0 I + l_0 + \mathcal{O}_{\mathcal{C}^2}(|k_0|\varepsilon^{1-4\alpha}), \quad x = k_0\varphi + l_0 s + \mathcal{O}_{\mathcal{C}^2}(|k_0|\varepsilon^{1-4\alpha}),$$

whose inverse in terms of the I variable can be written in the form

$$I = -\frac{l_0}{k_0} + \frac{1}{k_0}y + \zeta(y, x, s; \varepsilon),$$

where ζ is $\mathcal{O}_{\mathcal{C}^2}(\varepsilon^{1-4\alpha})$.

Using expression (105) and (106) these invariant objects are given by the level sets of a $\mathcal{C}^{4-\varrho}$ function F, for any $\varrho > 0$, which has the form

$$F(I,\varphi,s;\varepsilon) = \frac{(k_0 I + l_0)^2}{2}(1+\varepsilon k_0^2 \tilde{h}(k_0 I + l_0;\varepsilon)) + \varepsilon^{\gamma}\widetilde{U}(\theta;\varepsilon) + \mathcal{O}_{\mathcal{C}^2}(|k_0|^3 \varepsilon^{\gamma/2+1+\nu/2}), \tag{139}$$

where $\theta = k_0\varphi + l_0 s$. By the definition of γ in (110) jointly with $\widetilde{U}$ and U in (112) and (107), respectively, we get the expression (94) given in Theorem 3.28.

Moreover, from items (1), (2) and (3), together with the estimates in item (4a) in Corollary 3.31 we have that KAM tori can be written as graphs in the variables (y,x,s) of functions of the form

$$y = f_E^{\pm}(x,s;\varepsilon) = \mathcal{Y}_{\pm}(x,E) + \mathcal{O}_{\mathcal{C}^1}(|k_0|\varepsilon^{1+\eta}).$$

Using the mentioned changes, we obtain that the tori inside the region $\mathcal{D}_{\mathrm{BG}}$, are given in the original variables (I,φ,s) by

$$I = \lambda_E^{\pm}(\varphi,s;\varepsilon) = -\frac{l_0}{k_0} + \frac{1}{k_0}\mathcal{Y}_{\pm}(\theta,E) + \mathcal{O}_{\mathcal{C}^0}(\varepsilon^{1+\eta})$$

with $\theta = k_0\varphi + l_0 s$, where $\mathcal{Y}_{\pm}$ is given (117).

Finally, from Corollary 3.31 we know that there exist invariant tori $\mathcal{T}_E, \mathcal{T}_{E'}$ of energies E, E' such that

$$|E - E'| = \mathcal{O}(|k_0|\varepsilon^{\gamma/2+1+\nu/2}) = \mathcal{O}(|k_0|^2\varepsilon^{3/2+\nu/2}|(k_0,l_0)|^{-r/2})$$

and there exist also points $(y_1,x,s) \in \mathcal{T}_E$ and $(y_2,x,s) \in \mathcal{T}_{E'}$ with

$$|y_1 - y_2| = \mathcal{O}_{\mathcal{C}^1}(|k_0|\varepsilon^{1+\nu/2}),$$

so in term of their I variables it follows that

$$\begin{aligned} |I_1 - I_2| &\leq \frac{1}{|k_0|}|y_1-y_2| + \frac{1}{|k_0|}\left|\frac{\partial\zeta}{\partial y}\right||y_1-y_2| \\ &\preceq \varepsilon^{1+\nu/2} + |k_0|\varepsilon^{1-4\alpha}\varepsilon^{1+\nu/2} \\ &\preceq \varepsilon^{1+\nu/2}. \end{aligned}$$

and by the definition of γ given in (110), we obtain the claimed results in item (iii) of Theorem 3.28. □

3.4. Proof of Theorem 3.1

The proof of Theorem 3.1 follows directly from the results obtained in Propositions 3.24, 3.26 and Theorem 3.28.

Choosing $n = 2m+6$ and assuming $m \geq 10$ and $r > 2(m+1)^2$, the hypotheses on r, n and m in the mentioned Propositions and Theorem are satisfied. Moreover, the choice $\eta = \min((m-1-\alpha n)/2, \nu/2 - 3(1+\nu)/r)$ with $1/(r/6-1) < \nu \leq 1/16$, fits clearly with the assumptions on η in Propositions 3.24 and 3.26, and also with the one in Theorem 3.28.

By Propositions 3.24 and 3.26, the tori obtained in the non resonant region and in the resonant region with small gaps are primary and they are given by the level sets of the same function $F = I + \mathcal{O}_{\mathcal{C}^2}(\varepsilon^{1+\eta})$, so they are flat up to $\mathcal{O}_{\mathcal{C}^2}(\varepsilon^{1+\eta})$. Both regions form the flat tori region. The explicit approximate expressions for the invariant tori are given implicitly by the function (87) and as a graph of the action I over the variables (φ, s) by (88), both functions in Proposition 3.26.

By hypotheses **H3'**, Theorem 3.28 provides a sequence of invariant KAM tori (both primary and secondary) for the big gaps region. In a connected component of this region of the form (82), these tori are given by the level sets of a function F in (94) and as a graph of the action I over the angle variables (φ, s), in (95). Moreover, the distance between consecutive tori is $\mathcal{O}(\varepsilon^{1+\eta})$ in terms of the action variable and $\mathcal{O}(\varepsilon^{3/2+\eta}|(k_0, l_0)|^{-r/2+1})$ in terms of the energy. □

4. Construction of a transition chain

In the previous section, we have proved that in the NHIM $\widetilde{\Lambda}_\varepsilon$ there exists a discrete foliation of invariant tori $\mathcal{T}_i$ (primary and secondary) with graphs at a distance $\mathcal{O}_{\mathcal{C}^1}(\varepsilon^{1+\eta})$, for some $\eta > 0$. We have also shown that these tori are close to being the level sets of the averaged Hamiltonian, and we have given its first order perturbative calculation for the flat tori region $\mathcal{D}_{\mathrm{F}}$ and the big gaps region $\mathcal{D}_{\mathrm{BG}}$.

The goal of this section is to prove Proposition 4.1, which states that, assuming that the non-degeneracy conditions **H2"**, **H3"** and **H3"'** in Theorem 2.1 hold, there exists transversality between the foliation of invariant tori in $\widetilde{\Lambda}_\varepsilon$ provided by Theorem 3.1 and its image under the scattering map S_ε given in (20) and it is possible to construct a transition chain.

Recall that, as we said in Section 2.3.4, by Lemma 10.4 in [DLS06a] two submanifolds, like the invariant tori $\mathcal{T}_i$, $\mathcal{T}_{i+1}$ of the NHIM $\widetilde{\Lambda}_\varepsilon$, have a transverse heteroclinic intersection provided they are transversal under the scattering map S_ε as submanifolds of $\widetilde{\Lambda}_\varepsilon$:

$$S_\varepsilon(\mathcal{T}_i) \pitchfork_{\widetilde{\Lambda}_\varepsilon} \mathcal{T}_{i+1} \Rightarrow W^u_{\mathcal{T}_i} \pitchfork W^s_{\mathcal{T}_{i+1}}$$

.

Hence, Proposition 4.1 provides a transition chain through applications of the scattering map.

Proposition 4.1. *Consider Hamiltonian (1) satisfying the hypotheses of Theorem 2.1. Pick two KAM tori $\mathcal{T}_\pm$ such that $|I(x_\pm) - I_\pm| \leq \varepsilon^{1+\eta}$ for some $x_\pm \in \mathcal{T}_\pm$ and $\eta > 0$ (these tori exist thanks to Theorem 3.1). Then, there exists a transition chain $\{\mathcal{T}_i\}_{i=0}^{N(\varepsilon)}$, where $N(\varepsilon) = C/\varepsilon$, in such a way that*

1. *The transition chain is obtained through applications of the scattering map. That is,*

$$S_\varepsilon(\mathcal{T}_i) \pitchfork_{\widetilde{\Lambda}_\varepsilon} \mathcal{T}_{i+1}.$$

2. $\mathcal{T}_0 = \mathcal{T}_-$, $\mathcal{T}_{N(\varepsilon)} = \mathcal{T}_+$.

Proof. The proof of Proposition 4.1 is postponed to Section 4.2 and is based on the results in the following Section 4.1.

4.1. The scattering map and the transversality of heteroclinic intersections

The main result of this section is Lemma 4.2, stated below, which considers a foliation $\mathcal{F}_F$ whose leaves are the level sets of a certain function F and provides an expression for the action of the scattering map S_ε on this foliation in terms of the Hamiltonian function $\mathcal{S}_\varepsilon$ given in (19), generating its deformation. Moreover, it gives criteria to establish transversality between the foliation $\mathcal{F}_F$ and its image under the scattering map S_ε.

Lemma 4.2. *Consider the foliation $\mathcal{F}_F$ whose leaves L^F_E are the level sets of a certain function F:*

$$L^F_E = \{(I, \varphi, s) \in (I_-, I_+) \times \mathbb{T}^2, F(I, \varphi, s; \varepsilon) = E\}, \ E \in (E_1, E_2).$$

Let S_ε be the scattering map introduced in (17), and $\mathcal{S}_\varepsilon = \mathcal{S}_0 + \varepsilon\mathcal{S}_1 + \mathcal{O}(\varepsilon^2)$ its Hamiltonian function given in (19) with $\mathcal{S}_0 = -\mathcal{L}^*$, where $\mathcal{L}^*$ is the reduced Poincaré function introduced in (11). Then, $S_\varepsilon(L^F_E)$, the image sets of the leaves L^F_E of $\mathcal{F}_F$ under the scattering map S_ε, satisfy $S_\varepsilon(L^F_E) = L^{F\circ S_\varepsilon^{-1}}_E$ and therefore the equation $F \circ S_\varepsilon^{-1} = E$, where the expression $F \circ S_\varepsilon^{-1}$ is given by

$$F \circ S_\varepsilon^{-1} = F - \varepsilon\{F, \mathcal{S}_0\} + \frac{\varepsilon^2}{2}(\{\{F, \mathcal{S}_0\}, \mathcal{S}_0\} - \{F, \mathcal{S}_1\}) + \mathcal{O}(\varepsilon^3), \qquad (140)$$

where $\{F, \mathcal{S}_i\} = \partial_\varphi F \partial_I \mathcal{S}_i - \partial_I F \partial_\varphi \mathcal{S}_i$ is the Poisson bracket of the functions F and $\mathcal{S}_i$. Moreover, the image of a leaf L^F_E under the scattering map S_ε intersects another leaf $L^F_{E'}$, for some E', if and only if there exist $x \in L^F_E$ such that $F \circ S_\varepsilon(x) = E'$, where the expression $F \circ S_\varepsilon$ is given by

$$F \circ S_\varepsilon = F + \varepsilon\{F, \mathcal{S}_0\} + \frac{\varepsilon^2}{2}(\{\{F, \mathcal{S}_0\}, \mathcal{S}_0\} + \{F, \mathcal{S}_1\}) + \mathcal{O}(\varepsilon^3). \qquad (141)$$

Assuming that

$$\frac{|\{F, F \circ S_\varepsilon^{-1}\}|}{|\nabla F|^2} \geq C\varepsilon, \qquad (142)$$

where C is a constant independent of ε and E, the angle between the surfaces $L^F_{E'}$ and $S_\varepsilon(L^F_E)$ at the intersection points is bounded from below by $C\varepsilon$. Therefore, foliations $\mathcal{F}_F$ and $\mathcal{F}_{F\circ S_\varepsilon^{-1}}$ intersect transversally.

Remark 4.3. For the case of a function F which is $\mathcal{O}_{\mathcal{C}^2}(1)$, the scattering map increases (decreases) the energy E by order ε, provided that the first order term $\{F, \mathcal{L}^*\}$ in (141) satisfies

$$\{F, \mathcal{L}^*\} \not\equiv 0.$$

Remark 4.4. Using expression (140) and $\mathcal{S}_0 = -\mathcal{L}^*$, the condition for the transversality of the foliations (142) reads out

$$\frac{|\{F, \{F, \mathcal{L}^*\}\} + \varepsilon/2(-\{F, \{\{F, \mathcal{L}^*\}, \mathcal{L}^*\}\} + \{F, \{F, \mathcal{S}_1\}\}) + \mathcal{O}(\varepsilon^2)|}{|\nabla F|^2} \geq C. \tag{143}$$

Notice that if F is $\mathcal{O}_{\mathcal{C}^2}(1)$ the term ε can be neglected and the condition reduces to

$$\frac{|\{F, \{F, \mathcal{L}^*\}\}|}{|\nabla F|^2} \geq C. \tag{144}$$

Also notice that an equivalent condition to (142) is

$$\frac{|\{F, F \circ S_\varepsilon\}|}{|\nabla F|^2} \geq C\varepsilon. \tag{145}$$

Proof: In Section 2.3.2 we have shown that there exists a Hamiltonian function $\mathcal{S}_\varepsilon$ generating the deformation of the scattering map S_ε and we have given its first order perturbative computation in equation (19). Hence, taking into account that $\mathcal{S}_\varepsilon = \mathcal{S}_0 + \varepsilon\mathcal{S}_1 + \mathcal{O}(\varepsilon^2)$, it is clear that (see [CH82] for instance) $F \circ S_\varepsilon$ is given by

$$F \circ S_\varepsilon = F + \varepsilon\{F, \mathcal{S}_0\} + \frac{\varepsilon^2}{2}(\{\{F, \mathcal{S}_0\}, \mathcal{S}_0\} + \{F, \mathcal{S}_1\}) + \mathcal{O}(\varepsilon^3),$$

with $\mathcal{S}_0 = -\mathcal{L}^*$. The expression for $F \circ S_\varepsilon^{-1}$ follows identically.

