Planar undulator motion excited by a fixed traveling wave: Quasiperiodic averaging, normal forms, and the free electron laser pendulum

James A. Ellison,[1] Klaus Heinemann,[1] Mathias Vogt,[2] and Matthew Gooden[3]

[1]*Department of Mathematics and Statistics, The University of New Mexico, Albuquerque, New Mexico 87131, USA*
[2]*Deutsches Elektronen–Synchrotron, DESY, 22607 Hamburg, Germany*
[3]*Department of Physics, North Carolina State University, Raleigh, North Carolina 27695, USA*

(Received 14 March 2013; published 24 September 2013)

We present a mathematical analysis of planar motion of energetic electrons moving through a planar dipole undulator, excited by a fixed planar polarized plane wave Maxwell field in the x-ray free electron laser (FEL) regime. Our starting point is the 6D Lorentz system, which allows planar motions, and we examine this dynamical system as the wavelength λ of the traveling wave varies. By scalings and transformations the 6D system is reduced, without approximation, to a 2D system in a form for a rigorous asymptotic analysis using the method of averaging (MoA), a long-time perturbation theory. The two dependent variables are a scaled energy deviation and a generalization of the so-called ponderomotive phase. As λ varies the system passes through resonant and nonresonant (NonR) intervals and we develop NonR and near-to-resonant (NearR) MoA normal form approximations to the exact equations. The NearR normal forms contain a parameter which measures the distance from a resonance. For the planar motion, with the special initial condition that matches into the undulator design trajectory, and on resonance, the NearR normal form reduces to the well-known FEL pendulum system. We then state and prove NonR and NearR first-order averaging theorems which give explicit error bounds for the normal form approximations. We prove the theorems in great detail, giving the interested reader a tutorial on mathematically rigorous perturbation theory in a context where the proofs are easily understood. The proofs are novel in that they do not use a near-identity transformation and they use a *system* of differential inequalities. The NonR case is an example of quasiperiodic averaging where the small divisor problem enters in the simplest possible way. To our knowledge the planar problem has not been analyzed with the generality we aspire to here nor has the standard FEL pendulum system been derived with associated error bounds as we do here. We briefly discuss the low gain theory in light of our NearR normal form. Our mathematical treatment of the *noncollective* FEL beam dynamics problem in the framework of *dynamical systems theory* sets the stage for our mathematical investigation of the *collective* high gain regime.

DOI: 10.1103/PhysRevSTAB.16.090702 PACS numbers: 02.30.Hq, 02.30.Mv, 29.27.Bd, 41.60.Cr

I. INTRODUCTION

A. Basic ideas and parameters

We present a normal form analysis of the three-degree-of-freedom Lorentz force system of six ODEs (ordinary differential equations) governing the planar (x, $y = 0$, z) motion of relativistic electrons moving through a planar dipole undulator along the z axis perturbed by a horizontally polarized plane wave traveling in the z direction. We are interested in the parameter range for an x-ray FEL.

Our normal form analysis is based on the method of averaging (MoA) at first order. The method has four steps. The first step is to put ODEs into a standard form. The second step is to identify the normal form approximations. The third step is the derivation of error bounds relating the exact and normal form solutions. The final step is the transformation back to the original variables. In the first step new variables are typically introduced using scalings and transformations. In this process we discover that the exact problem can be formulated, without approximation, in terms of two ODEs for the normalized energy deviation and a generalized ponderomotive phase. Important in this process is the identification of an appropriate small dimensionless parameter, often denoted by ε, so that the system can be written as $\dot{u} = \varepsilon f(u, t) + O(\varepsilon^2)$. In the present context this is the most complicated step. The normal form approximation is obtained by dropping the $O(\varepsilon^2)$ term and replacing f by its t average. The third step is often the most difficult, however here the system in standard form is fairly simple and we use this opportunity to give very detailed proofs of two averaging theorems, partly as a tutorial on the methods of proof, rather than applying general theorems from the literature. The latter allows us to obtain quite explicit error bounds which are likely near optimal.

An electron, as a member of an electron bunch, will enter the undulator with a given angle in the $y = 0$ plane and a given Lorentz factor. Here the normalized angle will be given by ΔP_{x0} and the Lorentz factor will be written $\gamma = \gamma_c(1 + \eta)$, where γ_c is a characteristic value of γ for the electron bunch, e.g., the mean, and η is the so-called normalized energy deviation. We will replace η by χ via the relation $\eta = \varepsilon\chi$, where *a posteriori* ε will be a measure of the spread of η values which lead to an FEL pendulum-type behavior. We let B_u, k_u denote the undulator field strength and wave number and let E_r, νk_r denote the Maxwell field strength and wave number of the fixed traveling wave radiation field. Thus, our basic parameters are eight, namely ΔP_{x0}, γ_c, ε, B_u, k_u, E_r, k_r, ν, and all will be assumed positive except ΔP_{x0}. We will study the electron response to the radiation field as ν varies.

For an x-ray FEL, ε is small, γ_c is large, and the dimensionless undulator parameter,

$$K := \frac{eB_u}{mck_u} = 0.934\frac{\lambda_u}{1\ \text{cm}}\frac{B_u}{1\ \text{T}}, \tag{1.1}$$

is $O(1)$. Also $k_r = O(k_u\gamma_c^2)$ and we define the $O(1)$ constant K_r by

$$K_r := \frac{k_r}{k_u\gamma_c^2}. \tag{1.2}$$

In Sec. III B it will be clear that it is convenient to fix K_r by setting

$$K_r = 2\left[1 + \frac{1}{2}K^2 + K^2(\Delta P_{x0})^2\right]^{-1}. \tag{1.3}$$

For $\Delta P_{x0} = 0$ we obtain $k_r = 4k_u\gamma_c^2/(2 + K^2)$ which is the usual so-called resonant wave number (see, e.g., [1]). The dependence of K_r on ΔP_{x0} will be a consequence of our analysis. Typical parameters for the Linac Coherent Light Source (LCLS) are $\lambda_u = 3$ cm, $mc^2\gamma_c = 15$ GeV, and $B_u = 1.32$ T so that $K = 3.70$ (see [2]) and for FLASH (Free-Electron Laser in Hamburg) are $\lambda_u = 2.73$ cm, $mc^2\gamma_c = 0.7$ GeV and $B_u = 0.48$ T so that $K = 1.23$ (see [3]).

Mathematically then, we are interested in an asymptotic analysis of the electron motion for ε small and γ_c large as ν varies. In particular we are interested in the (ε, γ_c) regime that gives rise to the pendulum-type behavior important for the functioning of an x-ray FEL. We find that in order to obtain this behavior, in the MoA at first order, there must be a relation between ε and γ_c. Introducing the normalized field strength,

$$\mathcal{E} := \frac{E_r}{cB_u}, \tag{1.4}$$

we show a pendulum-type behavior emerges when $\varepsilon = O(\sqrt{\mathcal{E}}/\gamma_c)$ for ε small. Without loss of generality we will take the order constant to be 1, and choose

$$\varepsilon = \sqrt{\mathcal{E}}\frac{1}{\gamma_c}. \tag{1.5}$$

We also show that, for ε small, the system associated with (1.5) has a resonant structure, such that, as ν varies, the system goes through a sequence of nonresonant (NonR) and near-to-resonant (NearR) intervals. The associated NearR approximating normal forms while being nonautonomous have an underlying pendulum structure and reduce to the standard FEL pendulum system for $\Delta P_{x0} = 0$ and ν an odd integer. This behavior is not present for $\varepsilon \ll 1/\gamma_c$ or $\varepsilon \gg 1/\gamma_c$ and so we refer to (1.5) as a *distinguished case*. This turns out to be a very simple example of the concept of a "distinguished limit" in the singular perturbation literature. This can be seen in action in the context of our equations (2.49) and (2.50).

When it comes to our main results in Sec. III we have the following situation. We have the eight basic parameters ΔP_{x0}, γ_c, ε, B_u, k_u, E_r, k_r, ν and the two constraints (1.3) and (1.5), leaving us with six basic parameters out of which we construct the five nondimensional parameters K, ΔP_{x0}, $\mathcal{E}$, ε, ν (see also [4]). The NearR normal forms can be understood in terms of the simple pendulum system and reduce to the usual FEL pendulum equations for $\Delta P_{x0} = 0$ and ν an odd integer (see Secs. III D 2 and III D 3). The NearR normal form allows us to study the effect of ν being slightly off resonance. This completes the first two steps in the MoA. In the third step we state two theorems which give error bounds, relating the exact and normal form solutions, which go to zero as $\varepsilon \to 0+$. The theorems are then proved. Our goal is to present a mathematically rigorous analysis that is self-contained. However, the reader can understand the results of the paper without understanding the proofs of the theorems. With this in mind the proofs are isolated in a separate section.

B. Comment on normalized field strength

For the results of this paper the normalized field strength $\mathcal{E}$ cannot be too big (or ε will not be small) and it cannot be too small or another distinguished case will come into play. Of course for a seeded FEL, $\mathcal{E}$ will be set by the seeding field. In Appendix G we present two very crude bounds that have some relevance to the beginning stages of a high gain FEL. Here we simply note that for $\mathcal{E} = 1000$, ε is approximately 0.001 for the 15 GeV (e.g., LCLS) and 0.025 for the 700 MeV (e.g., FLASH). In an early approach to this problem we built a normal form analysis assuming $\mathcal{E}$ small, so that the radiation field was a small perturbation of the undulator motion. We thus considered $\mathcal{E}$ as a small parameter in addition to $1/\gamma_c$. This led to another distinguished case, which also had resonant and NonR intervals but with a different underlying pendulum structure. Later we realized that $\mathcal{E}$ is not necessarily small for cases of interest and we were led to the current case of (1.5). In fact, since $\mathcal{E}$

does not have to be small our results may have some relevance to the high gain saturation regime [5].

C. General comment on method of averaging

For ODEs, the MoA is the most robust of the long-time perturbation theories which include, e.g., Lindstedt series [6], multiple scales [6], renormalization group methods [7], and Hamiltonian perturbation theory [8]. For example, Hamiltonian perturbation theory has the advantage that one is transforming a scalar function, however the MoA is more robust in that transformations and scalings are not restricted to canonical transformations. Central to the MoA, and in contrast to those just mentioned, is the derivation of error bounds. We emphasize these are true bounds and not just estimates. The MoA is a mature subject and there are several good books; see [6,9,10] for example as well as the Scholarpedia articles [11,12]. We refer to the MoA approximation as a normal form. Generally, a normal form of a mathematical object is a simplified form of the object obtained with the aid of, for example, scalings and transformations such that the essential features of the object are preserved. Here we not only preserve the essential features of the exact ODEs but bound the errors in the approximation with a bound proportional to the small parameter ε. See [11] for the use of normal form in a similar context.

D. Paper outline

In Sec. II we start with the three-degree-of-freedom Lorentz equations with the horizontally polarized plane wave of (2.9) and then introduce z as the independent variable. The system has planar solutions where $0 = y = p_y$ and using a conservation law we arrive at a system of two ODEs (2.30) and (2.31) for the energy and a precursor to a generalization of the so-called ponderomotive phase. By scalings and transformations we discover the distinguished case of (1.5) which then leads to a standard form for the method of averaging in (2.52) and (2.53). The two dependent variables are now a scaled energy deviation and a generalization of the so-called ponderomotive phase.

In Sec. III we present our main results. We begin by introducing the monochromatic plane wave, the case of main physical interest. The system is carefully defined in Sec. III A. In Sec. III B we discuss the topic of resonance in the MoA context. We emphasize that as ν varies the system passes through resonant and NonR intervals. In particular we introduce the resonant, NonR, Δ-nonresonant (Δ-NonR), and NearR cases. The NonR case, its first-order averaging normal form, and associated solutions are presented in Sec. III C as well as an appropriate domain for the associated vector field. The NearR case is discussed in Sec. III D. The system is carefully defined in Sec. III D 1 and an appropriate domain for the associated vector field is found. The first-order averaging normal form is derived in Sec. III D 2 and the normal form system is transformed into the simple pendulum system in Sec. III D 3. The structure of the NearR normal form solutions are then discussed in detail. For the special initial condition, $\Delta P_{x0} = 0$, which matches into the design trajectory of the undulator, we recover the result of standard approaches which focus on the energy transfer equations alone and do not consider the phase space variables. This completes the first two steps in the MoA.

Some readers may only care about the exact equations in the form for averaging, the associated normal form approximations and the rough statement that the approximation errors are $O(\varepsilon)$ on $O(1/\varepsilon)$ "time" intervals. They will find these in (3.18), (3.19), (3.46), (3.47), (3.39), (3.40), (3.61), and (3.62) and in the beginning of Sec. III E. The statements of our first-order averaging theorems, which give an order ε bound on the error for long times, i.e., intervals of $O(1/\varepsilon)$, are presented in Sec. III E and applied to the six variables in Sec. III F. This completes the third and fourth steps of the MoA. We emphasize that the averaging theorems can be understood independently of the proofs in Sec. IV. Finally in Sec. III G we use our results in a low gain calculation and compare the result with [13].

While the paper is aimed at the FEL community we believe that newcomers to the field and mathematically inclined readers will find Secs. II and III a good introduction to the noncollective case of an FEL.

The proofs of the two averaging theorems, stated in Secs. III E 1 and III E 2, are presented in Sec. IV and this section can be skipped if the reader only wants to know the results as given in Sec. III. However, Sec. IV has a pedagogical aspect, giving the reader, who may not be familiar with modern long-time perturbation theory, an introduction in a context where the proofs are easily understood. The proofs are given in detail so the reader needs no prior knowledge of the MoA. They are based on an idea of Besjes (see [14–16]) which leads to proofs without using a so-called near-identity transformation, as in usual treatments of, e.g., [6,9,10]. The Δ-NonR case is an example of quasiperiodic averaging with a rigorous treatment of a small divisor problem in what is surely the simplest setting. Here we keep ν a distance Δ away from a resonance, where Δ turns out to be the small divisor. The NearR case is an example of periodic averaging. A novelty of our approach in Sec. IV is that we use a *system* of differential inequalities, rather than the usual Gronwall inequality, to obtain better error bounds. Furthermore we obtain better results as our theorems are tuned to the problem at hand. In addition, to our knowledge, the treatment of the undulator problem in the mathematically rigorous and self-contained way that we do here has not been done before. Our mathematical analysis is straightforward, using only undergraduate mathematics as commonly taught in advanced calculus courses, however the proofs are somewhat intricate in spots. The reader who studies Sec. IV will be rewarded with an understanding of the computation of error bounds in the quasiperiodic method of averaging in what is likely the most elementary context. The general quasiperiodic case [14] is more difficult and for contrast the proofs of the KAM and Nekhoroshev theorems are much more difficult [8]. Finally, for us, the work on the noncollective case sets the stage for our more serious goal of obtaining a deep mathematical understanding of the collective high gain FEL theory.

Following the summary of Sec. V the Appendices contain calculations needed in the main text. Appendix A provides properties of the Bessel expansion of the function jj which is introduced in Sec. III B. In Appendices B and C we study the next-to-leading order terms g_1, g_2 used in Theorem 1 and in Appendices D and E we study the next-to-leading order terms g_1^R, g_2^R used in Theorem 2. Appendix F provides some formulas used in Sec. III G. In Appendix G we discuss $\mathcal{E} = E_r/cB_u$ in the high gain regime and obtain a crude upper bound estimate of it. Finally, in Appendix H we derive some properties of the system of differential inequalities that are used in the proof of both averaging theorems.

E. Putting our work in context

Standard derivations of the FEL pendulum equations can be found in, e.g., [13,17–19]. They differ from our approach in that they start from the ODE for the normalized energy deviation, η, and use physical reasoning to introduce approximations leading to the FEL pendulum normal form for $\Delta P_{x0} = 0$. They do not derive the pendulum equation using perturbation theory as we do nor do they obtain error bounds on their approximations.

In contrast, we start with the full 6D noncollective system of Lorentz equations of motion for our undulator with a planar polarized traveling wave in the z direction and show that it contains planar $y = 0$ motion. Writing γ as $\gamma = \gamma_c(1 + \varepsilon\chi)$, where γ_c is a characteristic value of γ for the electron bunch and $\eta = \varepsilon\chi$ is the so-called normalized energy deviation, we find a distinguished relation in (ε, γ_c) given by $\varepsilon = \sqrt{\mathcal{E}}/\gamma_c$ that leads to a pair of exact equations for a scaled energy deviation and a generalized ponderomotive phase (which depends on ΔP_{x0}) in a form for the application of the MoA, with small parameter ε. These equations are exact in the sense that there are no approximations in deriving them from the full 6D Lorentz system. We note that ε is small even if the traveling wave is *not* a small perturbation to the undulator motion. When we specialize to the single frequency, νk_r, monochromatic case we see there is a resonant structure depending on ν. Using the MoA formalism we find resonant and NonR normal forms as ν varies. We then ask the question, "When are these good approximations to the exact problem?" We show that the NonR normal form gives a good approximation sufficiently far from resonance, on $O(1/\varepsilon)$ time intervals. We then consider neighborhoods of the resonances and construct NearR normal form approximations which reduce to the resonant normal form for resonant ν. We show that they give an $O(\varepsilon)$ approximation to the exact problem, on $O(1/\varepsilon)$ time intervals, in $O(\varepsilon)$ neighborhoods of the resonances. On resonance and for $\Delta P_{x0} = 0$ our generalized ponderomotive phase reduces to the well-known ponderomotive phase and in addition we obtain the usual FEL pendulum system. Thus, we have a new view of the ponderomotive phase and a new derivation of the FEL pendulum equation.

Our approach gives a clear picture of the system response, for small ε as ν varies through resonant and NonR intervals with the associated NonR and NearR normal form approximations to the generalized ponderomotive phase and scaled energy deviation. To our knowledge this is new. Our explicit error bounds, covered by our averaging theorems, are nearly optimal and can be used to examine parameter ranges for validity of the low gain regime. The fact that $\mathcal{E}$ does not have to be small led to the suggestion that our results may be relevant in the high gain saturation regime. Finally, we have discovered that a planar polarized traveling wave with a continuous distribution of wave numbers near resonance (e.g., a narrow Gaussian centered on the resonance) may wash out the resonant effect and thus the FEL pendulum behavior, at least in first-order averaging. We discuss this briefly in Sec. V.

Our definition of resonance is intimately linked to the derivation of our averaging normal forms, whereas in the standard derivations resonance is introduced in the context of maximizing energy exchange. We emphasize that we obtain more than the pendulum normal form; we also obtain the more general NearR normal forms as well as the NonR normal forms.

We do not intend to minimize the importance of the standard derivations; the physical derivations are certainly important and as is often the case show great physical insight. Here we want to show what can be done in a mathematically rigorous way in the context of dynamical systems theory, but in that we have been guided by and are indebted to previous works.

II. GENERAL PLANAR UNDULATOR MODEL

In this section we state the basic problem and put the equations of motion in a standard form for the MoA.