In order to show the transversality between the foliations $\mathcal{F}_F$ and $\mathcal{F}_{F\circ S_\varepsilon^{-1}}$, we need to obtain lower bounds for the angle of intersection. More precisely, the angle α between the normal vectors to the tangent planes to the surfaces $S_\varepsilon(L_E^F)$ and $L_{E'}^F$ is given by

$$\sin(\alpha) = \frac{|\nabla(F \circ S_\varepsilon^{-1}) \times \nabla F|}{|\nabla(F \circ S_\varepsilon^{-1})||\nabla F|} = \frac{|\{F, F \circ S_\varepsilon^{-1}\}|}{|\nabla(F \circ S_\varepsilon^{-1})||\nabla F|},$$

where $F \circ S_\varepsilon^{-1}$ is given in expression (140). From this expression one can see that $\sin(\alpha)$ is $\mathcal{O}(\varepsilon)$ and condition (142) gives the required transversality. □

As we have argued in the previous section the tori in $\widetilde{\Lambda}_\varepsilon$ have different behavior depending whether they are close to or far from the separatrix. Thus, the tori in the flat tori region and in the big gaps region far from the resonance are rather flat, whereas they are bent in the big gaps region close to a resonance. The fact that the tori are not flat has the consequence that the dominant effect of comparing a torus with the image under the scattering map of another torus, will include some extra terms. For this reason, we will divide the study in three cases: on the one hand, the flat tori region and on the other hand the resonant region with big gaps, where we will distinguish between far from and close to the resonance.

4.1.1. The flat tori region In Lemma 4.5, we apply Lemma 4.2 to the flat tori region $\mathcal{D}_{\mathrm{F}}$. By Theorem 3.1, in one connected component of this region the invariant tori are given by the leaves L_E^F of a foliation $\mathcal{F}_F$, where F is of the form (87). Moreover they can be written as a graph of the action I over the angle variables (φ, s): $I = \lambda_E(\varphi, s; \varepsilon)$, where λ_E is given in (88).

Lemma 4.5. *Let us consider a foliation $\mathcal{F}_F$ contained in a connected component of the flat tori region $\mathcal{D}_{\mathrm{F}}$, where the function F is of the form (87), so that the equation $F(I,\varphi,s;\varepsilon)=E$ defines a smooth surface given as a graph $\lambda_E(\varphi,s;\varepsilon)$, with λ_E as in (88).*

Assume that hypothesis **H2"** *is fulfilled. More precisely, the reduced Poincaré function $\mathcal{L}^*$ defined in (11) verifies, for any value of $(I,\varphi,s)\in H_+\cap\mathcal{D}_{\mathrm{F}}$ that the function*

$$\widetilde{\theta}\mapsto\frac{\partial\mathcal{L}^*}{\partial\widetilde{\theta}}(I,\widetilde{\theta})$$

for $\widetilde{\theta}=\varphi-Is$ is positive (resp. negative) and non-constant for $\widetilde{\theta}$ on some set $\mathcal{J}^*_E$ (see (12)). Then the foliations $\mathcal{F}_F$ and $\mathcal{F}_{F\circ S^{-1}}$ intersect transversally.

More precisely, any surface $S_\varepsilon(L^F_E)$ intersects at some point the surface $L^F_{E'}$ for any $E'>0$ (resp. $E'<0$), $|E'-E|=\mathcal{O}(\varepsilon)$. The angle between the surfaces $S_\varepsilon(L^F_E)$ and $L^F_{E'}$ at the intersection can be bounded from below by $C\varepsilon$, where C is a constant independent of ε and E.

Proof: We will apply Lemma 4.2 with $F(I,\varphi,s;\varepsilon)=I+\mathcal{O}_{\mathcal{C}^2}(\varepsilon^{1+\eta})$ and $I=\lambda_E(I,\varphi,s;\varepsilon)=E+\mathcal{O}_{\mathcal{C}^2}(\varepsilon^{1+\eta})$ for some $\eta>0$. We will see that provided hypothesis **H2"** is fulfilled, condition (142) of Lemma 4.2 is satisfied.

We first apply the scattering map to the implicit surface

$$L^F_E=\{(I,\varphi,s)\in\mathcal{D}_{\mathrm{F}},F(I,\varphi,s)=E\},$$

and recall that $S_\varepsilon(L^F_E)$ intersects a leaf $L^F_{E'}$ at a point $(I,\varphi,s)\in L^F_E$ if $F\circ S_\varepsilon(I,\varphi,s;\varepsilon)=E'$, where, using expression (141), $F\circ S_\varepsilon$ is given by

$$F\circ S_\varepsilon=E-\varepsilon\{F,\mathcal{L}^*\}+\mathcal{O}(\varepsilon^2). \tag{146}$$

with,

$$\begin{aligned}\{F,\mathcal{L}^*\}&=-\frac{\partial\mathcal{L}^*}{\partial\varphi}\frac{\partial F}{\partial I}+\frac{\partial F}{\partial\varphi}\frac{\partial\mathcal{L}^*}{\partial I}\\&=-(1+\mathcal{O}_{\mathcal{C}^1}(\varepsilon^{1+\eta}))\frac{\partial\mathcal{L}^*}{\partial\widetilde{\theta}}+\mathcal{O}_{\mathcal{C}^1}(\varepsilon^{1+\eta})\\&=-\frac{\partial\mathcal{L}^*}{\partial\widetilde{\theta}}+\mathcal{O}_{\mathcal{C}^1}(\varepsilon^{1+\eta})\end{aligned}$$

with $\widetilde{\theta}=\varphi-Is$. Evaluating on $I=E+\mathcal{O}_{\mathcal{C}^0}(\varepsilon^{1+\eta})$, equation (146) reads out

$$(F\circ S_\varepsilon)(I,\varphi,s;\varepsilon)=E+\varepsilon\frac{\partial\mathcal{L}^*}{\partial\widetilde{\theta}}(E,\varphi-Es)+\mathcal{O}(\varepsilon^{1+\eta}).$$

By hypothesis **H2"** in Theorem 2.1 the scattering map increases for $(I,\varphi,s)\in H_+\cap\mathcal{D}_{\mathrm{F}}$ (resp. decreases) the energy by order ε. In particular, the surface $S_\varepsilon(L^F_E)$ intersects all surfaces $L^F_{E'}$ such that $|E'-E|=\mathcal{O}(\varepsilon)$.

Moreover, in order to see that they intersect transversally we need to check that condition (142) is satisfied. Notice that in this case, by Remark 4.4, condition (144) implies (142). Thus, we first compute

$$\begin{aligned}\{F,\{F,\mathcal{L}^*\}\} &= \left(\frac{\partial F}{\partial I}\right)^2 \frac{\partial^2 \mathcal{L}^*}{\partial \varphi^2} + \mathcal{O}_{\mathcal{C}^0}(\varepsilon^{1+\eta}) \\ &= (1+\mathcal{O}_{\mathcal{C}^0}(\varepsilon^{1+\eta}))^2 \frac{\partial^2 \mathcal{L}^*}{\partial \widetilde{\theta}^2} + \mathcal{O}_{\mathcal{C}^0}(\varepsilon^{1+\eta}) \\ &= \frac{\partial^2 \mathcal{L}^*}{\partial \widetilde{\theta}^2} + \mathcal{O}_{\mathcal{C}^0}(\varepsilon^{1+\eta}).\end{aligned}$$

Since, by assumption, the function $\frac{\partial \mathcal{L}^*}{\partial \widetilde{\theta}}(E,\widetilde{\theta})$ is non-constant for $\widetilde{\theta}$ in $\mathcal{J}_E^*$, there exists an interval $\bar{\mathcal{J}}_E \subset \mathcal{J}_E^*$ where

$$\left|\frac{\partial^2 \mathcal{L}^*}{\partial \widetilde{\theta}^2}\right| \geq C > 0,$$

and using

$$|\nabla F| = 1 + \mathcal{O}_{\mathcal{C}^1}(\varepsilon^{1+\eta}),$$

we have that condition (142) is satisfied and the angle between the surfaces $S_\varepsilon(L_E^F)$ and $L_{E'}^F$ at the intersection can be bounded from below by $C\varepsilon$, where C is a constant independent of ε and E. □

Remark 4.6. By Theorem 3.1, two consecutive tori are, at most, at distance of $\mathcal{O}(\varepsilon^{1+\eta})$, for some $\eta > 0$, in terms of the I variable. Moreover, these tori are $\mathcal{O}_{\mathcal{C}^0}(\varepsilon^{1+\eta})$ close to the level sets of the action I.

Hence, we conclude that the image under the scattering map of a torus $\mathcal{T}_i$ in the flat tori region, given by $I = I_i + \mathcal{O}(\varepsilon^{1+\eta})$ intersects transversally another torus of this region given by $I = I_{i+1} + \mathcal{O}(\varepsilon^{1+\eta})$ with $|I_{i+1} - I_i| = \mathcal{O}(\varepsilon)$:

$$S_\varepsilon(\mathcal{T}_i) \pitchfork \mathcal{T}_{i+1}.$$

4.1.2. Big gaps region In Lemma 4.7 we are going to apply Lemma 4.2 in one connected component of the big gaps region $\mathcal{D}_{\mathrm{BG}}$. By Theorem 3.1, the invariant tori are given by the leaves L_E^F of a foliation $\mathcal{F}_F$ for a certain function F of the form (94). Moreover, they can be written as a graph of the action I over the angle variables (φ, s): $I = \lambda_E^\pm(\varphi, s; \varepsilon)$, with $\lambda_E^\pm$ as in (95). Recall that in this foliation, the leaves with $E > 0$ are primary KAM tori whereas the leaves with $E < 0$ are secondary.

The dominant terms in F and in the expressions $\lambda_E^\pm$ of these tori depend on the resonance $-l_0/k_0$ and the distance to the separatrix, which is measured in terms of E. Thus, on the one hand tori are bent when they approach the separatrix, that is, when $E \to 0$, and on the other hand tori are flatter when the size $\varepsilon|(k_0, l_0)|^{-1/r}$ of the gap decreases, which is controlled by k_0 and therefore by γ (see (110) for a definition of γ).

In the following Lemma 4.7 we consider the different cases and we prove that conditions **H2"**, **H3"** and **H3"'** ensure the existence of a transversal intersection between the foliation $\mathcal{F}_F$ and its image under the scattering map $\mathcal{F}_{F\circ S_\varepsilon^{-1}}$.

Lemma 4.7. *Let us consider a connected component of the big gaps region $\mathcal{D}_{\rm BG}$ defined in (82). Recall from formula (94) together with expressions (110) and (112) that, in this component, the function F defining the foliation is of the form*

$$F(I,\varphi,s;\varepsilon)=\frac{(k_0I+l_0)^2}{2}(1+\varepsilon k_0^2\tilde{h}(k_0I+l_0;\varepsilon))+\varepsilon^\gamma\widetilde{U}(\theta;\varepsilon)+\mathcal{O}_{\mathcal{C}^2}(|k_0|^3\varepsilon^{\gamma/2+1+\eta}),\quad(147)$$

where $\theta=k_0\varphi+l_0s$, and for some $0\le\rho<\pi$ and some range of energies $-\varepsilon^\gamma\le E\le L^2$, the equation $F(I,\varphi,s;\varepsilon)=E$ defines two smooth surfaces $L_E^{F\pm}$ given as graphs $I=\lambda_E^\pm(\varphi,s;\varepsilon)$, with $\lambda_E^\pm$ given in (95), which are of the form

$$\lambda_E^\pm(\varphi,s;\varepsilon)=-\frac{l_0}{k_0}+\frac{1}{k_0}\mathcal{Y}_\pm(\theta,E)+\mathcal{O}_{\mathcal{C}^0}(\varepsilon^{1+\eta}),\qquad(148)$$

where

$$\mathcal{Y}_\pm(\theta,E)=\pm(1+\varepsilon b)\ell(\theta,E)+\varepsilon\widetilde{\mathcal{Y}}_\pm(\ell(\theta,E)),\qquad(149)$$

for $\rho\le\theta=k_0\varphi+l_0s\le2\pi-\rho$ and $\ell(\theta,E)=\sqrt{2(E-\varepsilon^\gamma\widetilde{U}(\theta;0))}$ with $\widetilde{U}(\theta;\varepsilon)$ defined in (112) and $\widetilde{\mathcal{Y}}_\pm$ satisfying (118).