A. Lorentz force equations

Using SI units, the Lorentz equations for motion of a relativistic electron in an electromagnetic field, $(\mathbf{E}, \mathbf{B})$, are

$$\dot{\mathbf{r}} = \mathbf{v}(\mathbf{p}), \tag{2.1}$$

$$\dot{\mathbf{p}} = -e[\mathbf{E} + \mathbf{v}(\mathbf{p}) \times \mathbf{B}], \tag{2.2}$$

with $\dot{} = d/dt$ and where

$$\mathbf{v}(\mathbf{p}) = \frac{\mathbf{p}}{m\gamma}, \tag{2.3}$$

is the velocity, γ is the Lorentz factor defined by

$$\gamma^2 = 1 + \mathbf{p}\cdot\mathbf{p}/m^2c^2, \tag{2.4}$$

and m and $-e$ are the electron mass and charge, respectively. We denote the undulator magnetic field by $\mathbf{B}_u$ and the radiation field by (($\mathbf{E}_r$, $\mathbf{B}_r$) whence

$$\mathbf{E} = \mathbf{E}_r, \qquad \mathbf{B} = \mathbf{B}_r + \mathbf{B}_u. \tag{2.5}$$

We introduce Cartesian coordinates by

$$\mathbf{r} = x\mathbf{e}_x + y\mathbf{e}_y + z\mathbf{e}_z, \tag{2.6}$$

$$\mathbf{p} = p_x\mathbf{e}_x + p_y\mathbf{e}_y + p_z\mathbf{e}_z, \tag{2.7}$$

where $\mathbf{e}_x$, $\mathbf{e}_y$, $\mathbf{e}_z$ are the standard unit vectors. A simple planar undulator model magnetic field which satisfies the Maxwell equations, $\nabla\cdot\mathbf{B}_u = 0$ and $\nabla\times\mathbf{B}_u = 0$, as in [17], is

$$\mathbf{B}_u = -B_u[\cosh(k_u y)\sin(k_u z)\mathbf{e}_y + \sinh(k_u y)\cos(k_u z)\mathbf{e}_z], \tag{2.8}$$

where $B_u > 0$. Since $\nabla\times\mathbf{B}_u = 0$ there is a scalar potential ϕ such that $\mathbf{B}_u = \nabla\phi$. To satisfy $\nabla\cdot\mathbf{B}_u = 0$, ϕ must satisfy Laplace's equation. The field (2.8) is easily constructed by separation of variables and requiring periodicity in z with period λ_u and then taking the first eigenmode (see, e.g., page 145 of [20]). The scalar field is $\phi = -(B_u/k_u)\sinh(k_u y)\sin(k_u z)$. The traveling wave radiation field we choose is also a Maxwell field and is given by

$$\mathbf{E}_r = E_r h(\breve{\alpha})\mathbf{e}_x, \quad \mathbf{B}_r = \frac{1}{c}(\mathbf{e}_z\times\mathbf{E}_r) = \frac{E_r}{c}h(\breve{\alpha})\mathbf{e}_y, \tag{2.9}$$

where E_r is a positive constant, h is a real valued function on $\mathbb{R}$,

$$\breve{\alpha}(z,t) = k_r(z - ct), \tag{2.10}$$

and k_r is the positive parameter mentioned in the Introduction. In the present Sec. II we will put the equations of motions in a standard form for the MoA staying with a general h. It is easy to carry through this part of the analysis without restricting h and we do want to make a comment on this general case.

However, our primary emphasis is on the standard monochromatic example and from Sec. III onward we will take the monochromatic case, i.e.,

$$H(\breve{\alpha}) = (1/\nu)\sin(\nu\breve{\alpha}), \qquad h(\breve{\alpha}) = H'(\breve{\alpha}) = \cos(\nu\breve{\alpha}), \tag{2.11}$$

and $\nu \geq 1/2$ thus $h[\breve{\alpha}(z,t)] = \cos[\nu k_r(z - ct)]$. Note that the prime $'$ always indicates a derivative of a function with respect to its only argument. In this monochromatic case k_r will be fixed by (1.2) and (1.3) and the ν will allow for a variable wave number for the traveling wave. With the choice of K_r mentioned in the Introduction, it will be seen that $\nu = 1$ gives the primary resonance with the concomitant pendulum normal form.

The Lorentz system can now be written in Cartesian coordinates as

$$\dot{x} = \frac{p_x}{m\gamma}, \qquad \dot{y} = \frac{p_y}{m\gamma}, \qquad \dot{z} = \frac{p_z}{m\gamma}, \tag{2.12}$$

$$\dot{p}_x = -e\bigg[\frac{p_z}{m\gamma}B_u\cosh(k_u y)\sin(k_u z) - \frac{p_y}{m\gamma}B_u\sinh(k_u y) \times\cos(k_u z) + E_r\Big(1 - \frac{p_z}{m\gamma c}\Big)h[\breve{\alpha}(z,t)]\bigg], \tag{2.13}$$

$$\dot{p}_y = -e\frac{p_x}{m\gamma}B_u\sinh(k_u y)\cos(k_u z), \tag{2.14}$$

$$\dot{p}_z = -e\bigg[-\frac{p_x}{m\gamma}B_u\cosh(k_u y)\sin(k_u z) + E_r\frac{p_x}{m\gamma c}h[\breve{\alpha}(z,t)]\bigg]. \tag{2.15}$$

It is easy to check that (2.12)–(2.15) is a Hamiltonian system with Hamiltonian $\mathcal{H}$:

$$\mathcal{H} = c\sqrt{[\mathbf{P}_c + e\mathbf{A}(\mathbf{r},t)]^2 + m^2c^2} = mc^2\gamma, \tag{2.16}$$

where the canonical momentum vector $\mathbf{P}_c$ is related to $\mathbf{p}$ by $\mathbf{p} = \mathbf{P}_c + e\mathbf{A}$ and the vector potential $\mathbf{A}$ is given by

$$\mathbf{A}(y,z,t) = \bigg[\frac{B_u}{k_u}\cosh(k_u y)\cos(k_u z) + \frac{E_r}{k_r c}H[\breve{\alpha}(z,t)]\bigg]\mathbf{e}_x. \tag{2.17}$$

Since $\mathbf{A}$ is independent of x the x component, $P_{c,x}$, of the canonical momentum vector $\mathbf{P}_c$ is conserved, i.e.,

$$p_x - eA_x(y,z,t), \tag{2.18}$$

is constant along solutions of (2.12)–(2.15) as is easily confirmed directly. We will not make explicit use of the Hamiltonian structure in the following. The MoA does not require a Hamiltonian structure and this frees us from having to deal only with canonical transformations as we proceed to put (2.12)–(2.15) in an averaging standard form.

B. Motion in the $y = 0$ plane with z as the independent variable

It is common to take the distance z along the undulator as the independent variable, rather than the time t. With the usual abuse of notation, and from now on, we write $x(z)$, $y(z)$, $p_x(z)$, $p_y(z)$, $p_z(z)$ instead of $x[t(z)]$, $y[t(z)]$, $p_x[t(z)]$, $p_y[t(z)]$, $p_z[t(z)]$ whence the ODEs (2.12)–(2.15) become

$$\frac{dx}{dz} = \frac{p_x}{p_z}, \qquad \frac{dy}{dz} = \frac{p_y}{p_z}, \qquad \frac{dt}{dz} = \frac{m\gamma}{p_z}, \tag{2.19}$$

$$\frac{dp_x}{dz} = -\frac{e}{c}\left[cB_u\cosh(k_u y)\sin(k_u z) - \frac{p_y}{p_z}cB_u\sinh(k_u y)\right. \\ \left.\times\cos(k_u z) + E_r\left(\frac{m\gamma c}{p_z} - 1\right)h[\check{\alpha}(z,t)]\right], \tag{2.20}$$

$$\frac{dp_y}{dz} = -\frac{e}{c}\frac{p_x}{p_z}cB_u\sinh(k_u y)\cos(k_u z), \tag{2.21}$$

$$\frac{dp_z}{dz} = -\frac{e}{c}\left[-\frac{p_x}{p_z}cB_u\cosh(k_u y)\sin(k_u z) + E_r\frac{p_x}{p_z}h[\check{\alpha}(z,t)]\right]. \tag{2.22}$$

The initial conditions at $z = 0$ will be denoted by a subscript 0, e.g., $t(0) = t_0$. Clearly t_0 is the arrival time of an electron at the entrance, $z = 0$, of the undulator.

Here and in the rest of the paper we consider the initial value problem (IVP) with $y_0 = p_{y0} = 0$. It follows, with no approximation, that $y(z) = p_y(z) = 0$ for all z and the six ODEs (2.19)–(2.22) reduce to four. The right-hand sides (rhs's) of (2.19)–(2.22) are independent of x and so we do not need to consider the x equation until Sec. III F. It is standard, and also quite convenient, to replace p_z by the energy variable γ. With $\gamma(z)$ defined in terms of $p_x(z)$ and $p_z(z)$ by (2.4) and using (2.20) and (2.22), we obtain $\gamma' = (p_x p_x' + p_z p_z')/m^2c^2\gamma = -(eE_r/mc^2)(p_x/p_z)h[\check{\alpha}(z,t)]$. Finally, we take $\check{\alpha}$ as a dependent variable in place of t and we define

$$\alpha(z) := \check{\alpha}[z, t(z)] = k_r[z - ct(z)]. \tag{2.23}$$

Later it will be seen that α is a precursor to a generalization of the so-called ponderomotive phase which emerges naturally as we put the ODEs in a standard form for averaging.

With the above four changes the ODEs for t, p_x, p_z in (2.19), (2.20), and (2.22) become

$$\frac{d\alpha}{dz} = k_r\left(1 - \frac{m\gamma c}{p_z}\right), \tag{2.24}$$

$$\frac{dp_x}{dz} = -\frac{e}{c}\left[cB_u\sin(k_u z) + E_r\left(\frac{m\gamma c}{p_z} - 1\right)h(\alpha)\right], \tag{2.25}$$

$$\frac{d\gamma}{dz} = -\frac{eE_r}{mc^2}\frac{p_x}{p_z}h(\alpha), \tag{2.26}$$

where the initial conditions are $\alpha(0) = \alpha_0 := -k_r c t_0$, $p_x(0) =: p_{x0}$, $\gamma(0) =: \gamma_0$. Here p_z must be replaced by

$$p_z = \sqrt{m^2c^2(\gamma^2 - 1) - p_x^2}, \tag{2.27}$$

and it is easy to see that (2.24)–(2.26) are then self-contained. From now on, we restrict p_z to be positive:

$$p_z > 0. \tag{2.28}$$

Thus, the argument of the square root in (2.27) is positive and this entails $\gamma > 1$, as it should be. Note that, by (2.24), α is a strictly decreasing function whence, as one expects, $z < c[t(z) - t_0]$. It is also easy to check that

$$\frac{p_x}{mcK} - \cos(k_u z) - \frac{E_r}{cB_u}\frac{k_u}{k_r}H(\alpha) \tag{2.29}$$

is conserved along solutions of (2.24)–(2.26). This conservation law is identical to (2.18) with $y = 0$. Recall that K was defined by (1.1).

To complete the solution of (2.19)–(2.22) it suffices to note that $t(z)$ is determined from (2.23) in terms of $\alpha(z)$ and $x(z)$ is determined from (2.19) by integration.

We can now state a first formulation of the *basic 2D system* which we study in the rest of the paper. We have replaced the Lorentz system defined at the beginning of Sec. II A by the IVP for (2.19)–(2.22) with $y_0 = p_{y0} = 0$, which entails $y = p_y = 0$. The solutions of the IVP are given in terms of solutions of (2.24) and (2.26). We now write our basic system as the complete IVP for (2.24) and (2.26), namely,

$$\frac{d\alpha}{dz} = k_r\left(1 - \frac{m\gamma c}{p_z}\right), \qquad \alpha(0) = \alpha_0, \tag{2.30}$$

$$\frac{d\gamma}{dz} = -\frac{eE_r}{mc^2}\frac{p_x}{p_z}h(\alpha), \qquad \gamma(0) = \gamma_0, \tag{2.31}$$

with p_x and p_z replaced by

$$p_x = p_{x0} + mcK\left(\cos(k_u z) - 1 + \frac{E_r}{cB_u}\frac{k_u}{k_r}[H(\alpha) - H(\alpha_0)]\right), \tag{2.32}$$

$$p_z = \sqrt{m^2c^2(\gamma^2 - 1) - p_x^2}. \tag{2.33}$$

C. Standard form for method of averaging

We begin by introducing the normalized energy deviation η and its $O(1)$ counterpart χ via

$$\gamma = \gamma_c(1 + \eta) = \gamma_c(1 + \varepsilon\chi), \tag{2.34}$$

as mentioned in the Introduction. Here γ_c is a characteristic value of γ, e.g., its mean over the bunch, and ε is a characteristic spread of η so that χ becomes the new $O(1)$ dependent variable replacing γ in (2.30) and (2.31). We are interested in an asymptotic analysis for γ_c large and η small as in an x-ray FEL. Here we determine a relation between ε and γ_c which leads to a standard form for the MoA and which will contain the FEL pendulum system at first order in the monochromatic case of (2.11).

A natural scaling for z is

$$z = \zeta/k_u, \tag{2.35}$$

so that the undulator period is 2π in ζ. Introducing

$$\theta_{\rm aux}(\zeta) := \alpha(\zeta/k_u), \tag{2.36}$$

and considering $\chi = \chi(\zeta)$ in (2.34), we obtain

$$\theta'_{\rm aux} = K_r\gamma_c^2\Big(1 - \frac{1+\varepsilon\chi}{P_z}\Big), \tag{2.37}$$

$$\chi' = -K^2\frac{\mathcal{E}}{\varepsilon\gamma_c^2}\frac{P_x}{P_z}h(\theta_{\rm aux}), \tag{2.38}$$

where $' = d/d\zeta$, K_r and $\mathcal{E}$ are given in (1.2) and (1.4), and where

$$P_x := \cos\zeta + \Delta P_{x0} + \frac{\mathcal{E}}{K_r\gamma_c^2}[H(\theta_{\rm aux}) - H(\theta_0)], \tag{2.39}$$

$$\Delta P_{x0} := \frac{p_{x0}}{mcK} - 1, \tag{2.40}$$

$$P_z := \Big[(1+\varepsilon\chi)^2 - \frac{1}{\gamma_c^2}(1 + K^2P_x^2)\Big]^{1/2}. \tag{2.41}$$

Since we require $p_z > 0$ and since $P_z^2 = (p_z/mc\gamma_c)^2$, as is easily checked, the argument of the square root in (2.41) must be positive and thus equals $p_z/mc\gamma_c$. This along with the fact that γ is positive leads to the maximal domain in the extended phase space $(\theta_{\rm aux}, \chi, \zeta)$ for (2.37) and (2.38), as defined by $\gamma_c(1+\varepsilon\chi) > [1 + K^2P_x^2]^{1/2}$. We note that most derivations of the FEL pendulum take $\Delta P_{x0} = 0$, see [13,17–19], which is only necessary for perfect matching of the incoming bunch.

It is easy to show that

$$\frac{P_x}{P_z} = \cos\zeta + \Delta P_{x0} + O(\varepsilon) + O\Big(\frac{1}{\gamma_c^2}\Big), \tag{2.42}$$

$$\frac{1+\varepsilon\chi}{P_z} = 1 + \frac{q(\zeta)}{2\gamma_c^2}(1 - 2\varepsilon\chi) + O\Big(\frac{1}{\gamma_c^4}\Big) + O\Big(\frac{\varepsilon^2}{\gamma_c^2}\Big), \tag{2.43}$$

where

$$q(\zeta) := 1 + K^2(\cos\zeta + \Delta P_{x0})^2. \tag{2.44}$$

Thus (2.37) and (2.38) become

$$\theta'_{\rm aux} = -\frac{K_rq(\zeta)}{2} + \varepsilon K_rq(\zeta)\chi + O(1/\gamma_c^2) + O(\varepsilon^2), \tag{2.45}$$

$$\chi' = -K^2\frac{\mathcal{E}}{\varepsilon\gamma_c^2}(\cos\zeta + \Delta P_{x0})h(\theta_{\rm aux}) + O(1/\gamma_c^2) + O(1/\varepsilon\gamma_c^4). \tag{2.46}$$

To transform (2.45) and (2.46) into a standard form for the MoA we need to introduce dependent variables that are slowly varying. We anticipate that χ will be slowly varying, i.e., $\mathcal{E}/\varepsilon\gamma_c^2$ will be small. To remove the $O(1)$ term in (2.45) we define

$$\theta := \theta_{\rm aux} + Q(\zeta), \tag{2.47}$$

where

$$Q'(\zeta) = \frac{K_r}{2}q(\zeta), \qquad Q(0) = 0 \tag{2.48}$$

(see [21] for comment). Thus, the system (2.45) and (2.46) becomes

$$\theta' = \varepsilon K_rq(\zeta)\chi + O(1/\gamma_c^2) + O(\varepsilon^2), \tag{2.49}$$

$$\chi' = -K^2\frac{\mathcal{E}}{\varepsilon\gamma_c^2}(\cos\zeta + \Delta P_{x0})h[\theta - Q(\zeta)] + O(1/\gamma_c^2) + O(1/\varepsilon\gamma_c^4). \tag{2.50}$$

To obtain a system where θ and χ interact with each other in first-order averaging we must balance the $O(\varepsilon)$ term in (2.49) with the $O(\mathcal{E}/\varepsilon\gamma_c^2)$ in (2.50). In this spirit we relate ε and γ_c by choosing

$$\varepsilon = \frac{\mathcal{E}}{\varepsilon\gamma_c^2}, \tag{2.51}$$

and so we obtain (1.5). It is this balance that will lead to the FEL pendulum equations in Sec. III and this is the distinguished case mentioned in the Introduction.

In summary, the *basic 2D system* at the end of Sec. II B has been transformed to

$$\theta' = \varepsilon K_rq(\zeta)\chi + \varepsilon^2g_1(\theta, \chi, \zeta, \varepsilon, \nu), \tag{2.52}$$

$$\chi' = -\varepsilon K^2(\cos\zeta + \Delta P_{x0})h[\theta - Q(\zeta)] + \varepsilon^2g_2(\theta, \chi, \zeta, \varepsilon, \nu), \tag{2.53}$$

which is now in standard form and is the *basic 2D system for our averaging study*. However, the $O(\varepsilon^2)$ terms are not explicit. To make them explicit we first rewrite (2.37) and (2.38) in terms of ε, K, and $\mathcal{E}$ and use (2.47) to obtain

$$\theta' = \frac{K_r\mathcal{E}}{\varepsilon^2}\Big(1 - \frac{1+\varepsilon\chi}{P_z}\Big) + \frac{K_rq(\zeta)}{2}, \tag{2.54}$$

$$\chi' = -\varepsilon K^2\frac{P_x}{P_z}h[\theta - Q(\zeta)], \tag{2.55}$$

$$P_x = \cos\zeta + \Delta P_{x0} + \frac{\varepsilon^2}{K_r}\{H[\theta - Q(\zeta)] - H(\theta_0)\}, \tag{2.56}$$

$$P_z = \Big[(1+\varepsilon\chi)^2 - \frac{\varepsilon^2}{\mathcal{E}}(1 + K^2P_x^2)\Big]^{1/2}, \tag{2.57}$$

replacing (2.30)–(2.33). Here K, K_r, $\mathcal{E}$, ε are given in (1.1), (1.2), (1.4), and (1.5), ΔP_{x0} is given in (2.40), q, Q are given in (2.44) and (2.48), and the independent variables θ, χ are defined in (2.34), (2.36), and (2.47).