Assume that hypothesis **H2"** *is fulfilled, more precisely, that the reduced Poincaré function $\mathcal{L}^*$ verifies, for any value of $(I,\varphi,s)\in H_+\cap\mathcal{D}_{\rm BG}$, that the function*

$$\widetilde{\theta}\mapsto\frac{\partial\mathcal{L}^*}{\partial\widetilde{\theta}}(I,\widetilde{\theta})\qquad(150)$$

for $\widetilde{\theta}=\varphi-Is$ is positive (resp. negative) and non-constant for $\widetilde{\theta}\in\mathcal{J}_I^*$.

For $|(k_0,l_0)|\prec\varepsilon^{-1/r}$ assume hypothesis **H3"** *on (k_0,l_0) in Theorem 2.1, which is that the function*

$$\theta\to\frac{k_0\widetilde{U}'^{k_0,l_0}(\theta;0)\frac{\partial\mathcal{L}^*}{\partial\widetilde{\theta}}\left(\frac{-l_0}{k_0},\frac{\theta}{k_0}\right)+2\widetilde{U}(\theta;0)\frac{\partial^2\mathcal{L}^*}{\partial\widetilde{\theta}^2}\left(\frac{-l_0}{k_0},\frac{\theta}{k_0}\right)}{2\frac{\partial^2\mathcal{L}^*}{\partial\widetilde{\theta}^2}\left(\frac{-l_0}{k_0},\frac{\theta}{k_0}\right)}\qquad(151)$$

is non-constant.

For $|(k_0,l_0)|\sim\varepsilon^{-1/r}$ we assume the following hypothesis, which is condition **H3"'** *on (k_0,l_0) in Theorem 2.1:*

There exists a constant C, independent of E and ε, and an interval $\mathcal{J}\subset\mathcal{J}^*_{-l_0/k_0}$ such that given any E,ε in this region and $\theta\in\mathcal{J}$,

$$\begin{aligned}\Bigg|&\frac{1}{2(E-\varepsilon^\gamma\widetilde{U}(\theta;0))}\Bigg(2E\frac{\partial^2\mathcal{L}^*}{\partial\widetilde{\theta}^2}(-\frac{l_0}{k_0},\frac{\theta}{k_0})\\&-\varepsilon^\gamma\left[k_0\widetilde{U}'^{k_0,l_0}(\theta;0)\frac{\partial\mathcal{L}^*}{\partial\widetilde{\theta}}(-\frac{l_0}{k_0},\frac{\theta}{k_0})+2\widetilde{U}(\theta;0)\frac{\partial^2\mathcal{L}^*}{\partial\widetilde{\theta}^2}(-\frac{l_0}{k_0},\frac{\theta}{k_0})\right]\\&\pm\varepsilon k_0\sqrt{2(E-\varepsilon^\gamma\widetilde{U}(\theta;0))}\frac{\partial\mathcal{L}^*}{\partial\widetilde{\theta}}(-\frac{l_0}{k_0},\frac{\theta}{k_0})\frac{\partial^2\mathcal{L}^*}{\partial\widetilde{\theta}^2}(-\frac{l_0}{k_0},\frac{\theta}{k_0})\Bigg)\Bigg|\ge C.\end{aligned}\qquad(152)$$

Then, the foliations $\mathcal{F}_F$ and $\mathcal{F}_{F\circ S_\varepsilon^{-1}}$ intersect transversally.

More precisely, any surface $S_\varepsilon(L_E^{F,-})$ intersects at some point the surface $L_{E'}^{F,-}$ for any $E'<E$ (resp. $E'>E$) such that $|E'-E|\le C|k_0|\varepsilon\max(|E|^{1/2},\varepsilon^{\gamma/2})$. Analogously, any surface $S_\varepsilon(L_E^{F,+})$ intersects at some point the surface $L_{E'}^{F,+}$ for any $E'>E$ (resp.

$E' < E$) such that $|E'-E| \leq C|k_0|\varepsilon \max(|E|^{1/2}, \varepsilon^{\gamma/2})$. (In some cases, it is also possible that a certain surface $S_\varepsilon(L_E^{F,-})$ intersects the surface $L_{E'}^{F,+}$ with E' such that $|E'-E| \leq C|k_0|\varepsilon \max(|E|^{1/2}, \varepsilon^{\gamma/2})$).

The angle between the surfaces $S_\varepsilon(L_E^{F,\pm})$ and $L_{E'}^{F,\pm}$ at the intersection is bounded from below by $C\varepsilon$, where C is a constant independent of ε and E.

Remark 4.8. Lemma 10.16 in [DLS06a] gives a computable sufficient condition that guarantees that hypothesis **H3"'** on (k_0, l_0) is verified independently of ε and E. Indeed, let

$$a(\theta) = \frac{\partial^2 \mathcal{L}^*}{\partial \widetilde{\theta}^2}(-\frac{l_0}{k_0}, \frac{\theta}{k_0}),$$

$$b(\theta) = -\frac{1}{2}\left(k_0 \widetilde{U}'^{k_0,l_0}(\theta;0)\frac{\partial \mathcal{L}^*}{\partial \widetilde{\theta}}(-\frac{l_0}{k_0}, \frac{\theta}{k_0}) + 2\widetilde{U}(\theta;0)\frac{\partial^2 \mathcal{L}^*}{\partial \widetilde{\theta}^2}(-\frac{l_0}{k_0}, \frac{\theta}{k_0})\right),$$

$$c(\theta) = \pm\frac{\sqrt{2}}{2}\frac{\partial \mathcal{L}^*}{\partial \widetilde{\theta}}(-\frac{l_0}{k_0}, \frac{\theta}{k_0})\frac{\partial^2 \mathcal{L}^*}{\partial \widetilde{\theta}^2}(-\frac{l_0}{k_0}, \frac{\theta}{k_0}),$$

if there exist θ_1, θ_2 and θ_3 in some interval $\mathcal{J}$ verifying

$$\begin{vmatrix} \widetilde{a}(\theta_1) & \widetilde{a}(\theta_2) & \widetilde{a}(\theta_3) \\ \widetilde{b}(\theta_1) & \widetilde{b}(\theta_2) & \widetilde{b}(\theta_3) \\ \widetilde{c}(\theta_1) & \widetilde{c}(\theta_2) & \widetilde{c}(\theta_3) \end{vmatrix} \neq 0, \tag{153}$$

where

$$\begin{aligned} \widetilde{a}(\theta) &= a(\theta)^2 \\ \widetilde{b}(\theta) &= 2a(\theta)b(\theta) - c(\theta)^2 \\ \widetilde{c}(\theta) &= b(\theta)^2 - c(\theta)^2\widetilde{U}(\theta;0), \end{aligned} \tag{154}$$

then there exists a constant C and three intervals $\theta_i \in \mathcal{J}_i \subset \mathcal{J}$, $i = 1,2,3$ such that for any $\theta \in \mathcal{J}_i$

$$\left|\frac{a(\theta)E + b(\theta)\varepsilon^\gamma + c(\theta)\varepsilon k_0\sqrt{E - \varepsilon^\gamma \widetilde{U}(\theta;0)}}{E - \varepsilon^\gamma \widetilde{U}(\theta;0)}\right| \geq C,$$

which is hypothesis **H3"'** on (k_0, l_0).

Proof: We will apply Lemma 4.2 to the foliation $\mathcal{F}_F$ given by the function F in (147).

We first apply the scattering map to the implicit surface

$$L_E^F = \{(I, \varphi, s) \in \mathcal{D}_{\mathrm{BG}}, F(I, \varphi, s; \varepsilon) = E\},$$

and recall that $S_\varepsilon(L_E^F)$ intersects a leaf $L_{E'}^F$ at a point $(I, \varphi, s) \in L_E^F$ if $F \circ S_\varepsilon(I, \varphi, s; \varepsilon) = E'$, where, using expression (141) with $\mathcal{S}_0 = -\mathcal{L}^*$, $F \circ S_\varepsilon$ on L_E^F is given by

$$(F \circ S_\varepsilon)(I, \varphi, s; \varepsilon) = E - \varepsilon\{F, \mathcal{L}^*\} + \frac{\varepsilon^2}{2}(\{\{F, \mathcal{L}^*\}, \mathcal{L}^*\} + \{F, \mathcal{S}_1\}) + \mathcal{O}(\varepsilon^3). \tag{155}$$

Notice that the terms in expression (155) involve the derivatives of F on L_E^F. Using the expression for F in (147) and the expression of the leaf $L_E^{F,\pm}$ as a graph of I over the angle variables given in (148), we have that

$$\begin{aligned}\frac{\partial F}{\partial I}(I,\varphi,s;\varepsilon) &= k_0(k_0I+l_0)(1+\varepsilon k_0^2\tilde{h}(k_0I+l_0;\varepsilon)) + \frac{(k_0I+l_0)^2}{2}\varepsilon k_0^3\tilde{h}'(k_0I+l_0;\varepsilon)\\ &\quad + \mathcal{O}_{\mathcal{C}^1}(|k_0|^3\varepsilon^{\gamma/2+1+\eta})\\ &= \pm k_0\ell(\theta,E) + \mathcal{O}(|k_0|^3\varepsilon|\ell| + k_0^2\varepsilon^{1+\eta}) \end{aligned}\tag{156}$$

and

$$\frac{\partial F}{\partial \varphi}(I,\varphi,s;\varepsilon) = \varepsilon^\gamma k_0 \widetilde{U}'^{k_0,l_0}(\theta,\varepsilon) + \mathcal{O}_{\mathcal{C}^1}(|k_0|^3\varepsilon^{\gamma/2+1+\eta}).\tag{157}$$

Hence,

$$\begin{aligned}\{F,\mathcal{L}^*\} &= -\frac{\partial F}{\partial I}\frac{\partial \mathcal{L}^*}{\partial \varphi} + \frac{\partial F}{\partial \varphi}\frac{\partial \mathcal{L}^*}{\partial I}\\ &= \mp k_0\ell(\theta,E)\frac{\partial \mathcal{L}^*}{\partial \widetilde{\theta}}\left(-\frac{l_0}{k_0} + \frac{1}{k_0}\ell(\theta,E), \varphi - \left(-\frac{l_0}{k_0}+\frac{1}{k_0}\ell(\theta,E)\right)s\right)\\ &\quad + \mathcal{O}(|k_0|^2\varepsilon|\ell| + k_0^2\varepsilon^{1+\eta} + |k_0|\varepsilon^\gamma). \end{aligned}\tag{158}$$

Regarding the term of order ε^2 in the expression (155), we will see that among all the terms in $\varepsilon^2/2(\{\{F,\mathcal{L}^*\},\mathcal{L}^*\} + \{F,\mathcal{S}_1\})$ there is a dominant one. To that end we notice first that all the terms that appear in the derivatives up to second order for F with respect to (I,φ,s) on $L_E^{F,\pm}$ are $\mathcal{O}(|k_0||\ell|, k_0^2\varepsilon^\gamma)$, except

$$\frac{\partial^2 F}{\partial I^2} = k_0^2(1+\mathcal{O}(|k_0|^2\varepsilon)).\tag{159}$$

Hence, in the expression $\{\{F,\mathcal{L}^*\},\mathcal{L}^*\}+\{F,\mathcal{S}_1\}$ on $L_E^{F,\pm}$, all the terms are of order $k_0^2\varepsilon^\varrho$, for some $\varrho>0$, except

$$\frac{\partial^2 F}{\partial I^2}\left(\frac{\partial \mathcal{L}^*}{\partial \widetilde{\theta}}(I,\widetilde{\theta})\right)^2.$$

Therefore, using this feature and (158), the expression (155) for $F\circ S_\varepsilon$ on $L_E^{F,\pm}$, is given by

$$\begin{aligned}F\circ S_\varepsilon(I,\varphi,s;\varepsilon) &= E \pm \varepsilon k_0\ell(\theta,E)\frac{\partial \mathcal{L}^*}{\partial \widetilde{\theta}}(-\frac{l_0}{k_0},\frac{\theta}{k_0})\\ &\quad + \frac{\varepsilon^2}{2}k_0^2\left(\frac{\partial \mathcal{L}^*}{\partial \widetilde{\theta}}(-\frac{l_0}{k_0},\frac{\theta}{k_0})\right)^2 + \mathcal{O}(k_0^2\varepsilon^{2+\varrho}, |k_0|\varepsilon^{\gamma+1}, \varepsilon|\ell|^2) \\ &= E + \varepsilon\mathcal{M}_\pm(\theta;\varepsilon) + \mathcal{O}(k_0^2\varepsilon^{2+\varrho}, |k_0|\varepsilon^{\gamma+1},\varepsilon|\ell|^2), \end{aligned}\tag{160}$$

where

$$\mathcal{M}_\pm(\theta;\varepsilon) = k_0\frac{\partial \mathcal{L}^*}{\partial \widetilde{\theta}}(-\frac{l_0}{k_0},\frac{\theta}{k_0})\left(\pm\sqrt{2(E-\varepsilon^\gamma\widetilde{U}(\theta;0))} + \varepsilon k_0\frac{1}{2}\frac{\partial \mathcal{L}^*}{\partial \widetilde{\theta}}(-\frac{l_0}{k_0},\frac{\theta}{k_0})\right).\tag{161}$$

Therefore, the size of the heteroclinic jumps provided by the scattering map is determined by the size of the term $\mathcal{M}_\pm$ in (161).