The $O(\varepsilon^2)$ terms in (2.52) and (2.53) can now be determined by comparison with (2.54) and (2.55). However, we have not proven that these terms are actually bounded by an ε-independent constant times ε^2. We will do this in the monochromatic case of Sec. III where we will show that the two $O(\varepsilon^2)$ terms are truly bounded by $\mathcal{C}\varepsilon^2$ on an appropriate domain for appropriate ε independent constants $\mathcal{C}$. This is to be expected because of the analysis leading to (2.52) and (2.53).

The θ defined by (2.47) is essentially the so-called ponderomotive phase. Here it arises naturally in the process of finding the distinguished relation between ε and γ_c and transforming to slowly varying coordinates. In standard treatments it is introduced heuristically to maximize energy transfer.

III. MONOCHROMATIC PLANE WAVES AND AVERAGING THEOREMS

We have the planar undulator in a standard form for the MoA in (2.52) and (2.53) where the $O(\varepsilon^2)$ terms can be determined from (2.54) and (2.55). We now specialize to a monochromatic traveling wave, write the system in Fourier form, discuss resonance as a normal form phenomenon, develop the NonR and NearR normal forms, and state one proposition and two theorems giving precise bounds on the normal form approximations. Thus, from now on the radiation field in (2.9) is monochromatic, i.e., h, H have the form (2.11).

A. The basic ODEs for the monochromatic radiation field

In this section we introduce the notation which will allow us to state and prove our proposition and two theorems.

In the monochromatic case of (2.11) the basic system in averaging form, (2.54) and (2.55), becomes

$$\theta' = \frac{K_r\mathcal{E}}{\varepsilon^2}\left(1 - \frac{1+\varepsilon\chi}{\Pi_z(\theta,\chi,\zeta,\varepsilon,\nu)}\right) + \frac{K_r q(\zeta)}{2}, \tag{3.1}$$

$$\chi' = -\varepsilon K^2 \frac{\Pi_x(\theta,\zeta,\varepsilon,\nu)}{\Pi_z(\theta,\chi,\zeta,\varepsilon,\nu)}\cos\{\nu[\theta - Q(\zeta)]\}, \tag{3.2}$$

with the initial conditions $\theta(0,\varepsilon) = \theta_0$, $\chi(0,\varepsilon) = \chi_0$. Recall that q and Q are defined in (2.44) and (2.48) and these can be rewritten

$$q(\zeta) = \bar{q} + 2K^2\Delta P_{x0}\cos\zeta + \frac{K^2}{2}\cos 2\zeta, \tag{3.3}$$

$$\bar{q} := 1 + \frac{1}{2}K^2 + K^2(\Delta P_{x0})^2, \tag{3.4}$$

$$Q(\zeta) = \frac{K_r\bar{q}}{2}\zeta + \Upsilon_0\sin\zeta + \Upsilon_1\sin 2\zeta, \tag{3.5}$$

$$\Upsilon_0 := K_r K^2\Delta P_{x0}, \qquad \Upsilon_1 := \frac{K_r K^2}{8}. \tag{3.6}$$

Clearly $\bar{q}$ is the average of $q(\zeta)$ over ζ. To make the arguments of P_x and P_z in (2.56) and (2.57) explicit, we have replaced them by Π_x and Π_z, where

$$\Pi_x(\theta,\zeta,\varepsilon,\nu) := \cos\zeta + \Delta P_{x0} + \frac{\varepsilon^2}{K_r\nu}(\sin\{\nu[\theta - Q(\zeta)]\} - \sin(\nu\theta_0)), \tag{3.7}$$

$$\Pi_z(\theta,\chi,\zeta,\varepsilon,\nu) := \sqrt{(1+\varepsilon\chi)^2 - \frac{\varepsilon^2}{\mathcal{E}}[1 + K^2\Pi_x^2(\theta,\zeta,\varepsilon,\nu)]}. \tag{3.8}$$

Note that because of the singularity for $\nu = 0$ in (3.7) we take $\nu \ge 1/2$ in the following. Equations (3.1) and (3.2) are the basic ODEs for the monochromatic case. They will lead us to the exact and approximate ODEs of both theorems [in fact (3.1) and (3.2) are the exact ODEs for Theorem 1].

From now on, we restrict ε to a finite interval $(0, \varepsilon_0]$. We are of course interested in ε small, i.e., $0 < \varepsilon \ll 1$, and so, without loss of generality, we take

$$0 < \varepsilon \le \varepsilon_0, \qquad 0 < \varepsilon_0 \le 1. \tag{3.9}$$

Consider the open set

$$\mathcal{D}(\varepsilon,\nu) := \left\{(\theta,\chi,\zeta) \in \mathbb{R}^3: \chi > -\frac{1}{\varepsilon} + \frac{1}{\sqrt{\mathcal{E}}}\sqrt{1 + K^2\Pi_x^2(\theta,\zeta,\varepsilon,\nu)}\right\}, \tag{3.10}$$

for $0 < \varepsilon \le \varepsilon_0$, $\nu \ge 1/2$, which was discussed after (2.41). The ODEs (3.1) and (3.2) are well defined on this domain as we now argue. We take the domain of Π_x to be $\{(\theta,\zeta,\varepsilon,\nu) \in \mathbb{R}^4: 0 < \varepsilon \le \varepsilon_0, \nu \ge 1/2\}$ and the domain of Π_z to be $\{(\theta,\chi,\zeta,\varepsilon,\nu) \in [\mathcal{D}(\varepsilon,\nu) \times \mathbb{R}^2]: 0 < \varepsilon \le \varepsilon_0, \nu \ge 1/2\}$. It is easy to check that on the domain of Π_z the argument of the square root in (3.8) is positive and, for $(\theta,\chi,\zeta) \in \mathcal{D}(\varepsilon,\nu)$, we have

$$0 < \Pi_z(\theta,\chi,\zeta,\varepsilon,\nu) < 1 + \varepsilon\chi. \tag{3.11}$$

The two singularities in (3.1) and (3.2) for $\varepsilon = 0$ and $\Pi_z = 0$ are excluded by (3.9) and (3.11). Thus, it is easy to check that the vector field associated with the ODEs (3.1) and (3.2) is well defined and C^∞ on $\mathcal{D}(\varepsilon,\nu)$ for $0 < \varepsilon \le \varepsilon_0 \le 1$, $\nu \ge 1/2$ [i.e., the vector field has partial derivatives in θ,

χ, ζ of all orders on $\mathcal{D}(\varepsilon, \nu)$]. Note that a function defined on an open set is said to be C^∞ if it has continuous partial derivatives of arbitrary order.

Since $\mathcal{D}(\varepsilon, \nu)$ is dependent on ε it is inconvenient to use it in our averaging theorems. Thus, we now restrict (θ, χ, ζ) to an ε independent subdomain $\mathcal{D}_0(\varepsilon_0)$ of $\mathcal{D}(\varepsilon, \nu)$, where

$$\mathcal{D}_0(\varepsilon_0) := \mathbb{R} \times [\chi_{lb}(\varepsilon_0), \infty] \times \mathbb{R} \subset \mathcal{D}(\varepsilon, \nu), \qquad (3.12)$$

$$\chi_{lb}(\varepsilon_0) := -\frac{1}{\varepsilon_0} + \frac{1}{\sqrt{\mathcal{E}}}\sqrt{1 + K^2\Pi^2_{x,ub}(\varepsilon_0)}, \qquad (3.13)$$

$$\Pi_{x,ub}(\varepsilon_0) := 1 + |\Delta P_{x0}| + 2\varepsilon_0^2\bar{q}. \qquad (3.14)$$

The subset condition in (3.12) and the fact that

$$|\Pi_x(\theta, \zeta, \varepsilon, \nu)| \le \Pi_{x,ub}(\varepsilon) \le \Pi_{x,ub}(\varepsilon_0) \qquad (3.15)$$

are easily checked. The domains $\mathcal{D}$ and $\mathcal{D}_0$ are illustrated in Fig. 1, where we show the "quasiperiodic" type surface $\chi = \chi_{\mathcal{D}}(\theta, \zeta, \varepsilon, \nu) = -\frac{1}{\varepsilon} + \frac{1}{\sqrt{\mathcal{E}}}\sqrt{1 + K^2\Pi_x^2(\theta, \zeta, \varepsilon, \nu)}$ of the lower boundary for $\mathcal{D}$ (labeled by $\chi_{\mathcal{D}}$) and the planar lower boundary $\chi = \chi_{lb}(\varepsilon_0)$ for $\mathcal{D}_0$ (labeled by χ_{lb}). Here $\chi_{\mathcal{D}}$ is given by the lower bound for χ in (3.10).

From (2.42) and (2.43) we have

$$\frac{1 + \varepsilon\chi}{\Pi_z} = 1 + \frac{q(\zeta)}{2\gamma_c^2}(1 - 2\varepsilon\chi) + O(\varepsilon^4), \qquad (3.16)$$

$$\frac{\Pi_x}{\Pi_z} = \cos\zeta + \Delta P_{x0} + O(\varepsilon). \qquad (3.17)$$

Thus, the basic 2D system of (3.1) and (3.2) can be written

$$\theta' = \varepsilon f_1(\chi, \zeta) + \varepsilon^2 g_1(\theta, \chi, \zeta, \varepsilon, \nu), \qquad (3.18)$$

$$\chi' = \varepsilon f_2(\theta, \zeta, \nu) + \varepsilon^2 g_2(\theta, \chi, \zeta, \varepsilon, \nu), \qquad (3.19)$$

where f_1, f_2 are given by

$$f_1(\chi, \zeta) := K_r q(\zeta)\chi, \qquad (3.20)$$

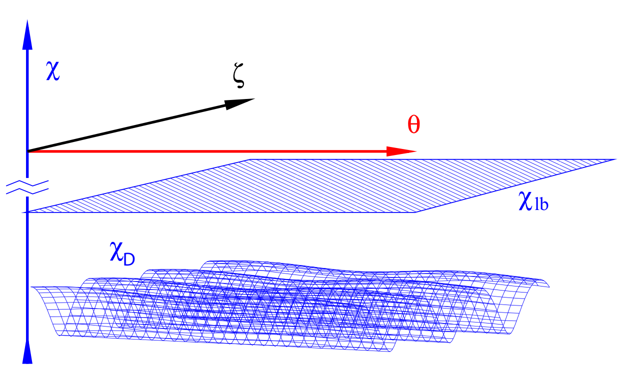

FIG. 1. A sketch of the domains $\mathcal{D}$ and $\mathcal{D}_0$.

$$f_2(\theta, \zeta, \nu) := -K^2(\cos\zeta + \Delta P_{x0})\cos\{\nu[\theta - Q(\zeta)]\}, \qquad (3.21)$$

so that the functions g_i: $\mathcal{D}_0(\varepsilon_0) \times (0, \varepsilon_0] \times [1/2, \infty) \to \mathbb{R}$ are given by

$$\varepsilon^2 g_1(\theta, \chi, \zeta, \varepsilon, \nu) := \frac{K_r\mathcal{E}}{\varepsilon^2}\left(1 - \frac{1 + \varepsilon\chi}{\Pi_z(\theta, \chi, \zeta, \varepsilon, \nu)}\right) + \frac{K_r q(\zeta)}{2}(1 - 2\varepsilon\chi), \qquad (3.22)$$

$$\varepsilon^2 g_2(\theta, \chi, \zeta, \varepsilon, \nu) := \varepsilon K^2\cos\{\nu[\theta - Q(\zeta)]\} \times\left[\cos\zeta + \Delta P_{x0} - \frac{\Pi_x(\theta, \zeta, \varepsilon, \nu)}{\Pi_z(\theta, \chi, \zeta, \varepsilon, \nu)}\right]. \qquad (3.23)$$

The ODEs (3.18) and (3.19) will be the subject of Theorem 1, i.e., the averaging theorem for the NonR case. In fact (3.18) and (3.19) are the exact ODEs for Theorem 1 and, unlike (3.1) and (3.2), they are written in a form which will allow us to derive the normal form system, i.e., the approximate ODEs for Theorem 1 (see also Sec. III C). The $\varepsilon = 0$ singularity in the definitions of g_1 and g_2 is removable. This is discussed at the end of Sec. III C and is proven in Appendix B. The ODEs (3.18) and (3.19) will also be used to obtain the exact and approximate ODEs for Theorem 2 (see Sec. III C).

B. Resonant, nonresonant, Δ-nonresonant, near-to-resonant

Here we discuss the resonant structure of the f_i defined in (3.20) and (3.21) needed for determining the normal form approximations. Most importantly the resonant structure of f_2 will lead us to the notions of resonant, nonresonant, Δ-nonresonant, and near-to-resonant normal form. The g_i play no role in the normal forms but are a central piece of the error bounds.

Clearly f_1 is 2π periodic in ζ. The periodic structure of f_2 can be made explicit using (3.5) which gives

$$f_2(\theta, \zeta, \nu) = -K^2(\cos\zeta + \Delta P_{x0}) \times\cos\left(\nu\theta - \nu\frac{K_r\bar{q}}{2}\zeta - \nu\Upsilon_0\sin\zeta - \nu\Upsilon_1\sin2\zeta\right) \qquad (3.24)$$

and shows the two base periodicities, 2π and $2\pi/\nu\frac{K_r\bar{q}}{2}$. We now choose

$$K_r = \frac{2}{\bar{q}}, \qquad (3.25)$$

so the base frequencies are 1 and ν, which is consistent with the FEL literature and with (1.3). As mentioned in Sec. I A, with (3.25) we have gone from seven to six parameters. Note also that f_2 is quasiperiodic [22].

To make the resonant structure explicit we write f_2 as

$$f_2(\theta,\zeta,\nu) = -\frac{K^2}{2}\exp[i\nu(\theta-\zeta)]\mathrm{jj}(\zeta;\nu,\Delta P_{x0}) + cc, \tag{3.26}$$

where cc denotes complex conjugate and

$$\mathrm{jj}(\zeta;\nu,\Delta P_{x0}) := (\cos\zeta + \Delta P_{x0})\exp(-i\nu[\Upsilon_0\sin\zeta + \Upsilon_1\sin 2\zeta]), \tag{3.27}$$

is 2π periodic in ζ. The Fourier series of jj is

$$\mathrm{jj}(\zeta;\nu,\Delta P_{x0}) \sim \sum_{n\in\mathbb{Z}} \hat{\hat{\mathrm{jj}}}(n;\nu,\Delta P_{x0})e^{in\zeta}, \tag{3.28}$$

with

$$\hat{\hat{\mathrm{jj}}}(n;\nu,\Delta P_{x0}) := \frac{1}{2\pi}\int_{[0,2\pi]} d\zeta\,\mathrm{jj}(\zeta;\nu,\Delta P_{x0})\mathrm{e}^{-in\zeta}, \tag{3.29}$$

and $\mathbb{Z}$ being the set of integers. Since $\mathrm{jj}(\cdot;\nu,\Delta P_{x0})$ is a 2π-periodic C^∞ function its Fourier series (3.28) is absolutely convergent, i.e., $\sum_{n\in\mathbb{Z}}|\hat{\hat{\mathrm{jj}}}(n;\nu,\Delta P_{x0})| < \infty$ whence $\sim$ in (3.28) can replaced by $=$. The f_2 in Eq. (3.19) can now be written

$$f_2(\theta,\zeta,\nu) = -\frac{K^2}{2}e^{i\nu\theta}\sum_{n\in\mathbb{Z}}\hat{\hat{\mathrm{jj}}}(n;\nu,\Delta P_{x0})e^{i(n-\nu)\zeta} + cc, \tag{3.30}$$

which shows the resonant structure in that the ζ average of f_2 is zero for $\nu \neq$ integer. In Appendix A we find

$$\begin{aligned}\hat{\hat{\mathrm{jj}}}(n;\nu,\Delta P_{x0}) = {}& \frac{1}{2}\mathcal{J}(n,1,\nu,\Upsilon_0,\Upsilon_1) \\ &+ \frac{1}{2}\mathcal{J}(n,-1,\nu,\Upsilon_0,\Upsilon_1) \\ &+ \Delta P_{x0}\mathcal{J}(n,0,\nu,\Upsilon_0,\Upsilon_1),\end{aligned} \tag{3.31}$$

where

$$\mathcal{J}(n,m,\nu,\Upsilon_0,\Upsilon_1) := \sum_{l\in\mathbb{Z}} J_{m-n-2l}(\nu\Upsilon_0)J_l(\nu\Upsilon_1), \tag{3.32}$$

and J_k is the kth order Bessel function of the first kind. Note that $\mathrm{jj}(-\zeta;\nu,\Delta P_{x0})$ is the complex conjugate of $\mathrm{jj}(\zeta;\nu,\Delta P_{x0})$ which implies $\hat{\hat{\mathrm{jj}}}(n;\nu,\Delta P_{x0})$ is real. This is confirmed in the explicit form of (3.31) and (3.32) since the J_k are real valued.

The time average of f_1 in (3.20) is

$$\bar{f}_1(\chi) := \lim_{T\to\infty}\left[\frac{1}{T}\int_0^T f_1(\chi,\zeta)d\zeta\right] = 2\chi. \tag{3.33}$$

Since $\overline{\exp[i(n-\nu)\zeta]} = \delta_{n,\nu}$, the time average of the quasiperiodic f_2 is

$$\bar{f}_2(\theta,\nu) := \lim_{T\to\infty}\left[\frac{1}{T}\int_0^T f_2(\theta,\zeta,\nu)d\zeta\right] = \begin{cases}0 & \text{if } \nu\in\mathbb{N} \quad \text{NonR} \\ -K^2\hat{\hat{\mathrm{jj}}}(k;k,\Delta P_{x0})\cos(k\theta) & \text{if } \nu = k\in\mathbb{N} \quad \text{resonant case,}\end{cases} \tag{3.34}$$

where $\mathbb{N}$ denotes the set of positive integers and where we have used the fact that $\hat{\hat{\mathrm{jj}}}$ is real. The integration in (3.34) can be done term by term since the series in (3.30) converges uniformly in ζ. The fact that $\bar{f}_2$ vanishes for $\nu\notin\mathbb{N}$ is due to the choice of K_r in (1.3) and (3.25). Without this choice $\bar{f}_2$ would vanish for $(\nu K_r\bar{q}/2)\notin\mathbb{N}$.