In order to check the transversality of the heteroclinic intersections we use condition (142), which involves, in any case, the computation of the Poisson bracket $\{F,\{F,\mathcal{L}^*\}\}$ and the gradient of F (see formula (143)).

From expressions (156) and (157) it follows that on $L_E^{F,\pm}$,

$$\nabla F(I,\varphi,s;\varepsilon) = k_0\ell(\theta,E) + \mathcal{O}(|k_0|^3\varepsilon|\ell| + k_0^2\varepsilon^{1+\eta} + |k_0|\varepsilon^\gamma)). \tag{162}$$

On the other hand, the computation of $\{F,\{F,\mathcal{L}^*\}\}$ involves several terms. However, using the expression for $\{F,\mathcal{L}^*\}$ obtained in (158) and the expression and estimates for the derivatives up to second order for F with respect to (I,φ,s) given in (156)-(157)-(159), one can see that the dominant terms in $\{F,\{F,\mathcal{L}^*\}\}$ involve

$$\left(\frac{\partial F}{\partial I}\right)^2 \frac{\partial^2\mathcal{L}^*}{\partial\varphi^2} - \frac{\partial F}{\partial\varphi}\frac{\partial^2 F}{\partial I^2}\frac{\partial\mathcal{L}^*}{\partial\varphi}$$

and therefore we have that on $L_E^{F,\pm}$,

$$\begin{aligned}\{F,\{F,\mathcal{L}^*\}\} = {} & k_0^2\ell(\theta,E)^2\frac{\partial^2\mathcal{L}^*}{\partial\widetilde{\theta}^2}(\widetilde{\theta},E) - k_0\varepsilon^\gamma\widetilde{U}'^{k_0,l_0}(\theta,\varepsilon)k_0^2\frac{\partial\mathcal{L}^*}{\partial\widetilde{\theta}}(\widetilde{\theta},E)\\ & + \mathcal{O}(|k_0|^2|\ell|(\varepsilon^\gamma + |k_0|^2\varepsilon|\ell| + |k_0|^2\varepsilon^{1+\eta})).\end{aligned} \tag{163}$$

In the expression (161) there appear two quantities that can be comparable or not depending on k_0 and E. Notice first that $|\ell(\cdot,E)| = \max(E^{1/2},\varepsilon^{\gamma/2})$, with $1\le\gamma<2+\nu$, for some $\nu>0$. In consequence, when the size of the energy is big ($|E|>\varepsilon^\gamma$), we have $\ell(\theta;E) = \mathcal{O}(E^{1/2})$ and therefore the term involving $\ell(\theta,E)$ in expression (161) dominates. On the other hand, if the energy is small, that is $|E|$ is smaller than or comparable to $\varepsilon^{\gamma/2}$, then $\ell(\theta;E) = \mathcal{O}(\varepsilon^{\gamma/2})$, which by expression (110) is also $\mathcal{O}(|k_0|\varepsilon^{1+\nu/2})$, for some $\nu>0$. In this case we, the dominant term in expression (161) will depend on the size of k_0.

Hence, we choose μ such that $0<\mu<\gamma$ and we distinguish two cases: the case when tori are close to the resonance, which corresponds to small values of the energy ($-\varepsilon^\gamma\le E\le\varepsilon^\mu$) and the case when they are reasonably far from a resonance, which corresponds to greater values of the energy ($\varepsilon^\mu\le E\le L^2$).

Far from the resonance: $\varepsilon^\mu\le E\le L^2$.

The case far from a resonance is analogous to the flat tori region, studied in the previous section, because in this case

$$\begin{aligned}\ell(\theta,E) &= \sqrt{2(E-\varepsilon^\gamma\widetilde{U}(\theta;\varepsilon))} = \sqrt{2E}\sqrt{1-\frac{\varepsilon^\gamma}{E}\widetilde{U}(\theta;\varepsilon)}\\ &= \sqrt{2E}(1+\mathcal{O}(\varepsilon^{\gamma-\mu})).\end{aligned}$$

Consequently, since $\sqrt{2E}\ge\sqrt{2}\varepsilon^{\mu/2}$ and $\varepsilon^{\mu/2}>\varepsilon^{\gamma/2}\ge|k_0|\varepsilon^{1/2}$, the expression (160) can be written as

$$F\circ S_\varepsilon(I,\varphi,s;\varepsilon) = E\pm\varepsilon k_0\sqrt{2E}\frac{\partial\mathcal{L}^*}{\partial\widetilde{\theta}}\left(-\frac{l_0}{k_0},\frac{\theta}{k_0}\right) + \mathcal{O}(k_0|E|^{1/2}\varepsilon^{1+\gamma-\mu},\varepsilon|E|).$$

Therefore, by the hypothesis **H2"** on $\frac{\partial\mathcal{L}^*}{\partial\widetilde{\theta}}(I,\widetilde{\theta})$, we have that image of $L_E^{F,\pm}$ under the scattering map, for E large, intersects all surfaces $L_{E'}^{F,\pm}$ such that $|E'-E| = \mathcal{O}(\varepsilon|k_0||E|^{1/2})$.

In order to prove the transversality of intersections, we need to check condition (144). Using that the term involving $\ell(\theta,E)$ is the dominant one in expression (163) for

$\{F, \{F, \mathcal{L}^*\}\}$ and the expression (162) for ∇F, condition (144) for the transversality of the intersections is

$$\left| \frac{\partial^2 \mathcal{L}^*}{\partial \widetilde{\theta}^2} \right| \geq C > 0,$$

which is clearly satisfied by the hypothesis on **H2"** on $\frac{\partial \mathcal{L}^*}{\partial \widetilde{\theta}}(I, \widetilde{\theta})$.

Close to the resonance: $-\varepsilon^\gamma \leq E \leq \varepsilon^\mu$.

The case close to a resonance is more technical because the size of the energy is now comparable to the term $\varepsilon^\gamma \widetilde{U}$ and therefore $\ell(\cdot, E) = \mathcal{O}(\varepsilon^{\gamma/2})$. Hence, in expression (161) there appear two quantities that can be comparable or not depending on k_0. On the one hand, there is $\sqrt{2(E - \varepsilon^\gamma \widetilde{U}(\theta; 0))}$, which is related to the size of the gap and the other one there is $\varepsilon k_0 \frac{1}{2} \frac{\partial \mathcal{L}^*}{\partial \widetilde{\theta}}(-\frac{l_0}{k_0}, \frac{\theta}{k_0})$, which is related to the size of the heteroclinic jumps provided by the scattering map S_ε. Hence we distinguish three situations depending on k_0:

i. If $\varepsilon^{\gamma/2} \prec k_0 \varepsilon$, that is $|(k_0, l_0)| \succ \varepsilon^{-1/r}$ (see definition for γ in (110)) we have that the expression (160) reduces to

$$F \circ S_\varepsilon(I, \varphi, s; \varepsilon) = E + \frac{\varepsilon^2}{2} k_0^2 \left(\frac{\partial \mathcal{L}^*}{\partial \widetilde{\theta}} \left(-\frac{l_0}{k_0}, \frac{\theta}{k_0} \right) \right)^2 + \mathcal{O}(k_0^2 \varepsilon^{2+\varrho}),$$

for any $\varrho > 0$. So, tori are essentially flat and this is equivalent to the flat tori case. Hence, condition **H2"** assures that the foliations intersect transversally.

ii. If $k_0 \varepsilon \prec \varepsilon^{\gamma/2}$, that is $|(k_0, l_0)| \prec \varepsilon^{-1/r}$ (see definition for γ in (110)), we have that the expression (160) reduces to

$$F \circ S_\varepsilon(I, \varphi, s; \varepsilon) = E \pm \varepsilon k_0 \sqrt{2(E - \varepsilon^\gamma \widetilde{U}(\theta; 0))} \frac{\partial \mathcal{L}^*}{\partial \widetilde{\theta}} \left(-\frac{l_0}{k_0}, \frac{\theta}{k_0} \right) + \mathcal{O}(k_0^2 \varepsilon^2). \qquad (164)$$

This is the case when the size of the gaps in the foliation of primary tori is bigger than the size of the heteroclinic jumps provided by the scattering map. Hence, if we consider the surface $L_E^{F,-}$, by hypothesis **H2"** we have that

$$-\varepsilon k_0 \sqrt{2(E - \varepsilon^\gamma \widetilde{U}(\theta; 0))} \frac{\partial \mathcal{L}^*}{\partial \widetilde{\theta}} \left(-\frac{l_0}{k_0}, \frac{\theta}{k_0} \right)$$

is a negative function, and therefore by equation (164) $S(L_E^{F,-})$ intersects surfaces $L_{E'}^{F,-}$ with $E' < E$ (resp. $E' > E$) such that $|E' - E| \preceq |k_0| \varepsilon^{1+\gamma/2}$. An analogous result is obtained for $L_E^{F,+}$ with $E' > E$ (resp. $E' < E$).

iii. If $\varepsilon^{\gamma/2} \sim k_0 \varepsilon$, which is the case when $|(k_0, l_0)| \sim \varepsilon^{-1/r}$ we have that the terms $\sqrt{2(E - \varepsilon^\gamma \widetilde{U}(\theta; 0))}$ and $\frac{1}{2} \varepsilon k_0 \frac{\partial \mathcal{L}^*}{\partial \widetilde{\theta}}(-\frac{l_0}{k_0}, \frac{\theta}{k_0})$ in the expression (161) are comparable. This case is the hardest to study because the size of the gap has the same order than the heteroclinic jumps. This causes that there are different geometries for $S_\varepsilon(L_E^{F,\pm})$ that could happen depending on the numerical values of the leading coefficients.
We focus in the case of $S_\varepsilon(L_E^{F,-})$ and the function (150) positive. The case for $S_\varepsilon(L_E^{F,+})$ and the function (150) negative is analogous. Hence, by hypothesis **H2"**, the main term $\mathcal{M}_-$ in F given in (161) can have different signs depending on the size of $\ell(\theta; \varepsilon)$. According to that, we distinguish the following cases:

(a) The first case is when

$$\left|-\sqrt{2(E-\varepsilon^\gamma \widetilde{U}(\theta;0))}+\varepsilon k_0 \frac{1}{2}\frac{\partial \mathcal{L}^*}{\partial \widetilde{\theta}}(-\frac{l_0}{k_0},\frac{\theta}{k_0})\right|_{\mathcal{C}^1} \leq \varepsilon^{1+\varrho}.$$

This case corresponds to points in $L_E^{F,-}$ that are $\mathcal{O}_{\mathcal{C}^2}(\varepsilon^{2+\varrho})$-close to homoclinic jumps $S_\varepsilon(L_E^{F,-}) \pitchfork L_E^{F,-}$. They are not good for diffusion.

(b) The second case is when

$$\left|-\sqrt{2(E-\varepsilon^\gamma \widetilde{U}(\theta;0))}+\varepsilon k_0 \frac{1}{2}\frac{\partial \mathcal{L}^*}{\partial \widetilde{\theta}}(-\frac{l_0}{k_0},\frac{\theta}{k_0})\right|_{\mathcal{C}^1} > \varepsilon^{1+\varrho}.$$

This case corresponds to points in heteroclinic jumps $S_\varepsilon(L_E^{F,-}) \pitchfork L_{E'}^{F,-}$ and we can distinguish two situations that can take place.

On the one hand, if

$$-\sqrt{2(E-\varepsilon^\gamma \widetilde{U}(\theta;0))} > \varepsilon k_0 \frac{1}{2}\frac{\partial \mathcal{L}^*}{\partial \widetilde{\theta}},$$

which is the case when the heteroclinic jumps are smaller than the gap, $S_\varepsilon(L_E^{F,-})$ intersects surfaces $L_{E'}^{F,-}$ with $E' < E$ and $|E'-E| \preceq |k_0|\varepsilon^{1+\gamma/2}$. Thus, for small values of energy $E>0$, the scattering map will connect a surface with energy $E>0$ with a surface $E'<0$, which corresponds to a heteroclinic connection of a primary tori with a secondary one.

On the other hand, when

$$-\sqrt{2(E-\varepsilon^\gamma \widetilde{U}(\theta;0))} < \varepsilon k_0 \frac{1}{2}\frac{\partial \mathcal{L}^*}{\partial \widetilde{\theta}},$$

which is the case when the heteroclinic jumps are bigger than the gaps created between primary tori, we obtain that $S_\varepsilon(L_E^{F,-})$ will intersect the surfaces $L_{E'}^{F,-}$ with $E' > E$ (resp. $E' < E$) and $|E'-E| \preceq |k_0|\varepsilon^{1+\gamma/2}$. In this case the scattering map will connect two tori with positive energy, that is, two primary tori, and cross the gap with just one application of the scattering map.

Once we have a heteroclinic connection that crosses the separatrix loop, we can consider $S_\varepsilon(L_E^{F,+})$, which corresponds to the upper branch of the level set $F(I,\varphi,s;\varepsilon)=E$, $E>0$. In this case, by hypothesis **H2"**, in expression (160) the main term $\mathcal{M}_+$ in F given in (161) is always positive, so $S_\varepsilon(L_E^{F,+})$ will intersect surfaces $L_{E'}^{F,+}$ with $E'>E$ (resp. $E'<E$)and $|E'-E| \preceq |k_0|\varepsilon^{1+\gamma/2}$.

Now, we want to check that the intersections for the cases (ii) and (iii) take place transversally by means of condition (142). For the case described in item (ii) in this proof, condition (144) implies condition (142). So, using expression (163) for $\{F,\{F,\mathcal{L}^*\}\}$ and expression (162) for ∇F on L_E^F, we have that the condition (142) is satisfied provided that

$$\left|\frac{\pm 1}{2(E-\varepsilon^\gamma \widetilde{U}(\theta;0))}\left(2E\frac{\partial^2 \mathcal{L}^*}{\partial \widetilde{\theta}^2}(-\frac{l_0}{k_0},\frac{\theta}{k_0}) \right.\right.$$
$$\left.\left. -\varepsilon^\gamma\left[k_0 \widetilde{U}'^{k_0,l_0}(\theta;0)\frac{\partial \mathcal{L}^*}{\partial \widetilde{\theta}}(-\frac{l_0}{k_0},\frac{\theta}{k_0}) + 2\widetilde{U}(\theta;0)\frac{\partial^2 \mathcal{L}^*}{\partial \widetilde{\theta}^2}(-\frac{l_0}{k_0},\frac{\theta}{k_0})\right]\right)\right| \geq C$$

By Lemma 10.10 in [DLS06a], hypothesis **H3"** on (k_0, l_0) implies the previous condition and therefore the angle between the surfaces $S_\varepsilon(L_E^F)$ and $L_{E'}^F$ at the intersection can be bounded from below by $C\varepsilon$, for some suitable constant independent of ε.

For the particular case $|(k_0, l_0)| \sim \varepsilon^{-1/r}$ described in item (iii), we need to check condition (143). Using the expression (161) for the dominant term $\mathcal{M}_\pm$ in F, it is not difficult to see that the dominant term in the numerator of (142) involves the terms

$$\begin{aligned}
&\left(\frac{\partial F}{\partial I}\right)^2 \frac{\partial^2 \mathcal{L}^*}{\partial \varphi^2} - \frac{\partial F}{\partial \varphi}\frac{\partial^2 F}{\partial I^2}\frac{\partial \mathcal{L}^*}{\partial \varphi} + \varepsilon \frac{\partial F}{\partial I}\frac{\partial \mathcal{L}^*}{\partial \varphi}\frac{\partial^2 \mathcal{L}^*}{\partial \varphi^2}\\
&= (k_0 I + l_0)^2 k_0^2 \frac{\partial^2 \mathcal{L}^*}{\partial \widetilde{\theta}^2}(\widetilde{\theta}, E) - \varepsilon^\gamma \widetilde{U}'^{k_0,l_0}(\theta,\varepsilon) k_0 k_0^2 \frac{\partial \mathcal{L}^*}{\partial \widetilde{\theta}}(\widetilde{\theta}, E)\\
&\quad + \varepsilon (k_0 I + l_0) k_0^2 \frac{\partial \mathcal{L}^*}{\partial \varphi}(\widetilde{\theta}, E)\frac{\partial^2 \mathcal{L}^*}{\partial \varphi^2}(\widetilde{\theta}, E)
\end{aligned}$$

Using the expression for ∇F in (162), we have that the condition (142) is satisfied provided that

$$\begin{aligned}
\Bigg| &\frac{\pm 1}{2(E - \varepsilon^\gamma \widetilde{U}(\theta;0))}\Bigg(2E\frac{\partial^2 \mathcal{L}^*}{\partial \widetilde{\theta}^2}(-\frac{l_0}{k_0}, \frac{\theta}{k_0})\\
&-\varepsilon^\gamma \left[k_0 \widetilde{U}'^{k_0,l_0}(\theta;0)\frac{\partial \mathcal{L}^*}{\partial \widetilde{\theta}}(-\frac{l_0}{k_0}, \frac{\theta}{k_0}) + 2\widetilde{U}(\theta;0)\frac{\partial^2 \mathcal{L}^*}{\partial \widetilde{\theta}^2}(-\frac{l_0}{k_0}, \frac{\theta}{k_0})\right]\\
&\pm \varepsilon k_0 \sqrt{2(E - \varepsilon^\gamma \widetilde{U}(\theta;0))}\frac{\partial \mathcal{L}^*}{\partial \widetilde{\theta}}(-\frac{l_0}{k_0}, \frac{\theta}{k_0})\frac{\partial^2 \mathcal{L}^*}{\partial \widetilde{\theta}^2}(-\frac{l_0}{k_0}, \frac{\theta}{k_0})\Bigg)\Bigg| \geq C,
\end{aligned}$$

for some constant C. By hypothesis **H3"'** on (k_0, l_0) in Theorem 2.1 we know that the previous condition is satisfied for $\theta \in \mathcal{J} \subset \mathcal{J}^*_{-l_0/k_0}$. Consequently, the angle of intersection can be bounded again from below by $C\varepsilon$, for some suitable constant independent of ε. □

Remark 4.9. By Theorem 3.1 we know that the tori in a connected component of the big gaps region are given by the expression $I = \lambda_E^\pm(\varphi, s; \varepsilon)$, for $E = E_i$ and $-\varepsilon^\gamma \leq E_i \leq L^2$, with $\lambda_E^\pm$ given in (95). Moreover, they satisfy

$$|E_i - E_{i+1}| \leq |k_0| \varepsilon^{\gamma/2+1+\eta} \leq |k_0| \max(|E_i|^{1/2}, \varepsilon^{\gamma/2})$$

and they are $\mathcal{O}(\varepsilon^{1+\eta})$-closely spaced, in terms of the I variable.

Hence, we conclude that the image under the scattering map of a torus $\mathcal{T}_i$, $S_\varepsilon(\mathcal{T}_i)$ in the big gaps region, given by $I = \lambda_{E_i}^\pm(\varphi, s; \varepsilon)$, intersects transversally another torus $\mathcal{T}_{i+1}$ of this region given by $\bar{I} = \lambda_{E_{i+1}}^\pm(I, \varphi, s; \varepsilon)$, with $|E_{i+1} - E_i| = \mathcal{O}(\varepsilon^{\gamma/2+1+\eta})$ (equivalently $|I - \bar{I}| \leq \varepsilon^{1+\eta}$):

$$S_\varepsilon(\mathcal{T}_i) \pitchfork \mathcal{T}_{i+1}.$$

4.2. Proof of Proposition 4.1

The proof is just a combination of the results obtained in Section 4.1.

We start with a torus $\mathcal{T}_0$, which is $\mathcal{O}(\varepsilon^{1+\eta})$-close to the submanifold $I = I_-$. Assume that this torus belongs to the flat tori region with averaged energy E_0. The case when $\mathcal{T}_0$ belongs to a big gaps region is analogous. Then, we apply Lemma 4.5 and Remark 4.6 and we get that $S_\varepsilon(\mathcal{T}_0)$ intersects transversally all primary tori with averaged energy in the mentioned interval $(E_0 - C\varepsilon, E_0 + C\varepsilon)$. We pick a primary KAM torus $\mathcal{T}_1$ provided by Theorem 3.1 with energy E_1 in the interval and we repeat the argument until we reach a big gaps region. Assuming that we have applied it K times, we have that the torus $\mathcal{T}_0$ has heteroclinic connections with all the tori whose energy lies in the interval $(E_0 - KC\varepsilon, E_0 + KC\varepsilon)$, or equivalently, in the interval $(I_- - K^*C\varepsilon, I_- + K^*C\varepsilon)$ in terms of action variables.

When the domain $(I_- - K^*C\varepsilon, I_- + K^*C\varepsilon) \times \mathbb{T}^2$ for which the torus $\mathcal{T}_0$ has a heteroclinic connection overlaps with a big gaps region $[-l_0/k_0 - L_{k_0}, -l_0/k_0 + L_{k_0}] \times \mathbb{T}^2$ we use Lemma 4.7 and Remark 4.9 to show that we can cross the gap created by the resonance $-l_0/k_0$ just connecting either a primary KAM torus with a secondary one and again with a primary one or two primary KAM tori. Hence, we can construct a piece of chain that starts in $\mathcal{T}_0$ and reaches all the way to $\mathcal{T}_i$, where $\mathcal{T}_i$ is a primary KAM torus whose equation is $I = -l_0/k_0 + L_{k_0} + \mathcal{O}(\varepsilon)$ and is contained again in the flat tori region.

Therefore, we can keep constructing a transition chain just repeating the procedure stated before for the primary KAM torus $\mathcal{T}_i$ until we reach $\mathcal{T}_{N(\varepsilon)}$. □

5. Example

Consider the Hamiltonian

$$H_\varepsilon(p, q, I, \varphi, t) = \pm \left(\frac{p^2}{2} + \cos q - 1 \right) + \frac{I^2}{2} + \varepsilon \cos q \; g(\varphi, t), \tag{165}$$

which is a generalization of the famous example introduced by V.I. Arnol'd in [Arn64]. This is the same Hamiltonian in the example discussed in [DLS06a], except that the function g is chosen as a periodic function with an *infinite number of harmonics* in the angles (φ, t),

$$g(\varphi, t) = \sum_{(k,l)\in\mathbb{N}^2} a_{k,l} \cos(k\varphi + lt), \tag{166}$$

where, for simplicity, we have chosen g to be an even function and with an explicit formula for its Fourier coefficients, say $a_{k,l} = \rho^k r^l$ and $0 < \rho, r < 1$ real numbers to be chosen small enough. Notice that

$$g(\varphi, t) = \frac{1 + \rho r \cos(\varphi + t) - \rho \cos\varphi - r \cos t}{(1 - 2\rho\cos\varphi + \rho^2)(1 - 2r\cos t + r^2)}.$$

The Hamiltonian of one degree of freedom $P_\pm(p, q) = \pm\,(p^2/2 + \cos q - 1)$ is the standard pendulum when we choose the $+$ sign, and its separatrix for positive p is given by

$$q_0(t) = 4\arctan e^{\pm t}, \quad p_0(t) = 2/\cosh t.$$

An important feature of the Hamiltonian (165) is that the 3-dimensional NHIM

$$\tilde{\Lambda} = \{(0,0,I,\varphi,s) : (I,\varphi,s) \in \mathbb{R} \times \mathbb{T}^2\}$$

is preserved without any deformation for any ε: $p = q = 0 \Rightarrow \dot{p} = \dot{q} = 0$. However, in contrast with the example in [Arn64], the perturbation does not vanish on $\tilde{\Lambda}$. Indeed, the Hamiltonian (165) restricted to $\tilde{\Lambda}$ takes the form $I^2/2 + \varepsilon g(\varphi, t)$. Hence, the 2-dimensional whiskered tori

$$\mathcal{T}_I^0 = \{(0,0,I,\varphi,s) : (\varphi,s) \in \mathbb{T}^2\}$$

are not preserved for $\varepsilon \neq 0$, and resonances (47) take place at $I = -l/k$ for each $(k,l) \in \mathbb{N}^2$, $\gcd(k,l) = 1$. Therefore, we have a dense set of gaps of size $\mathcal{O}(\varepsilon^{1/2}\sqrt{a_{k,l}})$ centered at $I = -l/k$ and, among them the ones such that $\sqrt{a_{k,l}} \succcurlyeq \varepsilon^{1/2}$ give rise to resonances with big gaps and the example (165) presents the large gap problem for $I < 0$.

Hence, for any finite range of I, $[I_-, I_+] \subset \mathbb{R}^-$ we will prove the existence of diffusing orbits.

The Melnikov potential (9) of the Hamiltonian (165) is given by

$$\mathcal{L}(I,\varphi,s) = \sum_{(k,l)\in\mathbb{N}^2} A_{k,l}(I)\cos(k\varphi + ls),$$

with

$$A_{k,l}(I) = 2\pi \frac{(kI + l)}{\sinh \frac{\pi}{2}(kI + l)} a_{k,l}. \tag{167}$$

Next, we will see that for $0 < \rho < r \ll 1$ we can find open sets of $(I,\varphi,s) \in [I_-, I_+] \times \mathbb{T}^2$, such that the function $\tau \in \mathbb{R} \mapsto \mathcal{L}(I, \varphi - I\tau, s - \tau)$ has non-degenerate critical points at $\tau = \tau^*(I,\varphi,s)$ which verify the hypothesis **H2'**.