An averaging normal form system for (3.18) and (3.19) is obtained by dropping the $O(\varepsilon^2)$ terms and averaging the $O(\varepsilon)$ terms over ζ. Thus, if $\nu\notin\mathbb{N}$ the normal form system is

$$\theta' = \varepsilon 2\chi, \qquad \chi' = 0, \tag{3.35}$$

and if $\nu\in\mathbb{N}$ the normal form system for $\nu = k$ is

$$\theta' = \varepsilon 2\chi, \qquad \chi' = -\varepsilon K^2\hat{\hat{\mathrm{jj}}}(k;k,\Delta P_{x0})\cos(k\theta). \tag{3.36}$$

From Appendix A we have, for $\Delta P_{x0} = 0$,

$$\hat{\hat{\mathrm{jj}}}(k;k,0) = \begin{cases}\frac{1}{2}(-1)^n[J_n(x_n) - J_{n+1}(x_n)] & \text{if } k = 2n+1 \\ 0 & \text{if } k \text{ even,}\end{cases} \tag{3.37}$$

where $x_n := (2n+1)\Upsilon_1$ and $n = 0, 1, \ldots$ with Υ_1 defined in (3.6). Thus, for $\Delta P_{x0} = 0$, (3.36) gives the standard FEL pendulum system (see also [13,18,19,23]:

$$\theta' = \varepsilon 2\chi, \qquad \chi' = -\varepsilon K^2\hat{\hat{\mathrm{jj}}}(k;k,0)\cos(k\theta), \tag{3.38}$$

where k is an odd integer.

The basic question we consider in this paper is, when do normal form systems give good approximations to the exact ODEs (3.18) and (3.19)? In the following we consider two cases.

In the first case of (3.37), the "nonresonant" (NonR) case, we consider the situation when the dynamics is well approximated in terms of the system (3.35), which we call the NonR normal form system. However, because of a small divisor issue, we are forced to keep ν away from resonance. More precisely, we obtain results for $\nu\in[k+\Delta, k+1-\Delta]$ with $\Delta\in(0,1/2)$ and $k\in\mathbb{N}$. We call this subcase of the NonR case the "Δ-NonR" case. We begin the discussion of the Δ-NonR case in Sec. III C, state the associated Theorem 1 in Sec. III E 1 and prove

Theorem 1 in Sec. IV A. We show that the error is $O(\varepsilon/\Delta)$. Thus, the error bound increases as $\Delta \to 0$.

In the second case of (3.37) we consider, in part, the situation when the dynamics is well approximated by the normal form system of (3.36), which we call the "resonant normal form system." However, we will do better in that we will explore the dynamics near the resonance. More precisely we explore $O(\varepsilon)$ neighborhoods of the $\nu = k$ resonances and parametrize ν by $\nu = k + \varepsilon a$, where $k \in \mathbb{N}$, $a \in [-1/2, 1/2]$. Recall that $0 < \varepsilon \leq \varepsilon_0 \leq 1$. The "near-to-resonant" (NearR) normal form system will be introduced in Sec. III D and in the subcase where $\nu = k$ we obtain the above resonant normal form system. The associated Theorem 2 will be stated in Sec. III E 2 and its proof will be given in Sec. IV B. We note that $\nu = 1$ is the primary resonance as discussed in the Introduction, further justifying the choice of K_r in (1.3) and (3.25). We call this case the NearR case. Note that as in the NonR case the exact ODEs are (3.18) and (3.19) where of course in the NearR case one takes $\nu = k + \varepsilon a$.

C. The nonresonant case and its normal form

The exact ODEs for (θ, χ) in the NonR case are (3.18) and (3.19). The vector field in (3.18) and (3.19) as well as the dependence of g_1, g_2 on (θ, χ, ζ) are C^∞ on $\mathcal{D}_0(\varepsilon_0)$.

As introduced in Sec. III B, the NonR normal form system is obtained from (3.18) and (3.19) by dropping the $O(\varepsilon^2)$ terms and averaging the rhs over ζ holding the slowly varying quantities θ, χ fixed. Using (3.33) and (3.34), we obtain as before

$$v_1' = \varepsilon \bar{f}_1(v_2) = \varepsilon 2 v_2, \tag{3.39}$$

$$v_2' = \varepsilon \bar{f}_2(v_1, \nu) = 0, \tag{3.40}$$

with the same initial conditions as in the exact ODEs, i.e., $v_1(0, \varepsilon) = \theta_0$, $v_2(0, \varepsilon) = \chi_0$. Here we have introduced new dependent variables in order to distinguish normal form solutions from the exact solutions.

The solution of the IVP is

$$v_1(\zeta, \varepsilon) = 2\chi_0 \varepsilon \zeta + \theta_0, \qquad v_2(\zeta, \varepsilon) = \chi_0. \tag{3.41}$$

This means that in the NonR case the normal form approximation to the energy deviation is constant while the normal form approximation to the ponderomotive phase advances linearly with ζ. Clearly the associated phase plane portrait is simply a family of horizontal lines.

The solutions of (3.39) and (3.40) with $\varepsilon = 1$ play an important role in the statement and proof of Theorem 1 and we refer to

$$\mathbf{v}(\cdot, 1) = [v_1(\cdot, 1), v_2(\cdot, 1)], \tag{3.42}$$

as the guiding solution at (θ_0, χ_0). Note that the $\mathbf{v}$ in (3.42) should not be confused with the velocity vector $\mathbf{v}$ in (2.3).

Our basic result in the NonR case will be Theorem 1 by which $|\theta(\zeta) - v_1(\zeta, \varepsilon)|$ and $|\chi(\zeta) - v_2(\zeta, \varepsilon)|$ are $O(\varepsilon/\Delta)$ in the Δ-NonR subcase, where $\nu \in [k + \Delta, k + 1 - \Delta]$. Note that Theorem 1 can be applied to every $\nu \geq 1/2$ with $\nu \notin \mathbb{N}$. However, for fixed ε, as $\nu \to k \in \mathbb{N}$, $\Delta \to 0^+$ and the error bound becomes large and thus useless. We can also consider Δ to be a function of ε as discussed in item (1) of Sec. III E 3. For example if $\Delta = O(\varepsilon)$ then the error is $O(1)$ and thus not very interesting.

The precise statement of Theorem 1 is given in Sec. III E 1 and its proof is given in Sec. IV A. It will become clear that the error bounds require g_1 and g_2 to be bounded independent of ε in a neighborhood of the normal form solutions, and while f_1 and f_2 were chosen with this in mind a proof of boundedness will be an important part of our proof of Theorem 1. We show in Appendix B that

$$\begin{aligned}\lim_{\varepsilon\to 0+}&[g_1(\theta, \chi, \zeta, \varepsilon, \nu)] \\ &= -\frac{q(\zeta)}{4\bar{q}}\left(\frac{3q(\zeta)}{\mathcal{E}} + 12\chi^2\right) \\ &\quad - \frac{K^2}{\nu}(\sin\{\nu[\theta - Q(\zeta)]\} - \sin(\nu\theta_0))(\cos\zeta + \Delta P_{x0}),\end{aligned} \tag{3.43}$$

$$\begin{aligned}\lim_{\varepsilon\to 0+}&[g_2(\theta, \chi, \zeta, \varepsilon, \nu)] \\ &= K^2 \chi \cos\{\nu[\theta - Q(\zeta)]\}(\cos\zeta + \Delta P_{x0}).\end{aligned} \tag{3.44}$$

Thus, the $\varepsilon = 0$ singularity for g_1 and g_2 in (3.22) and (3.23) is removable [in fact, g_1 and g_2 are rewritten in (B6) and (B13) of Appendix B so that the singularity is removed]. This makes second-order averaging possible and the normal form equations would be augmented by averages of these limits. Second-order averaging is discussed briefly in Sec. V.

D. The near-to-resonant case and its normal form

1. The near-to-resonant system

According to Sec. III B we have, in the NearR case,

$$\nu = k + \varepsilon a, \tag{3.45}$$

where $k \in \mathbb{N}$ and $a \in [-1/2, 1/2]$ is a measure of the distance of ν from k. The $O(\varepsilon)$ neighborhood of k is natural in first-order averaging, since if $|\nu - k|$ is too small [e.g., $O(\varepsilon^2)$] then the normal form will be the resonant normal form of (3.36) and if $|\nu - k|$ is too big then ν will be in the Δ-NonR regime. Equation (3.45) clearly includes the resonant case for $a = 0$.

To derive the NearR normal form system, we start from (3.18) and (3.19), use (3.45) and obtain

$$\theta' = \varepsilon f_1(\chi, \zeta) + \varepsilon^2 g_1(\theta, \chi, \zeta, \varepsilon, k + \varepsilon a), \tag{3.46}$$

$$\chi' = \varepsilon f_2(\theta, \zeta, k+\varepsilon a) + \varepsilon^2 g_2(\theta, \chi, \zeta, \varepsilon, k+\varepsilon a), \quad (3.47)$$

with initial conditions $\theta(0, \varepsilon) = \theta_0$, $\chi(0, \varepsilon) = \chi_0$. Note that (3.46) and (3.47) are the exact ODEs for the NearR case.

Since f_1 in (3.46) is independent of ε the normal form associated with it will be the same as in the NonR case. We now need to study the ε dependence of f_2 in (3.47). From (3.26),

$$\begin{aligned} f_2(\theta, \zeta, k+\varepsilon a) &= -\frac{K^2}{2}\exp[i(k+\varepsilon a)(\theta-\zeta)]\mathrm{jj}(\zeta; k+\varepsilon a, \Delta P_{x0}) + cc \\ &= -\frac{K^2}{2}\exp(i[k\theta - \varepsilon a\zeta]) \\ &\quad\times \exp(-ik\zeta)\mathrm{jj}(\zeta; k, \Delta P_{x0}) \\ &\quad\times \exp(i\varepsilon a[\theta - \Upsilon_0 \sin\zeta - \Upsilon_1 \sin 2\zeta]) + cc, \end{aligned} \quad (3.48)$$

where we have used from (3.27) that

$$\begin{aligned} \mathrm{jj}(\zeta; k+\varepsilon a, \Delta P_{x0}) &= (\cos\zeta + \Delta P_{x0})\exp\{-i(k+\varepsilon a)[\Upsilon_0 \sin\zeta + \Upsilon_1 \sin 2\zeta]\} \\ &= \mathrm{jj}(\zeta; k, \Delta P_{x0})\exp(-i\varepsilon a[\Upsilon_0 \sin\zeta + \Upsilon_1 \sin 2\zeta]). \end{aligned} \quad (3.49)$$

For $a = 0$ the resonant normal form of (3.36) can be obtained from (3.48). For $a \neq 0$ (3.48) displays two ε dependencies. The first is the $\varepsilon a\zeta$ one which cannot be expanded since it is $O(1)$ for $\zeta = O(1/\varepsilon)$ the upper range of our averaging theorem. The second is the εa factor in the final exponential which can be expanded and makes an $O(1)$ contribution to g_2 in (3.47) for all ζ. Therefore we rewrite f_2 as

$$f_2(\theta, \zeta, k+\varepsilon a) = f_2^R(\theta, \varepsilon\zeta, \zeta, k, a) + O(\varepsilon), \quad (3.50)$$

where

$$\begin{aligned} f_2^R(\theta, \tau, \zeta, k, a) &:= -\frac{K^2}{2}\exp(i[k\theta - a\tau])\exp(-ik\zeta)\mathrm{jj}(\zeta; k, \Delta P_{x0}) + cc \\ &= -\frac{K^2}{2}\exp(i[k\theta - a\tau])\sum_{n\in\mathbb{Z}}\widehat{\mathrm{jj}}(n; k, \Delta P_{x0})e^{i\zeta[n-k]} + cc. \end{aligned} \quad (3.51)$$

We can now write the exact NearR ODEs (3.46) and (3.47) in a form appropriate for the MoA. From (3.46)–(3.51) we obtain

$$\theta' = \varepsilon f_1^R(\chi, \zeta) + \varepsilon^2 g_1^R(\theta, \chi, \zeta, \varepsilon, k, a), \quad (3.52)$$

$$\chi' = \varepsilon f_2^R(\theta, \varepsilon\zeta, \zeta, k, a) + \varepsilon^2 g_2^R(\theta, \chi, \zeta, \varepsilon, k, a), \quad (3.53)$$

where

$$f_1^R(\chi, \zeta) := f_1(\chi, \zeta) = \frac{2q(\zeta)\chi}{\bar{q}}, \quad (3.54)$$

and where the functions $g_i^R\colon \mathcal{D}_0(\varepsilon_0) \times (0, \varepsilon_0] \times \mathbb{N} \times [-1/2, 1/2] \to \mathbb{R}$ are given by

$$g_1^R(\theta, \chi, \zeta, \varepsilon, k, a) := g_1(\theta, \chi, \zeta, \varepsilon, k+\varepsilon a), \quad (3.55)$$

$$\begin{aligned} g_2^R(\theta, \chi, \zeta, \varepsilon, k, a) &:= g_2(\theta, \chi, \zeta, \varepsilon, k+\varepsilon a) \\ &\quad + \frac{1}{\varepsilon}[f_2(\theta, \zeta, k+\varepsilon a) - f_2^R(\theta, \varepsilon\zeta, \zeta, k, a)]. \end{aligned} \quad (3.56)$$

By (3.24) we have

$$\begin{aligned} f_2(\theta, \zeta, k+\varepsilon a) &= -K^2(\cos\zeta + \Delta P_{x0})\cos\{(k+\varepsilon a)[\theta - \zeta - \Upsilon_0 \sin\zeta \\ &\quad - \Upsilon_1 \sin 2\zeta]\}, \end{aligned} \quad (3.57)$$

and, by (3.27) and (3.51),

$$\begin{aligned} f_2^R(\theta, \varepsilon\zeta, \zeta, k, a) &= -\frac{K^2}{2}\exp(i[k\theta - \varepsilon a\zeta])\exp(-ik\zeta)(\cos\zeta + \Delta P_{x0}) \\ &\quad\times \exp(-ik[\Upsilon_0 \sin\zeta + \Upsilon_1 \sin 2\zeta]) + cc \\ &= -K^2(\cos\zeta + \Delta P_{x0})\cos(k[\theta - \zeta - \Upsilon_0 \sin\zeta \\ &\quad - \Upsilon_1 \sin 2\zeta] - \varepsilon a\zeta). \end{aligned} \quad (3.58)$$

Note that the rhs of (3.46), (3.47), (3.52), and (3.53) are equal. Thus, by the remarks at the beginning of Sec. III C the vector field in (3.52) and (3.53) as well as the dependence of g_1^R, g_2^R on (θ, χ, ζ) are C^∞ on $\mathcal{D}_0(\varepsilon_0)$.

We show in Appendix D that

$$\begin{aligned} \lim_{\varepsilon\to 0+}[g_1^R(\theta, \chi, \zeta, \varepsilon, k, a)] &= -\frac{q(\zeta)}{4\bar{q}}\left(\frac{3}{\mathcal{E}}q(\zeta) + 12\chi^2\right) \\ &\quad - \frac{K^2}{k}(\sin\{k[\theta - Q(\zeta)]\} \\ &\quad - \sin(k\theta_0))(\cos\zeta + \Delta P_{x0}), \end{aligned} \quad (3.59)$$

$$\begin{aligned} \lim_{\varepsilon\to 0+}[g_2^R(\theta, \chi, \zeta, \varepsilon, k, a)] &= \chi K^2 \cos\{k[\theta - Q(\zeta)]\}(\cos\zeta + \Delta P_{x0}) \\ &\quad + K^2 a(\theta - \Upsilon_0 \sin\zeta - \Upsilon_1 \sin 2\zeta) \\ &\quad \times \sin(k[\theta - \zeta - \Upsilon_0 \sin\zeta - \Upsilon_1 \sin 2\zeta]) \\ &\quad \times (\cos\zeta + \Delta P_{x0}). \end{aligned} \quad (3.60)$$

Thus, the $\varepsilon = 0$ singularity for g_1^R and g_2^R in (3.55) and (3.56) is removable. This makes second-order averaging possible and the normal form equations would be augmented by averages of these limits. This will be discussed briefly in Sec. V.

2. The near-to-resonant normal form

The NearR normal form system is obtained from (3.52) and (3.53) by dropping the $O(\varepsilon^2)$ terms and averaging the rhs over ζ holding the slowly varying quantities θ, χ, $\varepsilon a\zeta$ fixed.

We thus obtain from (3.51)–(3.54) that

$$v_1' = \varepsilon \bar{f}_1^R(v_2) = 2\varepsilon v_2, \tag{3.61}$$

$$v_2' = \varepsilon \bar{f}_2^R(v_1, \varepsilon\zeta, k) = -\varepsilon \mathcal{A}(k, \Delta P_{x0}) \cos(kv_1 - \varepsilon a\zeta), \tag{3.62}$$

where

$$\mathcal{A}(k, \Delta P_{x0}) := K^2 \hat{\hat{\mathrm{jj}}}(k; k, \Delta P_{x0}), \tag{3.63}$$

and the same initial conditions as in the exact ODEs, i.e., $v_1(0, \varepsilon) = \theta_0$, $v_2(0, \varepsilon) = \chi_0$.

For $\Delta P_{x0} = a = 0$, (3.61) and (3.62) are the standard FEL pendulum equations, given by (3.37) and (3.38). In the special case when $\mathcal{A}(k, \Delta P_{x0}) = 0$ the ODEs (3.61) and (3.62) are the same as the NonR equations (3.39) and (3.40). We note that $\mathcal{A} = 0$ occurs when $\Delta P_{x0} = 0$ and k even [see the remark after (A11)] and thus the well-known fact that even harmonics vanish on axis emerges quite naturally.

The ultimate justification for the normal form (3.61) and (3.62) comes from the averaging theorem itself. However, if we replace $\varepsilon\zeta$ in (3.53) by τ and add the equation $\tau' = \varepsilon$ then this, together with (3.52) and (3.53), is in a standard form for "periodic averaging" (= averaging over a periodic function) and the normal form (3.61) and (3.62) is obtained by averaging over ζ holding θ, χ, τ fixed. In this θ, χ, τ formulation standard periodic averaging theorems apply for the 3D system of θ, χ, τ, see, e.g., [6,14] and Sec. 3.3 in [10]. We will however prove an averaging theorem directly tuned to (3.52) and (3.53) both to show the reader a proof in a simple context and in addition we obtain nearly optimal error bounds which are stronger than in those standard theorems.

Before discussing the NearR normal form solutions in detail in the next subsection we discuss them briefly here. Replacing $\varepsilon\zeta$ by τ in the normal form system we obtain

$$v_1' = 2v_2, \tag{3.64}$$

$$v_2' = -\mathcal{A}(k, \Delta P_{x0}) \cos(kv_1 - a\tau). \tag{3.65}$$

For $a = 0$, (3.64) and (3.65) become the resonant normal form of (3.36) with $\varepsilon = 1$. The phase plane portrait (PPP) for this autonomous case with $k = 1$, $\mathcal{A} = 2$ is shown by the solid magenta, blue, and red lines in Fig. 2 and is seen to have the pendulum phase plane structure with libration (magenta), separatrix motion (blue), and rotation (red). There is a special solution given by $v_1 = (a\tau + \pi/2 + 2\pi n)/k$, $v_2 = a/2k$. To help understand the NearR behavior we have superposed orbits for the nonautonomous case of $a = 1/2$ for four initial conditions with $v_1(0) = -5\pi/2$. For $v_2(0) = a/2$ we see the special solution just mentioned given by the green horizontal line, for $v_2(0)$ starting on top of the libration curve we see a spiral motion given by the dotted magenta curve, for $v_2(0)$ starting on the lower rotation curve we see the orbit moving to the left, given by the red dotted curve, so the rotation dominates over the $a\tau$ [this can be seen in (3.75) where in this case the evolution of X moving to the left dominates over the $a\varepsilon\zeta$ which moves to the right] and finally for $v_2(0)$ starting on the upper rotation curve we see a modification of the rotation curve moving to the right given by the red dotted curve [in (3.75) the evolution of X and $a\varepsilon\zeta$ both move to the right]. The time behavior of these orbits is shown in Figs. 3 and 4. In Fig. 3 we show v_1 as a function of τ and in Fig. 4 we show v_2 as a function of τ for the initial conditions in Fig. 2 and the same τ intervals as given in Fig. 2.

3. Structure of the near-to-resonant normal form solutions

Here we transform the nonautonomous normal form system (3.61) and (3.62) to an autonomous system which has a pendulum-type phase plane structure. We then transform the autonomous system to a simple pendulum system allowing us to write the solutions of (3.61) and (3.62) in terms of solutions of the simple pendulum system. We then discuss the normal form solutions in detail in terms of the simple pendulum solutions at the level necessary for the averaging theorems [24].

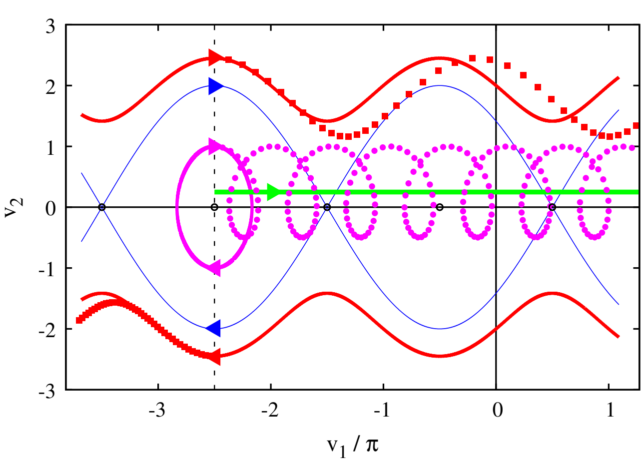

FIG. 2. Phase plane orbits on resonance ($a = 0$: solid magenta, blue, red curves, and five black fixed points) and NearR ($a = 1/2$: green solid and dotted magenta and red curves). $k = 1$, $\mathcal{A} = 2$.

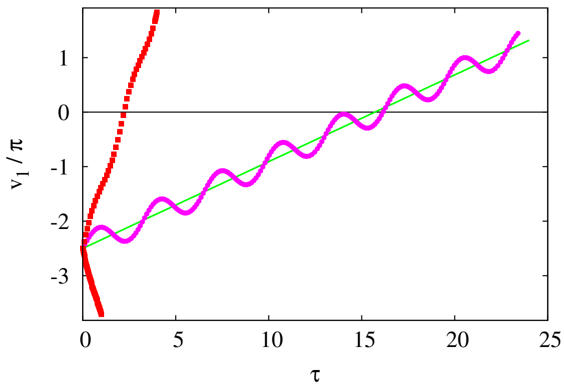

FIG. 3. $v_1(\tau)$ versus τ for $a = 1/2$, $k = 0$, $\mathcal{A} = 2$ as in Fig. 2.

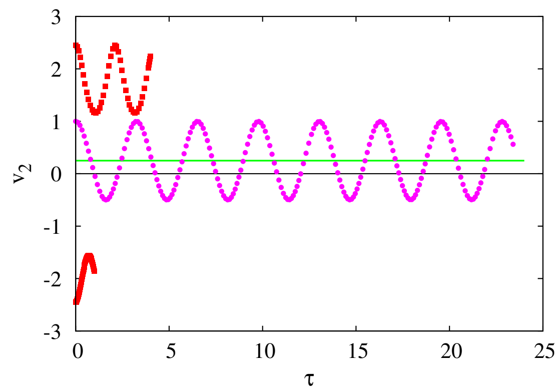

FIG. 4. $v_2(\tau)$ versus τ for $a = 1/2$, $k = 0$, $\mathcal{A} = 2$ as in Fig. 2.

Let $\mathbf{v} = (v_1, v_2)$, then it is easy to see that

$$\mathbf{v}(\zeta, \varepsilon) = \mathbf{v}(\varepsilon\zeta, 1). \tag{3.66}$$

The transformation $\mathbf{v}(\tau, 1) \longrightarrow \hat{\mathbf{v}}(\tau)$ via

$$\hat{\mathbf{v}}(\tau) = \begin{pmatrix} \hat{v}_1(\tau) \\ \hat{v}_2(\tau) \end{pmatrix} := \begin{pmatrix} kv_1(\tau, 1) - a\tau \\ v_2(\tau, 1) \end{pmatrix} \tag{3.67}$$

gives

$$\frac{d\hat{v}_1}{d\tau} = 2k\hat{v}_2 - a, \qquad \hat{v}_1(0) = k\theta_0, \tag{3.68}$$

$$\frac{d\hat{v}_2}{d\tau} = -\mathcal{A}(k, \Delta P_{x0})\cos\hat{v}_1, \qquad \hat{v}_2(0) = \chi_0. \tag{3.69}$$

Thus, we have scaled away the ε and made the transformed system autonomous. Solution properties of (3.68) and (3.69) are easily understood in terms of its phase plane portrait (PPP). However, it is more convenient to transform it to the simple pendulum system,

$$X' = Y, \qquad Y' = -\sin X, \tag{3.70}$$

$$X(0; Z_0) =: X_0, \qquad Y(0; Z_0) =: Y_0, \qquad Z_0 := \begin{pmatrix} X_0 \\ Y_0 \end{pmatrix}. \tag{3.71}$$

The required transformation is

$$\hat{v}_1(\tau) = X(\Omega\tau; Z_0) - \mathrm{sgn}(\mathcal{A})\frac{\pi}{2}, \tag{3.72}$$

$$\hat{v}_2(\tau) = \frac{\Omega Y(\Omega\tau; Z_0) + a}{2k}, \tag{3.73}$$

where

$$\Omega = \Omega(k) := \sqrt{2k|\mathcal{A}(k, \Delta P_{x0})|}. \tag{3.74}$$

From (3.66), (3.67), (3.72), and (3.73), the solutions of (3.61) and (3.62) are represented by

$$v_1(\zeta, \varepsilon) = \frac{X(\Omega\varepsilon\zeta; Z_0) - \mathrm{sgn}(\mathcal{A})\frac{\pi}{2} + \varepsilon a\zeta}{k}, \tag{3.75}$$

$$v_2(\zeta, \varepsilon) = \frac{\Omega Y(\Omega\varepsilon\zeta; Z_0) + a}{2k}, \tag{3.76}$$

where

$$Z_0(\theta_0, \chi_0, k, a) = \begin{pmatrix} X_0(\theta_0, k) \\ Y_0(\chi_0, k, a) \end{pmatrix} = \begin{pmatrix} k\theta_0 + \mathrm{sgn}(\mathcal{A}(k, \Delta P_{x0}))\frac{\pi}{2} \\ (2k\chi_0 - a)/\Omega(k) \end{pmatrix}. \tag{3.77}$$

We now discuss the solution properties of (3.61) and (3.62) in terms of the simple pendulum PPP, [25], for (3.70) using (3.75) and (3.76). The equilibria of (3.70) are at $(X, Y) = (\pi l, 0)$ with integer l. The systems obtained by linearizing about these equilibria are centers for l even and saddle points for l odd. From the theory of almost linear systems (see, e.g., [26]), it follows that the equilibria are centers and saddle points for the nonlinear system. A conservation law for the simple pendulum system is easily derived by first noting that the direction field is given by

$$\frac{dY}{dX} = -\frac{\sin X}{Y}. \tag{3.78}$$

This equation is separable and has solutions given implicitly by $\frac{1}{2}Y^2 + 1 - \cos X = \mathrm{const}$. Thus $\mathcal{E}_{\mathrm{Pen}}$, defined by

$$\mathcal{E}_{\mathrm{Pen}}(X, Y) := \frac{1}{2}Y^2 + U(X), \quad U(X) = 1 - \cos X, \tag{3.79}$$

is a constant of the motion which is easily checked directly. Incidentally $\mathcal{E}_{\mathrm{Pen}}$ is also a Hamiltonian for the ODEs (3.61) and (3.62) but this plays no role here. The PPP is easily constructed from the so-called potential plane which is simply a plot of the potential $U(X)$ vs X, see [27]. The PPP shows that the solutions of the simple pendulum system have four types of behavior, the equilibria mentioned above, libration, rotation, and separatrix motion.

These can be characterized in terms of $\mathcal{E}_{\rm Pen}$. Clearly, $\mathcal{E}_{\rm Pen}$ is non-negative, the centers correspond to $\mathcal{E}_{\rm Pen}(X, Y) = 0$ and the saddle points and separatrices to $\mathcal{E}_{\rm Pen}(X, Y) = 2$. The motion is libration for $0 < \mathcal{E}_{\rm Pen}(X, Y) < 2$, rotation for $\mathcal{E}_{\rm Pen}(X, Y) > 2$, and separatrix motion for $\mathcal{E}_{\rm Pen}(X, Y) = 2$ with $Y \neq 0$. In the libration case the solutions are periodic, which is easy to show, and the period as a function of amplitude [28] is given by

$$T(A) = 2\sqrt{2}\int_0^A \frac{dt}{[\cos t - \cos A]^{1/2}}, \quad (0 < A < \pi), \tag{3.80}$$

where $T(A)$ is the period associated with the initial conditions $X_0 = A$, $Y_0 = 0$. It is easy to show that $\lim_{A\to 0} T(A) = 2\pi$.

We denote by $\mathcal{B}_n$ the nth pendulum bucket which is defined by

$$\mathcal{B}_n := \{(X, Y) \in \mathbb{R}^2 \colon \mathcal{E}_{\rm Pen}(X, Y) < 2, |X - 2\pi n| < \pi\}, \tag{3.81}$$

with $n \in \mathbb{Z}$. Note that, by (3.77) and (3.79),

$$\begin{aligned}\mathcal{E}_{\rm Pen}[Z_0(\theta_0, \chi_0, k, a)] &= \mathcal{E}_R(\theta_0, \chi_0, k, a) := \frac{1}{2}\left[\frac{2k\chi_0 - a}{\Omega(k)}\right]^2 \\ &\quad + 1 + \mathrm{sgn}(\mathcal{A})\sin(k\theta_0).\end{aligned} \tag{3.82}$$

Note also that, by (3.75)–(3.77),

$$\begin{aligned}|v_1(\zeta, \varepsilon) - \theta_0| &= \left|\frac{X(\Omega\varepsilon\zeta; Z_0) - X_0 + \varepsilon a\zeta}{k}\right| \\ &\leq \frac{|X(\Omega\varepsilon\zeta; Z_0) - X_0| + \varepsilon|a|\zeta}{k},\end{aligned} \tag{3.83}$$

$$|v_2(\zeta, \varepsilon) - \chi_0| = \frac{\Omega}{2k}|Y(\Omega\varepsilon\zeta; Z_0) - Y_0|, \tag{3.84}$$

$$|v_2(\zeta, \varepsilon)| \leq \frac{\Omega|Y(\Omega\varepsilon\zeta; Z_0)| + |a|}{2k}. \tag{3.85}$$

We can now discuss the four cases of equilibria, libration, rotation, and separatrix motion. In each case, using (3.83)–(3.85), we will find $d_1^{\min}$, $d_2^{\min}$, $\chi_\infty \geq 0$ such that, for all $\zeta \geq 0$,

$$\begin{aligned}|v_1(\zeta, \varepsilon) - \theta_0| &\leq d_1^{\min}(\theta_0, \chi_0, \varepsilon\zeta, k, a), \\ |v_2(\zeta, \varepsilon) - \chi_0| &\leq d_2^{\min}(\theta_0, \chi_0, k, a),\end{aligned} \tag{3.86}$$

$$|v_2(\zeta, \varepsilon)| \leq \chi_\infty(\theta_0, \chi_0, k, a), \tag{3.87}$$

and we will at the same time observe that $d_1^{\min}(\theta_0, \chi_0, \tau, k, a)$ is increasing with respect to (wrt) τ. These quantities will be used in our statement and proofs of the averaging theorems.

(I) Equilibria regime.—$Y_0 = 0$ and either $\mathcal{E}_{\rm Pen}(X_0, Y_0) = 0$ or 2. Clearly $X_0 = \pi l$ where $l \in \mathbb{Z}$ and, by (3.77),

$$\begin{pmatrix} k\theta_0 + \mathrm{sgn}[\mathcal{A}(k, \Delta P_{x0})]\frac{\pi}{2} \\ (2k\chi_0 - a)/\Omega(k)\end{pmatrix} = Z_0(\theta_0, \chi_0, k, a) = \begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} \pi l \\ 0\end{pmatrix}, \tag{3.88}$$

so that $\theta_0 = \{\pi l - \mathrm{sgn}[\mathcal{A}(k, \Delta P_{x0})]\frac{\pi}{2}\}/k$ and $\chi_0 = a/2k$. Thus, by (3.75) and (3.76),

$$v_1(\zeta, \varepsilon) = \theta_0 + \frac{\varepsilon a\zeta}{k}, \tag{3.89}$$

$$v_2(\zeta, \varepsilon) = \chi_0. \tag{3.90}$$

Clearly, by direct substitution, these are solutions of (3.61) and (3.62). Incidentally these solutions are stable for l even and unstable for l odd.

Clearly, due to (3.86), (3.87), (3.89), and (3.90), we can choose

$$\begin{aligned} d_1^{\min}(\theta_0, \chi_0, \varepsilon\zeta, k, a) &:= \frac{\varepsilon|a|\zeta}{k}, \\ d_2^{\min}(\theta_0, \chi_0, k, a) &:= 0,\end{aligned} \tag{3.91}$$

$$\chi_\infty(\theta_0, \chi_0, k, a) := |\chi_0|. \tag{3.92}$$

(II) Libration regime.—$0 < \mathcal{E}_{\rm Pen}(X_0, Y_0) < 2$. In this case $Z_0(\theta_0, \chi_0, k, a) \in \mathcal{B}_{n(\theta_0, k)}$ where the integer $n = n(\theta_0, k)$ is determined by the condition $|X_0(\theta_0, k) - 2\pi n(\theta_0, k)| < \pi$. From (3.75) and (3.76) we see that

$$\mathbf{v}(\zeta, \varepsilon) = \mathbf{v}_{\rm per}(\zeta, \varepsilon) + \mathbf{v}_{\rm lin}(\varepsilon\zeta), \tag{3.93}$$

and it is easy to show that the periodic part has amplitude determined by the max and min values of X and Y and the linear growth term is

$$\mathbf{v}_{\rm lin}(\varepsilon\zeta) = \begin{pmatrix} \varepsilon a\zeta/k \\ 0 \end{pmatrix}. \tag{3.94}$$

The maximum values $X_{\max}$ and $Y_{\max}$ of X and Y satisfy, by (3.79),

$$\mathcal{E}_{\rm Pen}(Z_0) = \frac{1}{2}Y_0^2 + 1 - \cos X_0 = \frac{1}{2}Y_{\max}^2 = 1 - \cos X_{\max}, \tag{3.95}$$

whence

$$\begin{aligned} X_{\max}(\theta_0, \chi_0, k, a) &= 2\pi n(\theta_0, k) + \arccos\left(\cos X_0 - \frac{1}{2}Y_0^2\right) \\ &= 2\pi n(\theta_0, k) + \arccos[1 - \mathcal{E}_R(\theta_0, \chi_0, k, a)], \\ Y_{\max}(\theta_0, \chi_0, k, a) &:= \sqrt{2\mathcal{E}_{\rm Pen}[Z_0(\theta_0, \chi_0, k, a)]} \\ &= \sqrt{2\mathcal{E}_R(\theta_0, \chi_0, k, a)},\end{aligned} \tag{3.96}$$

and the minimum values $X_{\min}$ and $Y_{\min}$ of X and Y are given by

$$X_{\min} := 4\pi n - X_{\max}, \qquad Y_{\min} := -Y_{\max}. \tag{3.97}$$

Here arccos is the principle branch of the inverse cos mapping $[-1, 1] \to [0, \pi]$.

We now determine $d_1^{\min}$, $d_2^{\min}$, and χ_∞. It follows from (3.83)–(3.87), (3.96), and (3.97) that

$$\begin{aligned}|v_1(\zeta,\varepsilon)-\theta_0| &\le \frac{|X(\Omega\varepsilon\zeta;Z_0)-X_0|+\varepsilon|a|\zeta}{k}\\ &\le \frac{2X_{\max}(\theta_0,\chi_0,k,a)-4\pi n(\theta_0,k)+\varepsilon|a|\zeta}{k}\\ &= \frac{2\arccos[1-\mathcal{E}_R(\theta_0,\chi_0,k,a)]+\varepsilon|a|\zeta}{k}\\ &=: d_1^{\min}(\theta_0,\chi_0,\varepsilon\zeta,k,a), \end{aligned} \tag{3.98}$$

$$\begin{aligned}|v_2(\zeta,\varepsilon)-\chi_0| &= \frac{\Omega}{2k}|Y(\Omega\varepsilon\zeta;Z_0)-Y_0|\\ &\le \frac{\Omega}{k}Y_{\max}(\theta_0,\chi_0,k,a)\\ &= \frac{\Omega(k)}{k}\sqrt{2\mathcal{E}_R(\theta_0,\chi_0,k,a)}\\ &=: d_2^{\min}(\theta_0,\chi_0,k,a), \end{aligned} \tag{3.99}$$

$$\begin{aligned}|v_2(\zeta,\varepsilon)| &\le \frac{\Omega|Y(\Omega\varepsilon\zeta;Z_0)|+|a|}{2k}\\ &\le \frac{\Omega Y_{\max}(\theta_0,\chi_0,k,a)+|a|}{2k}\\ &= \frac{\Omega(k)\sqrt{2\mathcal{E}_R(\theta_0,\chi_0,k,a)}+|a|}{2k}\\ &=: \chi_\infty(\theta_0,\chi_0,k,a). \end{aligned} \tag{3.100}$$

(III) Separatrix regime.—$Y_0 \neq 0$ and $\mathcal{E}_{\rm Pen}(X_0, Y_0) = 2$. In this case $(X, Y) \in \overline{\mathcal{B}_{n(\theta_0,k)}}$ where the integer $n = n(\theta_0, k)$ is determined such that $|X_0(\theta_0, k) - 2\pi n(\theta_0, k)| < \pi$. Clearly

$$|X-X_0| \le 2\pi, \qquad |Y-Y_0| \le \sqrt{2\mathcal{E}_{\rm Pen}(X_0,Y_0)} = 2, \qquad |Y| \le 2. \tag{3.101}$$

For $Y_0 > 0$, $[X(t), Y(t)] \to [(2n+1)\pi, 0]$ as $t \to \infty$ and, for $Y_0 < 0$, $[X(t), Y(t)] \to [(2n-1)\pi, 0]$ as $t \to \infty$. Thus for large ζ

$$v(\varepsilon\zeta) \approx \frac{1}{k}\begin{pmatrix}(2n\pm 1)\pi - \mathrm{sgn}[\mathcal{A}(k,\Delta P_{x0})]\frac{\pi}{2}+\varepsilon a\zeta\\ a/2\end{pmatrix}, \tag{3.102}$$

which is the odd l solution in case (I).