Recall that hypothesis **H2'** deals with the existence of transverse intersections of the stable and unstable manifolds of $\widetilde{\Lambda}_\varepsilon$. Hence, the non-degenerate critical points of the function $\tau \mapsto \mathcal{L}(I, \varphi - I\tau, s - \tau)$ give rise to transverse intersections.

In order to check hypothesis **H2'**, we will use the results in the example given in Section 13 of [DLS06a] by means of the following argument. Assuming that ρ, r are small enough, the function $g(\varphi, s)$ is well approximated by its truncated first order trigonometric polynomial $g^{[\leq 1]}(\varphi, s) = 1 + \rho\cos\varphi + r\cos s$. More precisely,

$$\begin{aligned} g(\varphi,s) = &\ 1 + \rho\cos\varphi + r\cos s + \mathcal{O}_2(\rho, r) \\ := &\ g^{[\leq 1]}(\varphi,s) + g^{[>1]}(\varphi,s). \end{aligned}$$

Hence, as long as $0 < \rho, r \ll 1$, if hypothesis **H2'** is verified for the trigonometric polynomial $g^{[\leq 1]}(\varphi, s)$, it will be also verified for the perturbation $g(\varphi, s)$.

Notice that the Fourier coefficients $A_{k,l}(I)$ are nothing else but the Fourier coefficients $a_{k,l}$ multiplied by a certain function depending on I that decreases exponentially as $|I|$ goes to infinity. Hence, arguing as we did for the perturbation

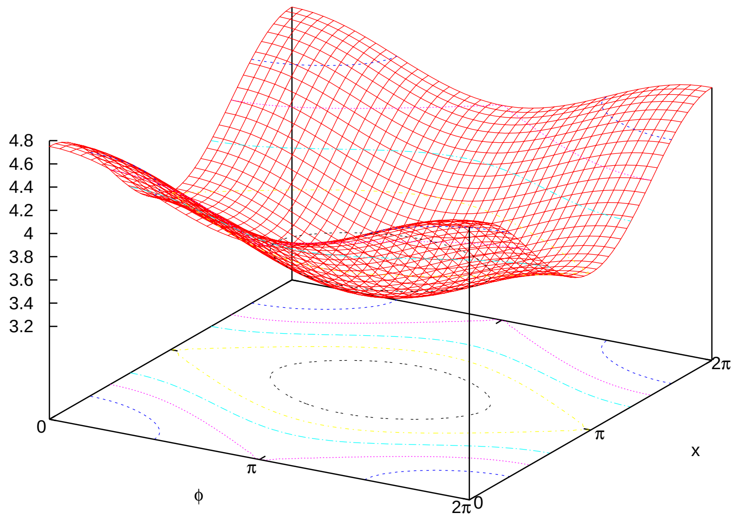

Figure 2. Graph and level curves of the Melnikov potential $\mathcal{L}^{[\leq 1]}(I,\varphi,s)$ with $\rho = 1/16$, $r = 1/8$ and $I = 0$. In this case, $A_{0,0} = 4$, $A_{1,0} = 4\rho = 1/4$ and $A_{0,1} = 4r = 1/2$.

g, we approximate the function $\mathcal{L}(I,\varphi,s)$ by its first order trigonometric polynomial $\mathcal{L}^{[\leq 1]}(I,\varphi,s) = A_{0,0} + A_{1,0}(I)\cos\varphi + A_{0,1}\cos s$, that is

$$\begin{aligned}\mathcal{L}(I,\varphi,s) = {} & A_{0,0} + A_{1,0}(I)\cos\varphi + A_{0,1}\cos s + \mathcal{O}_2(\rho,r)\\ := {} & \mathcal{L}^{[\leq 1]}(I,\varphi,s) + \mathcal{L}^{[>1]}(I,\varphi,s).\end{aligned} \tag{168}$$

Recall that we are looking for non-degenerate critical points of

$$\mathcal{L}(\tau) := \mathcal{L}(I,\varphi - I\tau, s-\tau) = \sum_{(k,l)\in\mathbb{N}^2} A_{k,l}(I)\cos(k\varphi + ls - \tau(Ik+l)), \tag{169}$$

with $A_{k,l}(I)$ as in (167).

Using that the Melnikov function $\mathcal{L}$ is well approximated by $\mathcal{L}^{[\leq 1]}$, fixed (I,φ,s), we only need to study the evolution of $\mathcal{L}^{[\leq 1]}$ along the straight lines

$$R : \tau \in \mathbb{R} \mapsto (\varphi - I\tau, s - \tau) \in \mathbb{T}^2 \tag{170}$$

on the torus.

This study has already been performed in the example in Section 13 in [DLS06a], where the reader can find more details. We just mention that since $0 < \rho < r$, for any fixed I, we have $A_{0,1} > A_{1,0}(I) > 0$ and therefore the function $(\varphi,s) \mapsto \mathcal{L}^{[\leq 1]}(I,\varphi,s)$ possesses exactly four non-degenerate critical points: a maximum at $(0,0)$, a minimum at (π,π) and two saddles at $(0,\pi)$ and $(\pi,0)$ (see Figure 2). Around the two extremum points, its level curves are closed (and indeed convex) curves which fill out a basin ending at the level curve of one of the saddle points.

Therefore, any straight line (170) that enters into some extremum basin is tangent to one of the convex closed level curves, giving rise to a non-degenerate extremum of $\tau \in \mathbb{R} \mapsto \mathcal{L}^{[\leq 1]}(I, \varphi - Is, s - \tau)$. So, degenerate extrema of $\tau \in \mathbb{R} \mapsto \mathcal{L}^{[\leq 1]}(I, \varphi - Is, s - \tau)$ can only exist for straight lines that never enter inside such extremum basins. It is clear that this never happens for irrational values of I because it implies a dense straight line (and infinite non-degenerate extrema for $\tau \in \mathbb{R} \mapsto \mathcal{L}^{[\leq 1]}(I, \varphi - Is, s - \tau)$). On the other hand, the straight lines with rational slopes enter inside both extremum basins at least twice, except for the slopes $I = 0, \pm 1$. In these cases, one can check directly, using that $A_{0,1} > A_{1,0}(I) > 0$, that the function $\tau \in \mathbb{R} \mapsto \mathcal{L}^{[\leq 1]}(I, \varphi - Is, s - \tau)$ has one non-degenerate maximum and one non-degenerate minimum in any interval of length 2π.

When we take into account $\mathcal{L}^{[>1]}$ in the Melnikov potential $\mathcal{L}$ in (168), it is clear that in the compact subset $[I_-, I_+] \times \mathbb{T}^2$, as long as $0 < \rho, r \ll 1$, the function $\tau \in \mathbb{R} \mapsto \mathcal{L}(I, \varphi - I\tau, s - \tau)$ has non-degenerate extrema, and for every I we can find a smooth function $\tau = \tau^*(I, \varphi, s)$ defined in an open set of $(\varphi, s) \in \mathbb{T}^2$.

Moreover, since $\mathcal{L}$ is periodic with respect to (φ, s) and non-constant with non-degenerate extrema along any straight line, $\partial_\varphi \mathcal{L}^*$, where $\mathcal{L}^*$ is given in (11), is also periodic and non-constant and indeed changes sign. Therefore, for every I, there exists a nonempty set $\mathcal{J}_I$ where $\partial_\varphi \mathcal{L}^* > 0$ (and a nonempty set $\mathcal{J}_I^-$ where $\partial_\varphi \mathcal{L}^* < 0$), so hypothesis **H2"** is fulfilled. Indeed the set of points where $\partial_\varphi \mathcal{L}^*$ vanishes is a discrete set.

Conditions **H3'**, **H3"** and **H3"'** can also be checked in the example (166) at the resonances $I = -l_0/k_0$.

If we consider $I = -l_0/k_0$ for any $(k_0, l_0) \in \mathbb{N}^2$, $k_0 \neq 0$ and $\gcd(k_0, l_0) = 1$, the function U^{k_0,l_0} in hypothesis **H3** on (k_0, l_0) has the following expression

$$U^{k_0,l_0}(\theta) = \sum_{t=1}^{M} a_{tk_0,tl_0} \cos(t\theta) = a_{k_0,l_0} \cos(\theta) + \mathcal{O}_2(\rho^{k_0}, r^{l_0}), \tag{171}$$

where $\theta = k_0\varphi + l_0 s$.

Therefore, $\theta_1 = 0$ and $\theta_2 = \pi$ are the unique critical points for the function $U^{k_0,l_0}(\theta)$. Hence hypothesis **H3'** on (k_0, l_0) is clearly verified.

Next, for $I = -l_0/k_0$ we want to check hypothesis **H3"** on (k_0, l_0). This condition requires to show that the function f in (13) is not constant. To that end, we will consider two values of θ and we will show that their images for this function are different. For instance, notice that the function f in (13) takes the same values as U^{k_0,l_0} evaluated on its critical points θ_1 and θ_2 as long as $\frac{\partial^2 \mathcal{L}^*}{\partial \varphi^2}(I, \theta_i/k_0) \neq 0$, for $i = 1, 2$. Hence, hypothesis **H3"** on (k_0, l_0) is clearly satisfied if the function U^{k_0,l_0} has two extrema θ_i taking different values which satisfy $\frac{\partial^2 \mathcal{L}^*}{\partial \varphi^2}(I, \theta_i/k_0) \neq 0$, which is the case as can be checked just looking at non-degenerate extrema of the function $\mathcal{L}$. They give rise to non-degenerate extrema of the function $\mathcal{L}^*$, which coincide with the ones of the function U^{k_0,l_0}.

Similarly, we can check hypothesis **H3"'** on (k_0, l_0). In this case we need to show that the determinant (153) given in Remark 4.8 does not vanish. It is clearly non-zero if we choose, for the two first columns, the two critical points θ_1 and θ_2 discussed above, and for the third column $\theta_3 \neq 0, \pi$, such that $\frac{\partial^2 \mathcal{L}^*}{\partial \varphi^2}(-l_0/k_0, \theta_3/k_0) = 0$, but otherwise $U'^{k_0,l_0}(\theta_3) \neq 0$ and $\frac{\partial \mathcal{L}^*}{\partial \varphi}(-l_0/k_0, \theta_3/k_0) \neq 0$. The existence of this point θ_3 is guaranteed by the fact that if one considers the first order trigonometric polynomial of the reduced Poincaré function $\mathcal{L}^{*[\leq 1]}$, one can see that its critical points are always non-degenerate.

Hence, we apply Theorem (2.1) and we conclude that

Proposition 5.1. *Given the Hamiltonian (165) with g as in (166), $0 < \rho < r \ll 1$ and $[I_-, I_+] \subset \mathbb{R}^-$, for $|\varepsilon| \leq \varepsilon^*(\rho, r)$ there exist orbits following the mechanism described in this paper and such that $I(0) \leq I_-$, $I(T) \geq I_+$, for any $T > 0$.*

Acknowledgments

We are very grateful to M. Gidea, R. de la Llave, T. M. Seara and C. Simó for valuable comments and suggestions.

This work has been partially supported by the Spanish Grant MTM2006-00478. G.H. has also been supported by the Spanish Fellowship AP2003-3411 and the NSF Grant DMS 0354567 during her visit to UT Austin. The final version was written while we were visiting CRM during the Research Programme *Stability and Instability in Mechanical Systems (SIMS08)*, for whose hospitality we are very grateful.

Appendix A. Double Fouries Series

Proposition A.1. *Let f be a $\mathcal{C}^r$ function with respect to $(J, \varphi, s, \varepsilon)$, $r \geq 1$ and 2π-periodic with respect to (φ, s). Then its Fourier coefficients $f_{k,l}(J, \varepsilon)$, $(k, l) \in \mathbb{Z}^2$, satisfy, for $\ell = 0, \dots, r$*

$$|f_{k,l}|_{\mathcal{C}^\ell} \leq C \frac{|f|_{\mathcal{C}^r}}{|(k,l)|^{r-\ell}}, \tag{A.1}$$

where C is a constant that depends only on r and ℓ and $|(k,l)| = \max(|k|, |l|)$.