We now determine $d_1^{\min}$, $d_2^{\min}$, and χ_∞. By (3.83)–(3.87) and (3.101)

$$\begin{aligned}|v_1(\zeta,\varepsilon)-\theta_0| &\le \frac{|X(\Omega\varepsilon\zeta;Z_0)-X_0|+\varepsilon|a|\zeta}{k}\\ &\le \frac{2\pi+\varepsilon|a|\zeta}{k}\\ &=: d_1^{\min}(\theta_0,\chi_0,\varepsilon\zeta,k,a), \end{aligned} \tag{3.103}$$

$$\begin{aligned}|v_2(\zeta,\varepsilon)-\chi_0| &= \frac{\Omega}{2k}|Y(\Omega\varepsilon\zeta;Z_0)-Y_0| \le \frac{\Omega(k)}{k}\\ &=: d_2^{\min}(\theta_0,\chi_0,k,a), \end{aligned} \tag{3.104}$$

$$\begin{aligned}|v_2(\zeta,\varepsilon)| &\le \frac{\Omega|Y(\Omega\varepsilon\zeta;Z_0)|+|a|}{2k}\\ &\le \frac{2\Omega(k)+|a|}{2k}\\ &=: \chi_\infty(\theta_0,\chi_0,k,a). \end{aligned} \tag{3.105}$$

(IV) Rotation regime.—$\mathcal{E}_{\rm Pen}(X_0, Y_0) > 2$. For $Y_0 > 0$, X is increasing and Y is periodic such that

$$\sqrt{2}\sqrt{\mathcal{E}_{\rm Pen}(X_0,Y_0)-2} \le Y \le \sqrt{2}\sqrt{\mathcal{E}_{\rm Pen}(X_0,Y_0)}, \tag{3.106}$$

and for $Y_0 < 0$, X is decreasing and Y is periodic such that

$$-\sqrt{2}\sqrt{\mathcal{E}_{\rm Pen}(X_0,Y_0)} \le Y \le -\sqrt{2}\sqrt{\mathcal{E}_{\rm Pen}(X_0,Y_0)-2}. \tag{3.107}$$

Clearly $v_2(\cdot, \varepsilon)$ is periodic. We now determine $d_1^{\min}$, $d_2^{\min}$, and χ_∞. It follows from (3.106) and (3.107) that for any choice of Y_0

$$\begin{aligned}|Y-Y_0| \le \sqrt{2}\sqrt{\mathcal{E}_R(\theta_0,\chi_0,k,a)}\\ -\sqrt{2}\sqrt{\mathcal{E}_R(\theta_0,\chi_0,k,a)-2}, \end{aligned} \tag{3.108}$$

$$|Y| \le \sqrt{2\mathcal{E}_R(\theta_0,\chi_0,k,a)}. \tag{3.109}$$

It follows from (3.84), (3.85), (3.108), and (3.109) that

$$\begin{aligned}|v_2(\zeta,\varepsilon)-\chi_0| &= \frac{\Omega}{2k}|Y(\Omega\varepsilon\zeta;Z_0)-Y_0|\\ &\le \frac{\Omega}{2k}\Big[\sqrt{2}\sqrt{\mathcal{E}_R(\theta_0,\chi_0,k,a)}\\ &\quad -\sqrt{2}\sqrt{\mathcal{E}_R(\theta_0,\chi_0,k,a)-2}\Big]\\ &=: d_2^{\min}(\theta_0,\chi_0,k,a), \end{aligned} \tag{3.110}$$

$$\begin{aligned}|v_2(\zeta,\varepsilon)| &\le \frac{\Omega|Y(\Omega\varepsilon\zeta;Z_0)|+|a|}{2k}\\ &\le \frac{\Omega(k)\sqrt{2\mathcal{E}_R(\theta_0,\chi_0,k,a)}+|a|}{2k}\\ &=: \chi_\infty(\theta_0,\chi_0,k,a). \end{aligned} \tag{3.111}$$

It follows from (3.70) and (3.109) that

$$
\begin{aligned}
&|X(\Omega\varepsilon\zeta; Z_0) - X_0| \\
&= \left| \int_0^{\Omega\varepsilon\zeta} X'(s)ds \right| = \left| \int_0^{\Omega\varepsilon\zeta} Y(s)ds \right| \le \int_0^{\Omega\varepsilon\zeta} |Y(s)| ds \\
&\le \sqrt{2}\int_0^{\Omega\varepsilon\zeta} \sqrt{\mathcal{E}_{\rm Pen}[X(s), Y(s)]}ds \\
&= \sqrt{2}\Omega\varepsilon\zeta\sqrt{\mathcal{E}_{\rm Pen}(X_0, Y_0)} \\
&= \sqrt{2}\Omega\varepsilon\zeta\sqrt{\mathcal{E}_R(\theta_0, \chi_0, k, a)}, \qquad (3.112)
\end{aligned}
$$

whence, by (3.83),

$$
\begin{aligned}
&|v_1(\zeta, \varepsilon) - \theta_0| \\
&\le \frac{|X(\Omega\varepsilon\zeta; Z_0) - X_0| + \varepsilon|a|\zeta}{k} \\
&\le \frac{\sqrt{2}\Omega(k)\varepsilon\zeta\sqrt{\mathcal{E}_R(\theta_0, \chi_0, k, a)} + \varepsilon|a|\zeta}{k} \\
&=: d_1^{\min}(\theta_0, \chi_0, \varepsilon\zeta, k, a). \qquad (3.113)
\end{aligned}
$$

Clearly the simple pendulum system is central to our NearR normal form approximation. Every student who has taken a course in ODEs or classical mechanics has studied the pendulum equation at some level. However, not every reader of this paper may know the general settings of the equation. So, as an aside, we thought some might be interested in knowing how it fits in a broader context. First, the pendulum equation is a special case of the nonlinear oscillator $\ddot{x} + g(x) = 0$, where $g(x) = \sin x$. Second, the nonlinear oscillator is an important subclass of the class of second-order autonomous systems $\dot{x} = f(x, y)$, $\dot{y} = g(x, y)$. The nonlinear oscillator is discussed in many texts, and here we mention [26,29]. Its PPP is easily constructed from the potential plane as mentioned above and in [27]. After the class of linear systems, the class of second-order autonomous systems has the most well developed theory [30]. Here the qualitative behavior is completely captured in the PPP's. What is missing from a PPP is the time it takes to go from one point on an orbit to another, but this is easily determined using a good ODE solver. Note that in Figs. 2–4 we use the ode45 solver of MATLAB.

The limiting behavior of all solutions bounded in forward time is given by the celebrated Poincarè-Bendixson theorem and as a consequence existence of periodic solutions can be inferred and the possibility of chaotic behavior is eliminated. It also follows that a closed orbit in the phase plane corresponds to a periodic solution. In the pendulum case it is easy to show that orbits starting inside the separatrix correspond to periodic solutions, however this fact also follows from the Poincarè-Bendixson theorem.

E. Averaging theorems

Recall that we have gone from our basic Lorentz system, (2.19)–(2.22), to (3.18) and (3.19) with no approximations. We have also derived two related normal forms for $\nu \ge 1/2$ in the NonR (Sec. III C) and NearR (Sec. III D) cases. Here we state theorems which conclude that the solutions of these normal form systems yield good approximations to the solutions of (2.19)–(2.22) in the appropriate ν domains.

Our NonR theorem in Sec. III E 1 will cover the Δ-NonR case, i.e., closed subintervals $[k + \Delta, k + 1 - \Delta]$ of $(k, k + 1)$, where $k = 0, 1, \ldots, 0 < \Delta < 0.5$, and we will obtain error bounds of $O(\varepsilon/\Delta)$ (Here Δ can be small as mentioned in Secs. III B and III C). Our NearR theorem in Sec. III E 2 will cover the case where $\nu = k + \varepsilon a$ which includes the resonant $\nu = k$ case and we will obtain error bounds of $O(\varepsilon)$. In both cases the bounds will be valid on $O(1/\varepsilon)$ ζ intervals with one restriction on ε.

1. Δ-nonresonant case: $\nu \in [k + \Delta, k + 1 - \Delta]$ (quasiperiodic averaging)

The exact ODEs to be analyzed are (3.18) and (3.19) with the initial conditions $\theta(0, \varepsilon) = \theta_0$, $\chi(0, \varepsilon) = \chi_0$. These are well defined on $\mathcal{D}_0(\varepsilon_0)$ and f_1, f_2 are defined by (3.20) and (3.21) where $\widehat{\rm jj}(n; \nu, \Delta P_{x0})$ is given by (3.29) and g_1, g_2 are defined by (3.22) and (3.23) (see also the end of Sec. III C and Appendix B). The normal form system is (3.39) and (3.40) with initial conditions $v_1(0, \varepsilon) = \theta_0$, $v_2(0, \varepsilon) = \chi_0$ and solution (3.41). Note that $v_i(\zeta, \varepsilon) = v_i(\varepsilon\zeta, 1)$.

We are now ready to state the NonR theorem which roughly concludes that $|\theta(\zeta, \varepsilon) - 2\chi_0\varepsilon\zeta - \theta_0| = O(\varepsilon/\Delta)$ and $|\chi(\zeta, \varepsilon) - \chi_0| = O(\varepsilon/\Delta)$ for $0 \le \zeta \le O(1/\varepsilon)$ with ε sufficiently small. To make the statement of the theorem concise, we now set up the theorem in nine steps.

(1) Basic parameters and initial data.—Let $0 < \varepsilon \le \varepsilon_0 \le 1$, fix $0 < \Delta < 0.5$ and let $\nu \in [k + \Delta, k + 1 - \Delta]$ where k is a nonnegative integer. Choose $\theta_0, \chi_0 \in \mathbb{R}$.

(2) Guiding solution.—Choose $T > 0$ and define the compact (closed and bounded) subset S of $\mathbb{R}^2$ by

$$
\begin{aligned}
S &:= \{\mathbf{v}(\tau, 1): \tau \in [0, T]\} \\
&= \{(2\chi_0\tau + \theta_0, \chi_0): \tau \in [0, T]\}. \qquad (3.114)
\end{aligned}
$$

Recall that $\mathbf{v}(\zeta, \varepsilon) = \mathbf{v}(\varepsilon\zeta, 1)$.

(3) Rectangle around S (the basic domain for averaging theorem).—Let

$$
\begin{aligned}
W(\theta_0, \chi_0, d_1, d_2) &:= (\theta_0 - d_1, \theta_0 + d_1) \\
&\quad \times (\chi_0 - d_2, \chi_0 + d_2), \qquad (3.115)
\end{aligned}
$$

where d_1, d_2 are chosen such that

$$
2|\chi_0|T < d_1, \qquad 0 < d_2. \qquad (3.116)
$$

Clearly $(\theta_0, \chi_0) \in S \subset W(\theta_0, \chi_0, d_1, d_2)$ and for convenience the open rectangle $W(\theta_0, \chi_0, d_1, d_2)$ is big enough to allow for both signs of χ_0.

We denote the closure of W by $\bar{W}$ thus

$$\bar{W}(\theta_0, \chi_0, d_1, d_2) := [\theta_0 - d_1, \theta_0 + d_1] \times [\chi_0 - d_2, \chi_0 + d_2]. \quad (3.117)$$

(4) Restrictions on ε_0.—We want χ_0 negative to be admissible and so we choose ε_0 so small that

$$\chi_{lb}(\varepsilon_0) < 0. \quad (3.118)$$

We want the closure $\bar{W}$ of W to be such that $\bar{W}(\theta_0, \chi_0, d_1, d_2) \times \mathbb{R} \subset \mathcal{D}_0(\varepsilon_0)$ and so we further restrict ε_0 so that

$$d_2 < \chi_0 - \chi_{lb}(\varepsilon_0). \quad (3.119)$$

These two restrictions can be satisfied since, by (3.13), $\chi_{lb}(\varepsilon_0)$ is monotonically increasing in ε_0 and $\lim_{\varepsilon_0 \to 0+}[\chi_{lb}(\varepsilon_0)] = -\infty$. Thus we have three restrictions on ε_0

$$\varepsilon_0 \le 1 \quad \text{and} \quad \chi_{lb}(\varepsilon_0) < \min(0, \chi_0 - d_2). \quad (3.120)$$

By (3.119) and the remarks at the beginning of Sec. III C, the vector field of the ODEs (3.18) and (3.19) is C^∞ on $W(\theta_0, \chi_0, d_1, d_2) \times \mathbb{R} \subset \mathcal{D}_0(\varepsilon_0)$.

(5) Exact solution in rectangle W.—Since the vector fields in (3.18) and (3.19) are C^∞, solutions in $W(\theta_0, \chi_0, d_1, d_2)$ with initial condition $\theta(0, \varepsilon) = \theta_0$, $\chi(0, \varepsilon) = \chi_0$ exist uniquely in $W(\theta_0, \chi_0, d_1, d_2)$ on a maximum forward interval of existence $[0, \beta(\varepsilon))$. Either $\beta(\varepsilon) = \infty$ or the solution approaches the boundary of W as $\zeta \to \beta(\varepsilon)^-$. See Chapter 1 of [31] for a discussion of existence, uniqueness, and continuation to a maximum forward interval of existence.

For convenience we define $I(\varepsilon, T) := [0, T/\varepsilon] \cap [0, \beta(\varepsilon))$.

(6) Lipschitz constants for f_1, f_2 *on rectangle* W.—Let L_1, L_2 be defined by

$$L_1 := \frac{2}{\bar{q}} \max_{\zeta \in [0, 2\pi]} |q(\zeta)| = 2\left[1 + \frac{2K^2}{\bar{q}}|\Delta P_{x0}| + \frac{K^2}{2\bar{q}}\right], \quad (3.121)$$

$$L_2 := \nu K^2(1 + |\Delta P_{x0}|). \quad (3.122)$$

It follows by (3.20), (3.21), (3.121), and (3.122) and for θ_1, θ_2, χ_1, χ_2, $\zeta \in \mathbb{R}$, that

$$|f_1(\chi_2, \zeta) - f_1(\chi_1, \zeta)| \le \frac{2|q(\zeta)|}{\bar{q}}|\chi_2 - \chi_1| \le L_1|\chi_2 - \chi_1|, \quad (3.123)$$

$$\begin{aligned} &|f_2(\theta_2, \zeta, \nu) - f_2(\theta_1, \zeta, \nu)| \\ &\quad = K^2|\cos\zeta + \Delta P_{x0}||\cos\{\nu[\theta_2 - Q(\zeta)]\} \\ &\qquad - \cos\{\nu[\theta_1 - Q(\zeta)]\}| \\ &\quad \le K^2(1 + |\Delta P_{x0}|)|\nu[\theta_2 - Q(\zeta)] - \nu[\theta_1 - Q(\zeta)]| \\ &\quad = \nu K^2(1 + |\Delta P_{x0}|)|\theta_2 - \theta_1| = L_2|\theta_2 - \theta_1|, \end{aligned} \quad (3.124)$$

where we have also used the fact that $|\cos x - \cos y| \le |x - y|$. Thus L_1, L_2 are Lipschitz constants for f_1, f_2 on $W(\theta_0, \chi_0, d_1, d_2)$ respectively (in fact even on $\mathbb{R}^2$).

(7) Bounds for g_1, g_2 *on rectangle* $\bar{W}$.—Since $g_1(\cdot, \varepsilon, \nu)$, $g_2(\cdot, \varepsilon, \nu)$ are continuous on $\bar{W}(\theta_0, \chi_0, d_1, d_2) \times \mathbb{R}$ it is easy to show they are bounded. However Appendix C gives a very detailed derivation of quite explicit minimal bounds for g_1 and g_2. There we show, for (θ, χ, ζ) in $\bar{W}(\theta_0, \chi_0, d_1, d_2) \times \mathbb{R}$,

$$|g_i(\theta, \chi, \zeta, \varepsilon, \nu)| \le C_i(\chi_0, \varepsilon_0, \nu, d_2), \quad (3.125)$$

where $i = 1, 2$ and where the finite C_1 and C_2 are defined by (C26) and (C29).

(8) Besjes terms.—Let B_1, B_2 be defined by

$$\begin{aligned} B_1(\zeta) &:= \left|\int_0^\zeta \tilde{f}_1[v_2(s, \varepsilon), s]ds\right| = \left|\int_0^\zeta \tilde{f}_1(\chi_0, s)ds\right|, \\ B_2(\zeta) &:= \left|\int_0^\zeta \tilde{f}_2[v_1(s, \varepsilon), s, \nu]ds\right| \\ &= \left|\int_0^\zeta \tilde{f}_2(2\chi_0\varepsilon s + \theta_0, s, \nu)ds\right|, \end{aligned} \quad (3.126)$$

where

$$\begin{aligned} \tilde{f}_1(v_2, s) &:= f_1(v_2, s) - \bar{f}_1(v_2) = 2\left(\frac{q(s)}{\bar{q}} - 1\right)v_2, \\ \tilde{f}_2(v_1, s, \nu) &:= f_2(v_1, s, \nu) - \bar{f}_2(v_1, \nu) = f_2(v_1, s, \nu). \end{aligned} \quad (3.127)$$

In (3.126) we have used (3.41). We will also need $B_{1,\infty}$, $B_{2,\infty}$ defined by

$$B_{i,\infty}(\zeta) := \sup_{s \in [0, \zeta)} B_i(s), \quad (3.128)$$

for $i = 1, 2$.

We refer to B_1, B_2 as "Besjes terms" and their importance will be seen both in the bounds presented in Theorem 1 and in the proof of the theorem where they eliminate the need for a near-identity transformation (for the latter, see [6,9–12]).

With this setup we can now state the NonR approximation theorem.

Theorem 1 (averaging theorem in Δ*-NonR case:* $\nu \in [k + \Delta, k + 1 - \Delta]$, $k = 0, 1, \ldots, 0 < \Delta < 0.5$*).*—With the setup given by items (1)–(8) of the above preamble, we obtain, for $\zeta \in I(\varepsilon, T)$, that

$$|\theta(\zeta,\varepsilon) - 2\chi_0\varepsilon\zeta - \theta_0| = O(\varepsilon/\Delta),$$
$$|\chi(\zeta,\varepsilon) - \chi_0| = O(\varepsilon/\Delta). \tag{3.129}$$

More precisely

$$\begin{aligned}
&|\theta(\zeta,\varepsilon) - 2\chi_0\varepsilon\zeta - \theta_0| \\
&\quad \le \varepsilon\bigg([B_{1,\infty}(T/\varepsilon) + C_1T]\cosh(T\sqrt{L_1L_2}) \\
&\qquad + [B_{2,\infty}(T/\varepsilon) + C_2T]\sqrt{\frac{L_1}{L_2}}\sinh(T\sqrt{L_1L_2})\bigg),
\end{aligned} \tag{3.130}$$

$$\begin{aligned}
&|\chi(\zeta,\varepsilon) - \chi_0| \\
&\quad \le \varepsilon\bigg([B_{1,\infty}(T/\varepsilon) + C_1T]\sqrt{\frac{L_2}{L_1}}\sinh(T\sqrt{L_1L_2}) \\
&\qquad + [B_{2,\infty}(T/\varepsilon) + C_2T]\cosh(T\sqrt{L_1L_2})\bigg).
\end{aligned} \tag{3.131}$$

Moreover,

$$B_{1,\infty}(T/\varepsilon) \le \check{B}_1, \qquad B_{2,\infty}(T/\varepsilon) \le \check{B}_2(T,\Delta), \tag{3.132}$$

where $i = 1, 2$ and the $\check{B}_1$, $\check{B}_2(T,\Delta) \in [0,\infty)$ are finite, ε independent and are defined in terms of our basic parameters and initial conditions by

$$\check{B}_1 := \frac{2K^2|\chi_0|}{\bar{q}}\left(2|\Delta P_{x0}| + \frac{1}{4}\right), \tag{3.133}$$

$$\check{B}_2(T,\Delta) := \frac{1}{\Delta}\check{B}_{21}(T) + \check{B}_{22}(T), \tag{3.134}$$

$$\begin{aligned}
\check{B}_{21}(T) := 2K^2[1 + (k+1)|\chi_0|T][|\hat{\mathrm{jj}}(k;\nu,\Delta P_{x0})| \\
+ |\hat{\mathrm{jj}}(k+1;\nu,\Delta P_{x0})|],
\end{aligned} \tag{3.135}$$

$$\check{B}_{22}(T) := 2K^2[1 + (k+1)|\chi_0|T]\sum_{n\in(\mathbb{Z}\backslash\{k,k+1\})}|\hat{\mathrm{jj}}(n;\nu,\Delta P_{x0})|. \tag{3.136}$$

Furthermore, with possibly another restriction on ε_0, $[\theta(\zeta,\varepsilon), \chi(\zeta,\varepsilon)]$ can be made to stay away from the boundary of the rectangle $W(\theta_0,\chi_0,d_1,d_2)$ for $\zeta \in I(\varepsilon,T)$. Thus the ODE continuation theorem (see Sec. 1.2 in [31]) gives $\beta(\varepsilon) > T/\varepsilon$, and the error bounds hold on $I(\varepsilon,T) = [0,T/\varepsilon]$.

The proof of Theorem 1 is presented in Sec. IV A. Note that the symbol $O(\varepsilon/\Delta)$ conveys that the error contains the factor $1/\Delta$.

2. Near-to-resonant case: $\nu = k + \varepsilon a$ (periodic averaging)

The NearR case was defined in Sec. III B. The exact ODEs to be analyzed in this case were derived in Sec. III D and are given by (3.52) and (3.53) with initial conditions $\theta(0,\varepsilon) = \theta_0$, $\chi(0,\varepsilon) = \chi_0$. These are well defined on $\mathcal{D}_0(\varepsilon_0)$ while f_1^R, f_2^R are defined by (3.51) and (3.54) and g_1^R, g_2^R are defined by (3.55) and (3.56). The normal form system (3.61) and (3.62) with initial conditions $v_1(0,\varepsilon) = \theta_0$, $v_2(0,\varepsilon) = \chi_0$ is solved by (3.75) and (3.76) where X, Y satisfy the standard pendulum equations (3.70) with the initial conditions (3.77).

We are now ready to state the NonR theorem which roughly concludes that $|\theta(\zeta,\varepsilon) - v_1(\zeta,\varepsilon)| = O(\varepsilon)$ and $|\chi(\zeta,\varepsilon) - v_2(\zeta,\varepsilon)| = O(\varepsilon)$ for $0 \le \zeta \le O(1/\varepsilon)$ with ε sufficiently small. The setup for the theorem is as follows.

(1) Basic parameters and initial data.—Let $0 < \varepsilon \le \varepsilon_0 \le 1$, $a \in [-1/2, 1/2]$ and k be a positive integer. Choose $\theta_0, \chi_0 \in \mathbb{R}$.

(2) Guiding solution.—Choose $T > 0$ and define the compact subset $S_R := \{\mathbf{v}(\tau,1)\colon \tau \in [0,T]\}$ of $\mathbb{R}^2$ where $\mathbf{v} = (v_1, v_2)$ with v_1, v_2 given by (3.75) and (3.76). Recall that $\mathbf{v}(\zeta,\varepsilon) = \mathbf{v}(\varepsilon\zeta,1)$.

(3) Rectangle around S_R: the basic domain for averaging theorem.—Define an open rectangle $W_R(\theta_0,\chi_0,d_1,d_2)$ around S_R by

$$\begin{aligned}
W_R(\theta_0,\chi_0,d_1,d_2) := (\theta_0 - d_1, \theta_0 + d_1) \\
\times (\chi_0 - d_2, \chi_0 + d_2),
\end{aligned} \tag{3.137}$$

where d_1, d_2 satisfy

$$0 \le d_1^{\min}(\theta_0,\chi_0,T,k,a) < d_1, \tag{3.138}$$

$$0 \le d_2^{\min}(\theta_0,\chi_0,k,a) < d_2, \tag{3.139}$$

with $d_1^{\min}$, $d_2^{\min}$ defined in Sec. III D 3. Clearly $(\theta_0,\chi_0) \in S_R \subset W_R(\theta_0,\chi_0,d_1,d_2)$. Note that, by (3.86), (3.138), and (3.139),

$$\begin{aligned}
|v_1(\tau,1) - \theta_0| &\le d_1^{\min}(\theta_0,\chi_0,\tau,k,a) \\
&\le d_1^{\min}(\theta_0,\chi_0,T,k,a) < d_1, \\
|v_2(\tau,1) - \chi_0| &\le d_2^{min}(\theta_0,\chi_0,k,a) < d_2,
\end{aligned} \tag{3.140}$$

where we also used that $d_1^{\min}(\theta_0,\chi_0,\tau,k,a)$ is increasing wrt τ.

We denote the closure of W_R by $\bar{W}_R$; thus,

$$\begin{aligned}
\bar{W}_R(\theta_0,\chi_0,d_1,d_2) := [\theta_0 - d_1, \theta_0 + d_1] \\
\times [\chi_0 - d_2, \chi_0 + d_2].
\end{aligned} \tag{3.141}$$

(4) Restrictions on ε_0.—Choose ε_0 so small that $\chi_{lb}(\varepsilon_0) < 0$ and $d_2 < \chi_0 - \chi_{lb}(\varepsilon_0)$, i.e., such that (3.118) and (3.119) hold. This can be done since $\chi_{lb}(\varepsilon_0)$ is monotonically increasing in ε_0 and $\lim_{\varepsilon_0\to 0+}[\chi_{lb}(\varepsilon_0)] = -\infty$. Thus we have three restrictions on ε_0 as in (3.120).

Note that $\bar{W}_R(\theta_0,\chi_0,d_1,d_2) \times \mathbb{R} \subset \mathcal{D}_0(\varepsilon_0)$ and by (3.119) and the remarks after (3.58), the vector field of the ODEs (3.52) and (3.53) is C^∞ on $W_R(\theta_0,\chi_0,d_1,d_2) \times \mathbb{R} \subset \mathcal{D}_0(\varepsilon_0)$.

(5) Exact solution in rectangle W_R.—Since the vector fields in (3.52) and (3.53) are C^∞, solutions in $W_R(\theta_0, \chi_0, d_1, d_2)$ with initial condition $\theta(0,\varepsilon) = \theta_0$, $\chi(0,\varepsilon) = \chi_0$ exist uniquely on a maximum forward interval of existence $[0, \beta(\varepsilon))$. Here d_1, d_2 satisfy (3.119), (3.138), and (3.139). Either $\beta(\varepsilon) = \infty$ or the solution approaches the boundary of W_R as $\zeta \to \beta(\varepsilon)^-$. See Chapter 1 of [31] for a discussion of existence, uniqueness, and continuation to a maximum forward interval of existence.

It is convenient to introduce $I(\varepsilon, T) := [0, T/\varepsilon] \cap [0, \beta(\varepsilon))$.

(6) Lipschitz constants for f_1^R, f_2^R on rectangle W_R.—Let L_1^R, L_2^R be defined by

$$L_1^R := L_1 = 2\left[1 + \frac{2K^2}{\bar{q}}|\Delta P_{x0}| + \frac{K^2}{2\bar{q}}\right], \tag{3.142}$$

$$L_2^R := K^2 k(1 + |\Delta P_{x0}|), \tag{3.143}$$

where we have also used (3.121) and where d_1, d_2 satisfy (3.119), (3.138), and (3.139). It follows by (3.54), (3.58), (3.123), (3.142), and (3.143) and, for $\theta_1, \theta_2, \chi_1, \chi_2, \zeta \in \mathbb{R}$,

$$\begin{aligned} |f_1^R(\chi_2, \zeta) - f_1^R(\chi_1, \zeta)| &= |f_1(\chi_2, \zeta) - f_1(\chi_1, \zeta)| \\ &\le L_1|\chi_2 - \chi_1| = L_1^R|\chi_2 - \chi_1|, \end{aligned} \tag{3.144}$$

$$\begin{aligned} &|f_2^R(\theta_2, \varepsilon\zeta, \zeta, k, a) - f_2^R(\theta_1, \varepsilon\zeta, \zeta, k, a)| \\ &\quad = K^2|\cos\zeta + \Delta P_{x0}|| \\ &\qquad \times \cos(k[\theta_2 - \zeta - \Upsilon_0 \sin\zeta - \Upsilon_1 \sin 2\zeta] - \varepsilon a\zeta) \\ &\qquad - \cos(k[\theta_1 - \zeta - \Upsilon_0 \sin\zeta - \Upsilon_1 \sin 2\zeta] - \varepsilon a\zeta)| \\ &\quad \le kK^2(1 + |\Delta P_{x0}|)|\theta_2 - \theta_1| = L_2^R|\theta_2 - \theta_1|, \end{aligned} \tag{3.145}$$

where we have also used the fact that $|\cos x - \cos y| \le |x - y|$. Thus, L_1^R, L_2^R are Lipschitz constants for f_1^R, f_2^R on $W_R(\theta_0, \chi_0, d_1, d_2)$ (in fact even on $\mathbb{R}^2$).

(7) Bounds for g_1^R, g_2^R on rectangle $\bar{W}_R$.—Since $g_1^R(\cdot, \varepsilon, k, a)$, $g_2^R(\cdot, \varepsilon, k, a)$ are continuous on $\bar{W}_R(\theta_0, \chi_0, d_1, d_2) \times \mathbb{R}$ they are bounded. However, Appendix E gives a very detailed derivation of quite explicit minimal bounds for g_1^R and g_2^R. There we show that, for

$$\begin{aligned} &(\theta, \chi, \zeta) \in \bar{W}_R(\theta_0, \chi_0, d_1, d_2) \times \mathbb{R}, \\ &|g_1^R(\theta, \chi, \zeta, \varepsilon, k, a)| \le C_1^R(\chi_0, \varepsilon_0, k, d_2), \\ &|g_2^R(\theta, \chi, \zeta, \varepsilon, k, a)| \le C_2^R(\theta_0, \chi_0, \varepsilon_0, k, a, d_1, d_2), \end{aligned} \tag{3.146}$$

where $i = 1, 2$ and where the finite C_1^R and C_2^R are defined by (E5) and (E14).

(8) Besjes terms.—Let B_1^R, B_2^R be defined by

$$\begin{aligned} B_1^R(\zeta) &:= \left|\int_0^\zeta \tilde{f}_1^R[v_2(s,\varepsilon), s]ds\right|, \\ B_2^R(\zeta) &:= \left|\int_0^\zeta \tilde{f}_2^R[v_1(s,\varepsilon), \varepsilon s, s, k, a]ds\right|, \end{aligned} \tag{3.147}$$

where

$$\begin{aligned} \tilde{f}_1^R(\chi, s) &:= f_1^R(\chi, s) - \bar{f}_1^R(\chi), \\ \tilde{f}_2^R(\theta, \varepsilon s, s, k, a) &:= f_2^R(\theta, \varepsilon s, s, k, a) - \bar{f}_2^R(\theta, \varepsilon s, k). \end{aligned} \tag{3.148}$$

We will also need $B_{1,\infty}^R$, $B_{2,\infty}^R$ defined by

$$B_{i,\infty}^R(\zeta) := \sup_{s \in [0,\zeta)} B_i^R(s), \tag{3.149}$$

where $i = 1, 2$.

We refer to B_1^R, B_2^R as "Besjes terms" and their importance will be seen both in the bounds presented in Theorem 2 and in the proof of the theorem where they eliminate the need for a near-identity transformation.

With this setup we can now state the NearR approximation theorem.

Theorem 2 (averaging theorem in NearR case: $\nu = k + \varepsilon a$, $0 < \varepsilon \le \varepsilon_0$, $k \in \mathbb{N}$, $|a| \le 0.5$).—With the setup given by items (1)–(8) of the above preamble we obtain, for $\zeta \in I(\varepsilon, T)$, that

$$\begin{aligned} |\theta(\zeta, \varepsilon) - v_1(\zeta, \varepsilon)| &= O(\varepsilon), \\ |\chi(\zeta, \varepsilon) - v_2(\zeta, \varepsilon)| &= O(\varepsilon). \end{aligned}$$

More precisely

$$\begin{aligned} &|\theta(\zeta) - v_1(\zeta, \varepsilon)| \\ &\quad \le \varepsilon\Bigg([B_{1,\infty}^R(T/\varepsilon) + C_1^R T]\cosh(T\sqrt{L_1^R L_2^R}) \\ &\qquad + [B_{2,\infty}^R(T/\varepsilon) + C_2^R T]\sqrt{\frac{L_1^R}{L_2^R}}\sinh(T\sqrt{L_1^R L_2^R})\Bigg), \end{aligned} \tag{3.150}$$

$$\begin{aligned} &|\chi(\zeta) - v_2(\zeta, \varepsilon)| \\ &\quad \le \varepsilon\Bigg([B_{1,\infty}^R(T/\varepsilon) + C_1^R T]\sqrt{\frac{L_2^R}{L_1^R}}\sinh(T\sqrt{L_1^R L_2^R}) \\ &\qquad + [B_{2,\infty}^R(T/\varepsilon) + C_2^R T]\cosh(T\sqrt{L_1^R L_2^R})\Bigg). \end{aligned} \tag{3.151}$$

Moreover,

$$B_{i,\infty}^R(T/\varepsilon) \le \check{B}_i^R(T), \tag{3.152}$$

where and $\check{B}_i^R(T) \in [0, \infty)$ are independent of ε and defined by

$$\check{B}_1^R(T) := \frac{2K^2}{\bar{q}}\left[2|\Delta P_{x0}| + \frac{1}{4}\right][\chi_\infty(\theta_0, \chi_0, k, a) + K^2 T|\hat{\hat{\text{jj}}}(k; k, \Delta P_{x0})|], \quad (3.153)$$

$$\check{B}_2^R(T) := K^2\{2 + T[|a| + 2k\chi_\infty(\theta_0, \chi_0, k, a)]\} \times \sum_{n\in\mathbb{Z}\setminus\{k\}} \frac{|\hat{\hat{\text{jj}}}(n; k, \Delta P_{x0})|}{|n-k|}. \quad (3.154)$$

Furthermore, with possibly another restriction on ε_0, $[\theta(\zeta,\varepsilon), \chi(\zeta,\varepsilon)]$ can be made to stay away from the boundary of the rectangle $W_R(\theta_0, \chi_0, d_1, d_2)$ for $\zeta \in I(\varepsilon, T)$. Thus the ODE continuation theorem (see Sec. 1.2 of [31]) gives $\beta(\varepsilon) > T/\varepsilon$, and the error bounds hold on $I(\varepsilon, T) = [0, T/\varepsilon]$.

The proof of Theorem 2 is presented in Sec. IV B.

3. Remarks on the averaging theorems

(1) We have now explored the θ, χ dynamics as a function of ν in the Δ-NonR case and $\nu = k + \varepsilon a$ in the NearR case. However asymptotically there are gaps for $\nu \in (k + \varepsilon a, k + \Delta)$ when ε is small. For $\Delta = O(\varepsilon)$ the accuracy of the NonR normal form breaks down because the error is $O(1)$, however we can come close to the NearR neighborhood by letting $\Delta = O(\varepsilon^\beta)$ with β near 1 however the error in the NonR normal form does deteriorate to $O(\varepsilon^{1-\beta})$. It could be interesting to explore the dynamics in these gaps.

(2) Important for the functioning of the FEL is knowledge of the fraction of the bunch that occupies a bucket. From the analysis in Sec. III D 3 this occurs for initial conditions in the libration case, i.e., $0 < \mathcal{E}_{\text{Pen}}(Z_0) < 2$ where Z_0 is given in (3.71). One can thus determine the set of (θ_0, χ_0) for which Z_0 occupies the pendulum buckets. For more details on the pendulum motion and its impact on the low gain theory see Sec. III G.

(3) Mathematically we want to make sure the buckets are covered by our domain $\mathcal{D}_0(\varepsilon_0)$ for physically reasonable χ_0. From (3.76) the range of the v_2 values in the buckets for the NearR normal form is the interval $(-\frac{\Omega}{k} + \frac{a}{2k}, \frac{\Omega}{k} + \frac{a}{2k})$. Now $a \geq -1/2$ so, for every k, the smallest v_2 in a bucket is $-\frac{\Omega}{k} - \frac{1}{4k}$ whence, since $k \geq 1$, the very smallest v_2 in a bucket is $-\Omega - 1/4$. Thus requiring

$$\chi_b := -\Omega - \frac{1}{4} < 0 \quad (3.155)$$

entails that χ_b is smaller than any χ value inside the buckets and smaller than any χ value on the separatrix. It is plausible to restrict the physically interesting χ values to be greater than, say $3\chi_b$. The condition that $(\theta, 3\chi_b, \zeta) \in \mathcal{D}_0(\varepsilon_0)$ entails that the buckets are taken care of by $\mathcal{D}_0(\varepsilon_0)$ and that ε_0 satisfies the constraint $3\chi_b > \chi_{lb}(\varepsilon_0)$. This can be done since $\chi_{lb}(\varepsilon_0)$ is monotonically increasing in ε_0 and $\lim_{\varepsilon_0\to 0+}[\chi_{lb}(\varepsilon_0)] = -\infty$. The following proposition is a simple application of $\mathcal{D}_0(\varepsilon_0)$.