Proof. From the expression for the Fourier coefficients of a function f

$$f_{k,l}(J; \varepsilon) = \frac{1}{(2\pi)^2} \int_{\mathbb{T}^2} f(J, \varphi, s; \varepsilon) e^{\mathrm{i}(k\varphi + ls)} \mathrm{d}\varphi \mathrm{d}s,$$

taking into account that f is $\mathcal{C}^r$ in the variables (φ, s), we can integrate $r = n + m$ times by parts (n times with respect to φ and m times with respect to s) and express the Fourier coefficient $f_{k,l}(J, \varepsilon)$, with $(k, l) \neq (0, 0)$ in the form

$$f_{k,l}(J; \varepsilon) = (-1)^r \frac{1}{(2\pi)^2} \frac{1}{(\mathrm{i}k)^n (\mathrm{i}l)^m} \int_{\mathbb{T}^2} \frac{\partial^r f(J, \varphi, s; \varepsilon)}{\partial \varphi^n \partial s^m} e^{\mathrm{i}(k\varphi + ls)} \mathrm{d}\varphi \mathrm{d}s,$$

so that,

$$|f_{k,l}|_{\mathcal{C}^0} \leq \frac{1}{|k|^n |l|^m} \left| \frac{\partial^r f}{\partial \varphi^n \partial s^m} \right|_{\mathcal{C}^0} \leq \frac{n! m! |f|_{0,r}}{|k|^n |l|^m},$$

for any $0 \leq n, m \leq r$ such that $n + m = r$, where $|\cdot|_{\mathcal{C}^\ell}$ is the standard $\mathcal{C}^\ell$ norm defined in (2) and $|\cdot|_{\ell_1,\ell_2}$ is the seminorm defined in (3). Therefore,

$$|f_{k,l}|_{\mathcal{C}^0} \leq \frac{r!|f|_{0,r}}{|(k,l)|^r} \leq \frac{r!|f|_{\mathcal{C}^r}}{|(k,l)|^r}.$$

where $|(k,l)| = \max(|k|, |l|)$.

Now, taking into account that $D^\ell f_{k,l}(J;\varepsilon)$ is the Fourier coefficient of the function $\frac{\partial^\ell f(J,\varphi,s;\varepsilon)}{\partial J^\ell}$, which is a $\mathcal{C}^{r-\ell}$ function, and using the same argument as before we have that

$$|D^\ell f_{k,l}|_{\mathcal{C}^0} \leq \frac{\ell!(r-\ell)!|f|_{\ell,r-\ell}}{|(k,l)|^{r-\ell}} \leq \frac{\ell!(r-\ell)!|f|_{\mathcal{C}^r}}{|(k,l)|^{r-\ell}}.$$

From the definition of $|\cdot|_{\mathcal{C}^\ell}$ norm in (2) we have the estimate

$$|f_{k,l}|_{\mathcal{C}^\ell} = \sum_{i=0}^{\ell} \frac{|D^i f_{k,l}|_{\mathcal{C}^0}}{i!} \leq \sum_{i=0}^{\ell} \frac{(r-i)!|f|_{\mathcal{C}^r}}{|(k,l)|^{r-i}} \leq C\frac{|f|_{\mathcal{C}^r}}{|(k,l)|^{r-\ell}},$$

where C is a constant that only depends on ℓ and r, $C = r! + (r-1)! + \ldots + (r-\ell)!$, as we wanted to see. □

We consider the truncation of its Fourier series at order M in the following way:

$$f(J,\varphi,s;\varepsilon) = f^{[\leq M]}(J,\varphi,s;\varepsilon) + f^{[>M]}(J,\varphi,s;\varepsilon),$$

where

$$f^{[\leq M]}(J,\varphi,s;\varepsilon) = \sum_{\substack{(k,l)\in\mathbb{Z}^2,\\ |k|+|l|\leq M}} f_{k,l}(J;\varepsilon)e^{\mathrm{i}(k\varphi+ls)},$$

and

$$f^{[>M]}(J,\varphi,s;\varepsilon) = \sum_{\substack{(k,l)\in\mathbb{Z}^2\\ |k|+|l|>M}} f_{k,l}(J;\varepsilon)e^{\mathrm{i}(k\varphi+ls)}.$$

Proposition A.2. *Let f be of class $\mathcal{C}^r$ with respect to $(J,\varphi,s,\varepsilon)$, $r \geq 1$ and 2π-periodic with respect to (φ,s). The M-th order remainder $f^{[>M]}$ of the Fourier series of f is bounded in the standard $\mathcal{C}^\ell$ norm, for $\ell = 0,\ldots,r-3$ by*

$$\left|f^{[>M]}\right|_{\mathcal{C}^\ell} \leq C\frac{|f|_{\mathcal{C}^r}}{M^{r-(\ell+2)}}, \tag{A.2}$$

where C is a constant that depends only on r and ℓ.

Proof. The proof is very simple and follows from the estimate (A.1) for the Fourier coefficients of a $\mathcal{C}^r$ function obtained in the previous propositon. More precisely,

$$\begin{aligned}
\left|f^{[>M]}\right|_{\mathcal{C}^\ell} &\leq \sum_{\substack{(k,l)\in\mathbb{Z}^2,\\ |k|+|l|>M}} |f_{k,l}|_{\mathcal{C}^\ell} \\
&\leq C \sum_{\substack{(k,l)\in\mathbb{Z}^2,\\ |k|+|l|>M}} \frac{|f|_{\mathcal{C}^r}}{|(k,l)|^{r-\ell}} \\
&\leq C \sum_{t=M+1}^{\infty} 4t\frac{|f|_{\mathcal{C}^r}}{t^{r-\ell}} \\
&\leq 4C\,|f|_{\mathcal{C}^r} \int_M^\infty t^{\ell-r+1}\mathrm{d}t \\
&= 4\frac{C}{r-\ell-2}\,|f|_{\mathcal{C}^r}\,M^{\ell-r+2},
\end{aligned}$$

where C is a constant that depends only on r and ℓ. □

Appendix B. Weighted norms

We consider functions $u \in \tau_M(\mathcal{I}\times\mathbb{T}^2)$, where $\mathcal{I}\subset\mathbb{R}$, introduced in (28), and we can consider the different types of norms introduced in this paper: the standard $\mathcal{C}^r$ norm introduced in (2), the Fourier norm introduced in (29) and the Fourier norm with a weight introduced in (30).

The equivalence relations between all these norms are given in the following Lemmas:

Lemma B.1. *The norms $|\cdot|_{\mathcal{C}^\ell}$ and $\|\cdot\|_{\mathcal{C}^\ell}$ defined in (2) and (29), respectively, are equivalent and satisfy the following equivalence relation for $u \in \tau_M(\mathcal{I}\times\mathbb{T}^2)$ and $0 < L \leq 1$,*

$$L^\ell|u|_{\mathcal{C}^\ell} \leq \|u\|_{\mathcal{C}^\ell,L} \leq CM^2|u|_{\mathcal{C}^\ell}$$

where C is a constant depending on ℓ.

Proof. The first inequality is obvious using that $L\leq 1$. For the second one, using again that $L\leq 1$ we have

$$|u_{k,l}|_{\mathcal{C}^n,L} = \sum_{i=0}^{n} L^i\frac{|D^iu_{k,l}|_{\mathcal{C}^0}}{i!} \leq \sum_{i=0}^{n}\frac{|D^iu_{k,l}|_{\mathcal{C}^0}}{i!} = |u_{k,l}|_{\mathcal{C}^n},$$

for $0\leq n\leq\ell$. Therefore, the result follows directly from the estimate (A.1) for the $\mathcal{C}^\ell$

norm of the Fourier coefficients of a $\mathcal{C}^r$ function u, for $\ell = 0, \ldots, r$. More precisely,

$$\begin{aligned}
\|u\|_{\mathcal{C}^\ell,L} &= \sum_{m=0}^{\ell}\sum_{n=0}^{m} 2^\ell \sum_{\substack{(k,l)\in\mathbb{Z}^2,\\ |k|+|l|\leq M}} |u_{k,l}|_{\mathcal{C}^n,L}\,|(k,l)|^{m-n} \\
&\leq \sum_{m=0}^{\ell}\sum_{n=0}^{m} 2^\ell \sum_{\substack{(k,l)\in\mathbb{Z}^2,\\ |k|+|l|\leq M}} |u_{k,l}|_{\mathcal{C}^n}\,|(k,l)|^{m-n} \\
&\leq \sum_{m=0}^{\ell}\sum_{n=0}^{m} 2^\ell \sum_{\substack{(k,l)\in\mathbb{Z}^2,\\ |k|+|l|\leq M}} \widetilde{C}\frac{|u|_{\mathcal{C}^\ell}}{|(k,l)|^{\ell-n}}|(k,l)|^{m-n} \\
&\leq \sum_{m=0}^{\ell}\sum_{n=0}^{m} 2^\ell \sum_{\substack{(k,l)\in\mathbb{Z}^2,\\ |k|+|l|\leq M}} \widetilde{C}\,|u|_{\mathcal{C}^\ell} \\
&\leq CM^2\,|u|_{\mathcal{C}^\ell}
\end{aligned}$$

as we wanted to prove. □

Lemma B.2. *For the seminorm $|\cdot|_{j,\ell-j}$ defined in (2), one has that for all $0 \leq j \leq \ell$,*

$$L^j\,|u|_{j,\ell-j} \leq \|u\|_{\mathcal{C}^\ell,L} \tag{B.1}$$

Proof. Again, It follows directly from the fact that $L < 1$ and therefore,

$$L^j|u_{k,l}|_{\mathcal{C}^n} \leq \sum_{i=0}^{n} L^i\frac{|D^i u_{k,l}|_{\mathcal{C}^0}}{i!} = |u_{k,l}|_{\mathcal{C}^n,L}.$$

for $0 \leq n \leq j$. □

Lemma B.3. *For $0 < L \leq 1$, and $0 \leq \ell \leq r$ we have that for any $u \in \tau_M(\mathcal{I}\times\mathbb{T}^2)$ and $v \in \tau_N(\mathcal{I}\times\mathbb{T}^2)$*

$$\|uv\|_{\mathcal{C}^\ell,L} \leq \|u\|_{\mathcal{C}^\ell,L}\,\|v\|_{\mathcal{C}^\ell,L}\,. \tag{B.2}$$

Proof. Let us define

$$\|u\|_{n,m} = \sum_{\substack{(k,l)\in\mathbb{Z}^2,\\ |k|+|l|\leq M}} |u_{k,l}|_{\mathcal{C}^n,L}|(k,l)|^{m-n},$$

then,

$$\|u\|_{\mathcal{C}^\ell,L} = \sum_{m=0}^{\ell}\sum_{n=0}^{m} 2^\ell\|u\|_{n,m}. \tag{B.3}$$

The $\alpha - th$ Fourier coefficient of uv, where $\alpha \in \mathbb{Z}^2$, is

$$(uv)_\alpha = \sum_{\substack{\beta\in\mathbb{Z}^2,|\beta|\leq N\\ |\alpha-\beta|\leq M}} u_{\alpha-\beta}v_\beta.$$

Using the Leibniz rule for derivatives we have

$$
\begin{aligned}
|(uv)_\alpha|_{\mathcal{C}^n,L} &= \sum_{i=0}^{n} \frac{1}{i!} L^i |D^i (uv)_\alpha|_{\mathcal{C}^0} \\
&\le \sum_{i=0}^{n} \frac{1}{i!} \sum_{\substack{\beta\in\mathbb{Z}^2, |\beta|\le N\\ |\alpha-\beta|\le M}} L^i |D^i u_{\alpha-\beta} v_\beta|_{\mathcal{C}^0} \\
&\le \sum_{i=0}^{n} \frac{1}{i!} \sum_{\substack{\beta\in\mathbb{Z}^2, |\beta|\le N\\ |\alpha-\beta|\le M}} \sum_{j=0}^{i} \begin{pmatrix} i \\ j \end{pmatrix} L^{i-j} |D^{i-j} u_{\alpha-\beta}|_{\mathcal{C}^0} L^j |D^j v_\beta|_{\mathcal{C}^0} \\
&= \sum_{i=0}^{n} \sum_{\substack{\beta\in\mathbb{Z}^2, |\beta|\le N\\ |\alpha-\beta|\le M}} \sum_{j=0}^{i} L^{i-j} \frac{|D^{i-j} u_{\alpha-\beta}|_{\mathcal{C}^0}}{(i-j)!} L^j \frac{|D^j v_\beta|_{\mathcal{C}^0}}{j!} \\
&= \sum_{\substack{\beta\in\mathbb{Z}^2, |\beta|\le N\\ |\alpha-\beta|\le M}} \sum_{i=0}^{n} \sum_{j=0}^{i} L^{i-j} \frac{|D^{i-j} u_{\alpha-\beta}|_{\mathcal{C}^0}}{(i-j)!} L^j \frac{|D^j v_\beta|_{\mathcal{C}^0}}{j!} \\
&\le \sum_{\substack{\beta\in\mathbb{Z}^2, |\beta|\le N\\ |\alpha-\beta|\le M}} |u_{\alpha-\beta}|_{\mathcal{C}^n,L} \, |v_\beta|_{\mathcal{C}^n,L} \,.
\end{aligned}
$$

On the other hand, we have

$$
\begin{aligned}
|\alpha|^{m-n} \le (|\alpha-\beta| + |\beta|)^{m-n} &= \sum_{i=0}^{m-n} \begin{pmatrix} m-n \\ i \end{pmatrix} |\alpha-\beta|^i |\beta|^{m-n-i} \\
&\le \max\left(|\alpha|^{m-n}, \sum_{i=0}^{m-n} \begin{pmatrix} m-n \\ i \end{pmatrix} |\alpha-\beta|^{m-n} |\beta|^{m-n} \right) \\
&= \max(|\alpha|^{m-n}, 2^{m-n} |\alpha-\beta|^{m-n} |\beta|^{m-n}).
\end{aligned}
$$

Hence, using these two inequalities, we have that

$$
\begin{aligned}
\|uv\|_{n,m} &= \sum_{\substack{\alpha\in\mathbb{Z}^2,\\ |\alpha|\le M+N}} |(uv)_\alpha|_{\mathcal{C}^n,L} |\alpha|^{m-n} \\
&\le \sum_{\substack{\alpha\in\mathbb{Z}^2,\\ |\alpha|\le M+N}} \sum_{\substack{\beta\in\mathbb{Z}^2, |\beta|\le N\\ |\alpha-\beta|\le M}} |u_{\alpha-\beta}|_{\mathcal{C}^n,L} \, |v_\beta|_{\mathcal{C}^n,L} \, |\alpha|^{m-n} \\
&\le \sum_{\substack{\alpha\in\mathbb{Z}^2,\\ |\alpha|\le M+N}} |u_0|_{\mathcal{C}^n,L} \, |v_\alpha|_{\mathcal{C}^n,L} \, |\alpha|^{m-n} + |u_\alpha|_{\mathcal{C}^n,L} \, |\alpha|^{m-n} \, |v_0|_{\mathcal{C}^n,L} \\
&\quad + \sum_{\substack{\beta\in\mathbb{Z}^2, |\beta|\le N\\ |\alpha-\beta|\le M}} |u_{\alpha-\beta}|_{\mathcal{C}^n,L} \, |v_\beta|_{\mathcal{C}^n,L} \, 2^{m-n} |\alpha-\beta|^{m-n} |\beta|^{m-n} \\
&\le 2^{m-n} \sum_{\substack{\alpha\in\mathbb{Z}^2,\\ |\alpha|\le M}} |u_\alpha|_{\mathcal{C}^n,L} \, |\alpha|^{m-n} \sum_{\substack{\beta\in\mathbb{Z}^2\\ |\beta|\le M}} |v_\beta|_{\mathcal{C}^n,L} \, |\beta|^{m-n} \\
&= 2^{m-n} \|u\|_{n,m} \|v\|_{n,m}.
\end{aligned}
$$

Going back to the definition of $\|uv\|_{\mathcal{C}^\ell,L}$ in (B.3), we have

$$\begin{aligned}\|uv\|_{\mathcal{C}^\ell,L} &= \sum_{m=0}^{\ell}\sum_{n=0}^{m} 2^\ell \|uv\|_{n,m} \\ &\le \sum_{m=0}^{\ell}\sum_{n=0}^{m} 2^\ell 2^{m-n} \|u\|_{n,m}\|v\|_{n,m} \\ &\le \sum_{m=0}^{\ell}\sum_{n=0}^{m} 2^\ell 2^\ell \|u\|_{n,m}\|v\|_{n,m} \\ &\le \|u\|_{\mathcal{C}^\ell,L}\,\|v\|_{\mathcal{C}^\ell,L}\,,\end{aligned}$$

as claimed. □

Appendix C. Faa-di Bruno formula

Let g be a $\mathcal{C}^s(U,V)$ function, with $U \subset \mathbb{R}$ and $g(U) \subset W \subset \mathbb{R}$ and f be a $\mathcal{C}^r(W,\mathbb{R})$ function with $r,s > 0$. Then $f \circ g$ is a $\mathcal{C}^t(U,\mathbb{R})$ function, where $t = \min(r,s)$. By a repeated application of the chain rule, one gets

$$D^\ell(f\circ g)(x) = \sum_{k=1}^{\ell}\sum_{j_1+\cdots+j_k=\ell} c_{k,j_1,\cdots,j_k} D^k f(g(x)) D^{j_1} g(x)\cdots D^{j_k} g(x), \quad \text{(C.1)}$$

for $\ell = 1,\ldots,t$, where $c_{k,j_1,\cdots,j_k}$ are combinatorial coefficients. The formula (C.1) is called **Faa-di Bruno formula** (see [LO99]).

From equation (C.1), it is easy to see that there exists a constant C_t depending on t such that

$$|f\circ g|_{\mathcal{C}^t} \le C_t\,|f|_{\mathcal{C}^t}\,|g|^t_{\mathcal{C}^t}\,. \quad \text{(C.2)}$$

Since we are interested in multi-valued functions, we introduce now a generalized bound. Thus, let us consider a function g in $\mathcal{C}^s(U,V)$, with $U \subset \mathbb{R}^n$ and $g(U) \subset W \subset \mathbb{R}^m$ and a function f in $\mathcal{C}^r(W,\mathbb{R})$ with $r,s > 0$. As before, $f\circ g$ is a $\mathcal{C}^t(U,\mathbb{R})$ function, where $t = \min(r,s)$. Similarly, we can get an expression for the derivatives of $f\circ g$, such that for $\ell = 1,\ldots,t$,

$$|f\circ g|_{\mathcal{C}^\ell} \le C_\ell \sum_{k=1}^{\ell}\sum_{j_1+\cdots+j_k=\ell} |f|_{\mathcal{C}^k}\,|g|_{\mathcal{C}^{j_1}}\cdots|g|_{\mathcal{C}^{j_k}}\,, \quad \text{(C.3)}$$

for $\ell = 0,\ldots,t$, where C_ℓ is a constant depending on ℓ. As before, we can consider the following less precise but more compact bound,

$$|f\circ g|_{\mathcal{C}^\ell} \le C_\ell\,|f|_{\mathcal{C}^\ell}\,|g|^\ell_{\mathcal{C}^\ell}\,, \quad \text{(C.4)}$$

for $\ell = 1,\ldots,t$, where C_ℓ is a constant depending on ℓ.

For some other results related to this, we refer the reader to [LO99].

In some cases, it will be more convenient to use another estimate for the $|\cdot|_{\mathcal{C}^\ell}$ norm instead of the one obtained in (C.4). In formula (C.3) we can separate the term corresponding to $k = 1$ in the following way

$$|f \circ g|_{\mathcal{C}^\ell} \leq C_\ell \left(|f|_{\mathcal{C}^1} |g|_{\mathcal{C}^\ell} + \sum_{k=2}^{\ell} \sum_{j_1+\cdots+j_k=\ell} |f|_{\mathcal{C}^k} |g|_{\mathcal{C}^{j_1}} \cdots |g|_{\mathcal{C}^{j_k}} \right),$$

for $\ell = 1, \ldots, t$ and we can bound it in the $|\cdot|_{\mathcal{C}^\ell}$ norm

$$|f \circ g|_{\mathcal{C}^\ell} \leq C_\ell (|f|_{\mathcal{C}^1} |g|_{\mathcal{C}^\ell} + |f|_{\mathcal{C}^\ell} |g|^\ell_{\mathcal{C}^{\ell-1}}), \tag{C.5}$$

for $\ell = 1, \ldots, t$, where C_ℓ is a constant depending on ℓ.

References

[Arn63] V. I. Arnol'd. Proof of a theorem of A. N. Kolmogorov on the invariance of quasi-periodic motions under small perturbations. *Russian Math. Surveys*, 18(5):9–36, 1963.

[Arn64] V.I. Arnold. Instability of dynamical systems with several degrees of freedom. *Sov. Math. Doklady*, 5:581–585, 1964.

[CH82] Shui Nee Chow and Jack K. Hale. *Methods of bifurcation theory*, volume 251 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Science]*. Springer-Verlag, New York, 1982.

[Chi79] B.V. Chirikov. A universal instability of many-dimensional oscillator systems. *Phys. Rep.*, 52(5):264–379, 1979.

[Con90] John B. Conway. *A course in functional analysis*, volume 96 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 1990.

[CY04a] Chong-Qing Cheng and Jun Yan. Existence of diffusion orbits in a priori unstable Hamiltonian systems. *J. Differential Geom.*, 67(3):457–517, 2004.

[CY04b] Chong-Quin Cheng and Jun Yan. Arnold diffusion in Hamiltonian systems: 1 a priori unstable case. Preprint 04-265, `mp_arc@math.utexas.edu`, 2004.

[DGLS08] Amadeu Delshams, Marian Gidea, Rafael de la Llave, and Tere M. Seara. Geometric approaches to the problem of instability in Hamiltonian systems: an informal presentation. Craig, Walter (ed.), Hamiltonian dynamical systems and applications. Proceedings of the NATO Advanced Study Institute on Hamiltonian dynamical systems and applications, Montreal, Canada, June, 18-29, 2007. Berlin: Springer. NATO Science for Peace and Security Series B: Physics and Biophysics, 285-336 (2008)., 2008.

[DLS00] Amadeu Delshams, Rafael de la Llave, and Tere M. Seara. A geometric approach to the existence of orbits with unbounded energy in generic periodic perturbations by a potential of generic geodesic flows of $\mathbf{t}^2$. *Comm. Math. Phys.*, 209(2):353–392, 2000.

[DLS06a] A. Delshams, R. de la Llave, and T. M. Seara. A geometric mechanism for diffusion in Hamiltonian systems overcoming the large gap problem: heuristics and rigorous verification on a model. *Mem. Amer. Math. Soc.*, 179(844):viii+141, 2006.

[DLS06b] Amadeu Delshams, Rafael de la Llave, and Tere M. Seara. Orbits of unbounded energy in quasi-periodic perturbations of geodesic flows. *Adv. Math.*, 202(1):64–188, 2006.

[DLS07] Amadeu Delshams, Rafael de la Llave, and Tere M. Seara. Instability of high dimensional hamiltonian systems: Multiple resonances do not impede diffusion. Manuscript, 2007.

[DLS08] Amadeu Delshams, Rafael de la Llave, and Tere M. Seara. Geometric properties of the scattering map of a normally hyperbolic invariant manifold. *Adv. Math.*, 217(3):1096–1153, 2008.

[Fen77] N. Fenichel. Asymptotic stability with rate conditions. II. *Indiana Univ. Math. J.*, 26(1):81–93, 1977.

[Fen79] N. Fenichel. Geometric singular perturbation theory for ordinary differential equations. *J. Differential Equations*, 31(1):53–98, 1979.

[Fen74] N. Fenichel. Asymptotic stability with rate conditions. *Indiana Univ. Math. J.*, 23:1109–1137, 1973/74.

[FLS07] E. Fontich, R. de la Llave, and Y. Sire. Construction of invariant whiskered tori by a parametrization method. part I: Maps and flows in finite dimensions. Preprint, 2007.

[GL06a] Marian Gidea and Rafael de la Llave. Arnold diffusion with optimal time in the large gap problem. *Preprint*, 2006.

[GL06b] Marian Gidea and Rafael de la Llave. Topological methods in the instability problem of Hamiltonian systems. *Discrete Contin. Dyn. Syst.*, 14(2):295–328, 2006.

[Her83] Michael-R. Herman. *Sur les courbes invariantes par les difféomorphismes de l'anneau. Vol. 1*, volume 103 of *Astérisque*. Société Mathématique de France, Paris, 1983. With an appendix by Albert Fathi, With an English summary.

[HPS77] M.W. Hirsch, C.C. Pugh, and M. Shub. *Invariant manifolds*, volume 583 of *Lecture Notes in Math.* Springer-Verlag, Berlin, 1977.

[LGJV05] R. de la Llave, A. González, À. Jorba, and J. Villanueva. KAM theory without action-angle variables. *Nonlinearity*, 18(2):855–895, 2005.

[LHS08] Rafael de la Llave, Gemma Huguet, and Yannick Sire. Fast numerical algorithms for the computation of invariant tori in Hamiltonian systems. Preprint, 2008.

[Lla00] R. de la Llave. Persistence of hyperbolic manifolds, 2000. Preprint.

[Lla01] Rafael de la Llave. A tutorial on KAM theory. In *Smooth ergodic theory and its applications (Seattle, WA, 1999)*, pages 175–292. Amer. Math. Soc., Providence, RI, 2001.

[LM88] P. Lochak and C. Meunier. *Multiphase averaging for classical systems*, volume 72 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 1988. With applications to adiabatic theorems, Translated from the French by H. S. Dumas.

[LO99] R. de la Llave and R. Obaya. Regularity of the composition operator in spaces of Hölder functions. *Discrete Contin. Dynam. Systems*, 5(1):157–184, 1999.

[Moe96] Richard Moeckel. Transition tori in the five-body problem. *J. Differential Equations*, 129(2):290–314, 1996.

[PT07] G. N. Piftankin and D. V. Treshchëv. Separatrix mapping in Hamiltonian systems. *Uspekhi Mat. Nauk*, 62(2(374)):3–108, 2007.

[Ten82] Jeffrey Tennyson. Resonance transport in near-integrable systems with many degrees of freedom. *Phys. D*, 5(1):123–135, 1982.

[Tre04] D. Treschev. Evolution of slow variables in a priori unstable hamiltonian systems. *Nonlinearity*, 17(5):1803–1841, 2004.

[Wig90] S. Wiggins. *Global bifurcations and chaos: Analytical methods*, volume 73 of *Appl. Math. Sci.* Springer, New York, 1990.