Proposition 1.—Let $0 < \varepsilon \leq \varepsilon_0$, where $0 < \varepsilon_0 \leq 1$ and $\nu \in [1/2, \infty)$. Let also $\Delta\gamma$ be a positive constant and let

$$\varepsilon_0 < \sqrt{\mathcal{E}}\left[\Delta\gamma + \sqrt{1 + K^2\Pi_{x,ub}^2(1)}\right]^{-1}. \quad (3.156)$$

If $\chi \in \mathbb{R}$ satisfies the condition

$$1 < \gamma_c - \Delta\gamma \leq \gamma_c(1 + \varepsilon\chi) \leq \gamma_c + \Delta\gamma, \quad (3.157)$$

then

$$\chi > \chi_{lb}(\varepsilon_0). \quad (3.158)$$

In other words if ε_0 satisfies (3.156) then the γ values in $[\gamma_c - \Delta\gamma, \gamma_c + \Delta\gamma]$ are taken care of by $\mathcal{D}_0(\varepsilon_0)$.

The proposition guarantees, by choosing a sufficiently small ε_0, that the domain $\mathcal{D}_0(\varepsilon_0)$ is large enough to contain the physical relevant values of θ, χ, ζ.

Proof of Proposition 1.—Let $\chi \in \mathbb{R}$ satisfy (3.157). Then, by (1.5), $\chi \in [-\frac{1}{\sqrt{\mathcal{E}}}\Delta\gamma, \frac{1}{\sqrt{\mathcal{E}}}\Delta\gamma]$ whence, by (3.14), (3.13), and (3.156),

$$\begin{aligned}\chi_{lb}(\varepsilon_0) &= -\frac{1}{\varepsilon_0} + \frac{1}{\sqrt{\mathcal{E}}}\sqrt{1 + K^2\Pi_{x,ub}^2(\varepsilon_0)} \\ &\leq -\frac{1}{\varepsilon_0} + \frac{1}{\sqrt{\mathcal{E}}}\sqrt{1 + K^2\Pi_{x,ub}^2(1)} \\ &< -\frac{1}{\sqrt{\mathcal{E}}}\Delta\gamma \leq \chi,\end{aligned}$$

which entails (3.158). □

Note that the condition $1 < \gamma_c - \Delta\gamma$ in (3.157) is not used in the proof of Proposition 1 but serves to guarantee that χ satisfies the physical condition: $\gamma > 1$, i.e., $1 < \gamma_c(1 + \varepsilon\chi)$.

(4) In applications of Theorems 1 and 2, T should be chosen so that $z \in [0, T/\varepsilon k_u]$ is the domain of interest, e.g., so that $T/(\varepsilon k_u)$ is the length of the undulator.

(5) We note there are only four restrictions on the size of ε_0 and thus ε. The first is that we require $\varepsilon_0 \leq 1$. But this is only a matter of convenience and is really no restriction at all since the averaging theorems are only useful for ε small. The second and third restrictions are in item (4) of the preambles to the two theorems. The second one allows us to use negative χ_0. The third restriction gives us $\bar{W}(\theta_0, \chi_0, d_1, d_2) \times \mathbb{R} \subset \mathcal{D}_0(\varepsilon_0)$ and $\bar{W}_R(\theta_0, \chi_0, d_1, d_2) \times \mathbb{R} \subset \mathcal{D}_0(\varepsilon_0)$. The fourth restriction on ε_0 is given at the end of the two theorems in order for the error to be valid on $I(\varepsilon, T) = [0, T/\varepsilon]$. The third and fourth restrictions on ε_0 pose an optimization problem; by changing the size of W, W_R, the size of ε_0 varies as do the Lipschitz constants and the bounds on g_1, g_2, g_1^R, g_2^R.

Nonetheless, the situation is quite good in comparison to the KAM and Nekhoroshev theorems (see, e.g., [8]), where the restrictions on ε_0 are quite severe and it is with great effort that the restrictions on ε_0 have been improved in some applications, e.g., solar system problems.

(6) In many discussions of application of the MoA, researchers often just assert the existence of bounds, for example by using the well-known fact that a continuous function on a compact set is bounded, or bounds are obtained which are crude. Here we wanted to do more. By using, in the proofs of Theorems 1 and 2, a system of differential inequalities instead of the Gronwall inequality we have been able to use two Lipschitz constants in each proof instead of their maximum and in a similar manner can treat the two Besjes' terms independently as well as the components of g and g^R. Furthermore, we believe the Besjes bounds and the bounds on g_1, g_2, g_1^R, g_2^R are nearly optimal.

(7) We here clarify the contributions of $\hat{\text{jj}}$ to the error bounds of Theorems 1 and 2 by finding simple upper bounds for $\check{B}_{21}(T)$, $\check{B}_1^R(T)$, $\check{B}_{22}(T)$ and $\check{B}_2^R(T)$. First of all we note from (3.27) and (3.29) that

$$|\hat{\text{jj}}(n;\nu,\Delta P_{x0})| \leq 1 + |\Delta P_{x0}|, \tag{3.159}$$

where $\nu \geq 1/2$. Clearly (3.159) gives upper bounds for $\check{B}_{21}(T)$, $\check{B}_1^R(T)$ in (3.135) and (3.153). Second, we obtain from the Cauchy-Schwarz inequality that

$$\begin{aligned}\sum_{0\neq n\in\mathbb{Z}} &|\hat{\text{jj}}(n;\nu,\Delta P_{x0})| \\ &= \sum_{0\neq n\in\mathbb{Z}} \frac{1}{|n|}|n||\hat{\text{jj}}(n;\nu,\Delta P_{x0})| \\ &\leq \Big(\sum_{0\neq n\in\mathbb{Z}} n^2|\hat{\text{jj}}(n;\nu,\Delta P_{x0})|^2\Big)^{1/2}\Big(\sum_{0\neq n\in\mathbb{Z}}\frac{1}{n^2}\Big)^{1/2} \\ &= \frac{\pi}{\sqrt{3}}\Big(\sum_{0\neq n\in\mathbb{Z}} n^2|\hat{\text{jj}}(n;\nu,\Delta P_{x0})|^2\Big)^{1/2},\end{aligned} \tag{3.160}$$

where the finiteness of the rhs follows from the fact that the function $\text{jj}(\cdot;\nu,\Delta P_{x0})$ is C^∞. Since $\text{jj}(\cdot;\nu,\Delta P_{x0})$ is also 2π periodic we can apply Parseval's theorem to get

$$\begin{aligned}\frac{1}{2\pi}\int_{[0,2\pi]} d\zeta &\left|\frac{d}{d\zeta}\text{jj}(\zeta;\nu,\Delta P_{x0})\right|^2 \\ &= \sum_{0\neq n\in\mathbb{Z}} n^2|\hat{\text{jj}}(n;\nu,\Delta P_{x0})|^2.\end{aligned} \tag{3.161}$$

It also follows from (3.27) that

$$\begin{aligned}\frac{d}{d\zeta}&\text{jj}(\zeta;\nu,\Delta P_{x0}) \\ &= -\exp(-i\nu[\Upsilon_0\sin\zeta + \Upsilon_1\sin 2\zeta]) \\ &\quad\times\{\sin\zeta + i\nu(\cos\zeta + \Delta P_{x0})[\Upsilon_0\cos\zeta + 2\Upsilon_1\cos 2\zeta]\},\end{aligned}$$

whence

$$\left|\frac{d}{d\zeta}\text{jj}(\zeta;\nu,\Delta P_{x0})\right|^2 \leq 1 + \nu^2(1+|\Delta P_{x0}|)^2[|\Upsilon_0| + 2\Upsilon_1]^2,$$

so that, by (3.160) and (3.161),

$$\begin{aligned}\sum_{0\neq n\in\mathbb{Z}} &|\hat{\text{jj}}(n;\nu,\Delta P_{x0})| \\ &\leq \frac{\pi}{\sqrt{3}}[1 + \nu^2(1+|\Delta P_{x0}|)^2[|\Upsilon_0| + 2\Upsilon_1]^2]^{1/2},\end{aligned} \tag{3.162}$$

which entails, by (3.159),

$$\begin{aligned}\sum_{n\in(\mathbb{Z}\setminus\{k,k+1\})} &|\hat{\text{jj}}(n;\nu,\Delta P_{x0})| \\ &\leq 1 + |\Delta P_{x0}| + \sum_{0\neq n\in\mathbb{Z}}|\hat{\text{jj}}(n;\nu,\Delta P_{x0})| \\ &\leq 1 + |\Delta P_{x0}| + \frac{\pi}{\sqrt{3}}\{1 + \nu^2(1+|\Delta P_{x0}|)^2[|\Upsilon_0| + 2\Upsilon_1]^2\}^{1/2}.\end{aligned} \tag{3.163}$$

Clearly (3.163) gives an upper bound for $\check{B}_{22}(T)$ in (3.136). Moreover, by (3.159) and (3.162),

$$\begin{aligned}\sum_{n\in\mathbb{Z}\setminus\{k\}} &\frac{|\hat{\text{jj}}(n;k,\Delta P_{x0})|}{|n-k|} \\ &\leq |\hat{\text{jj}}(0;k,\Delta P_{x0})| + \sum_{0\neq n\in\mathbb{Z}}|\hat{\text{jj}}(n;k,\Delta P_{x0})| \\ &\leq 1 + |\Delta P_{x0}| + \frac{\pi}{\sqrt{3}}\{1 + \nu^2(1+|\Delta P_{x0}|)^2 \\ &\quad\times[|\Upsilon_0| + 2\Upsilon_1]^2\}^{1/2},\end{aligned}$$

which gives an upper bound for $\check{B}_2^R(T)$ in (3.154).

F. Approximation for the phase space variables in (2.19)–(2.22)

Here we discuss the approximate solutions of (2.19)–(2.22) and (2.26) in terms of the normal form approximations given in (3.41), (3.75), and (3.76), namely,

$$\theta_{\text{NF}}(\tau) := \begin{cases} 2\chi_0\tau + \theta_0 & \text{NonR case} \\ \{X(\Omega\tau;Z_0) - \text{sgn}[\mathcal{A}(k,\Delta P_{x0})]\pi/2 + a\tau\}/k & \text{NearR case,}\end{cases} \tag{3.164}$$

and

$$\chi_{\mathrm{NF}}(\tau) := \begin{cases} \chi_0 & \text{NonR case} \\ [\Omega Y(\Omega\tau; Z_0) + a]/2k & \text{NearR case,} \end{cases} \tag{3.165}$$

where $\mathcal{A}$ is given in (3.63) and Ω in (3.74). Recall from Theorems 1 and 2 that

$$\theta(\zeta, \varepsilon) = \theta_{\mathrm{NF}}(\varepsilon\zeta) + O(\varepsilon), \tag{3.166}$$

$$\chi(\zeta, \varepsilon) = \chi_{\mathrm{NF}}(\varepsilon\zeta) + O(\varepsilon), \tag{3.167}$$

for $\zeta \in I(\varepsilon, T)$. From (1.2), (1.5), (2.23), (2.36), (2.47), and (3.25) it follows that

$$\theta(\zeta, \varepsilon) = \frac{2\mathcal{E}}{\varepsilon^2 \bar{q}}[\zeta - k_u c t(\zeta/k_u)] + Q(\zeta), \tag{3.168}$$

and from (2.34)

$$\gamma(\zeta/k_u) = \gamma_c[1 + \varepsilon\chi(\zeta, \varepsilon)]. \tag{3.169}$$

Now we can determine the approximate solution of (2.19)–(2.22). From (3.166) and (3.168) the arrival time, $t(z)$, of a particle at z is given by

$$t(z) = \frac{z}{c} - \frac{\varepsilon^2\bar{q}}{2\mathcal{E}k_u c}[\theta_{\mathrm{NF}}(\varepsilon k_u z) - Q(k_u z) + O(\varepsilon)]. \tag{3.170}$$

Furthermore from (1.5), (3.167), and (3.169)

$$\gamma(z) = \sqrt{\mathcal{E}}\left(\frac{1}{\varepsilon} + \chi_{\mathrm{NF}}(\varepsilon k_u z) + O(\varepsilon)\right), \tag{3.171}$$

and is clearly slowly varying. From (1.2), (1.4), (1.5), (2.32), (2.39), and (2.40) we have $p_x = mcKP_x$ and

$$p_x(z) = mcK[\cos(k_u z) + \Delta P_{x0} + O(\varepsilon^2)]. \tag{3.172}$$

It is tedious but straightforward to derive from (1.5), (2.33), (2.34), (3.167), and (3.172)

$$p_z(z) = mc\sqrt{\mathcal{E}}\left(\frac{1}{\varepsilon} + \chi_{\mathrm{NF}}(\varepsilon k_u z) + O(\varepsilon)\right). \tag{3.173}$$

Finally we can now determine $x(z)$. From (2.19), (3.172), and (3.173)

$$\begin{aligned} \frac{d}{dz}x(z) = \frac{p_x(z)}{p_z(z)} &= \{mcK[\cos(k_u z) + \Delta P_{x0} + O(\varepsilon^2)]\}\Big/\left[mc\sqrt{\mathcal{E}}\left(\frac{1}{\varepsilon} + \chi_{\mathrm{NF}}(\varepsilon k_u z) + O(\varepsilon)\right)\right] \\ &= \varepsilon\frac{(K/\sqrt{\mathcal{E}})[\cos(k_u z) + \Delta P_{x0} + O(\varepsilon^2)]}{1 + \varepsilon\chi_{\mathrm{NF}}(\varepsilon k_u z) + O(\varepsilon^2)} \\ &= \frac{\varepsilon K}{\sqrt{\mathcal{E}}}[\cos(k_u z) + \Delta P_{x0} + O(\varepsilon^2)][1 - \varepsilon\chi_{\mathrm{NF}}(\varepsilon k_u z) + O(\varepsilon^2)] \\ &= \frac{\varepsilon K}{\sqrt{\mathcal{E}}}[\cos(k_u z) + \Delta P_{x0}][1 - \varepsilon\chi_{\mathrm{NF}}(\varepsilon k_u z)] + O(\varepsilon^3). \end{aligned} \tag{3.174}$$

Integrating (3.174) gives

$$x(z) = x(0) + \frac{\varepsilon K}{\sqrt{\mathcal{E}}}\left(\frac{\sin(k_u z)}{k_u} + z\Delta P_{x0} - \varepsilon\int_0^z [\cos(k_u s) + \Delta P_{x0}]\chi_{\mathrm{NF}}(\varepsilon k_u s)ds\right) + O(\varepsilon^3 z). \tag{3.175}$$

For ε sufficiently small, $I(\varepsilon, T) = [0, T/\varepsilon]$ and then (3.170)–(3.173) and (3.175) hold for $0 \le k_u z \le T/\varepsilon$.

G. Low gain calculation in the NearR regime

Low gain theories in [13,17,18] are done in the context of the pendulum equations, i.e., (3.61) and (3.62) with $a = 0$, $\Delta P_{x0} = 0$, and $k = 1$. Here we will not make those assumptions and we define the gain by

$$G(\zeta, \varepsilon) := \varepsilon\overline{[v_2(\zeta, \varepsilon) - \chi_0]}_{\theta_0} = \varepsilon\overline{[v_2(\varepsilon\zeta, 1) - \chi_0]}_{\theta_0}, \tag{3.176}$$

where v_2 is given in (3.76) and $\overline{(\,)}_{\theta_0}$ denotes the average over θ_0. This is consistent with [13,17,18].

The gain G could be calculated numerically using a quadrature formula and an ODE solver for v_1, v_2, however standard treatments calculate it perturbatively using a regular (and thus short time) perturbation expansion. We could do a regular perturbation expansion in (3.61) and (3.62) by letting $v_i = \sum_{k=0}^4 \varepsilon^k A_{ik} + O(\varepsilon^5)$ and using Gronwall techniques to make the $O(\varepsilon^5)$ error rigorous (see page 594 in [32] for an example of a regular perturbation theorem at first order and its proof). However, at the fourth order needed here this would be quite cumbersome. Because of the special scaling structure in (3.61) and (3.62) as given in (3.66) we can use a Taylor expansion. For $\varepsilon = 1$ we get from (3.61) and (3.62)

$$\begin{aligned} &v_1'(\cdot, 1) = 2v_2(\cdot, 1), \qquad v_1(0, 1) = \theta_0, \\ &v_2'(\cdot, 1) = -\mathcal{A}(k, \Delta P_{x0})\cos[kv_1(\cdot, 1) - a\tau], \\ &v_2(0, 1) = \chi_0, \end{aligned} \tag{3.177}$$

and we expand $v_2(\cdot, 1)$ about $\tau = 0$ so that

$$v_2(\tau,1) = \chi_0 + \sum_{k=1}^{4}\frac{1}{k!}v_2^{(k)}(0,1)\tau^k + \frac{\tau^5}{4!}\int_0^1 (1-t)^4 v_2^{(5)}(t\tau,1)dt. \quad (3.178)$$

From (F6) in Appendix F we have

$$\begin{aligned} v_2'(0,1) &= -\mathcal{A}(k,\Delta P_{x0})\cos(k\theta_0),\\ v_2''(0,1) &= \mathcal{A}(k,\Delta P_{x0})(2k\chi_0 - a)\sin(k\theta_0),\\ v_2'''(0,1) &= \mathcal{A}(k,\Delta P_{x0})(-k\mathcal{A}(k,\Delta P_{x0})\sin(2k\theta_0)\\ &\quad + [2k\chi_0 - a]^2\cos(k\theta_0)),\\ v_2''''(0,1) &= \mathcal{A}(k,\Delta P_{x0})(2k\mathcal{A}(k,\Delta P_{x0})(2k\chi_0 - a)\\ &\quad \times[\sin^2(k\theta_0) - 3\cos^2(k\theta_0)]\\ &\quad - [2k\chi_0 - a]^3\sin(k\theta_0)). \end{aligned} \quad (3.179)$$

It follows from (3.178) and (3.179) that the average over θ_0 leads to

$$\begin{aligned}\overline{[v_2(\tau,1)-\chi_0]}_{\theta_0} &= \frac{\tau^4}{4!}\overline{v_2''''(0,1)}_{\theta_0} + O(\tau^5)\\ &= -\frac{\tau^4}{12}k\mathcal{A}^2(k,\Delta P_{x0})[2k\chi_0 - a] + O(\tau^5),\end{aligned} \quad (3.180)$$

Thus by (3.176),

$$\begin{aligned} G(\zeta,\varepsilon) &= \varepsilon\overline{[v_2(\varepsilon\zeta,1)-\chi_0]}_{\theta_0}\\ &= -\frac{\varepsilon^5\zeta^4}{12}k\mathcal{A}^2(k,\Delta P_{x0})[2k\chi_0 - a] + O(\varepsilon^6). \end{aligned} \quad (3.181)$$

This shows the effect of a and k on the gain.

We now compare our gain formula in (3.181) with the corresponding calculation in [13], where $a = 0$, $\Delta P_{x0} = 0$, and $k = 1$. From our NearR normal form system (3.61) and (3.62) and letting $\theta = v_1$ and $\eta = \varepsilon v_2$ we obtain the IVP

$$\theta' = 2\eta, \qquad \theta(0) = \theta_0, \quad (3.182)$$

$$\eta' = -\epsilon\cos\theta, \qquad \eta(0) = \varepsilon\chi_0 =: \eta_0, \quad (3.183)$$

where $\epsilon = \varepsilon^2\mathcal{A}(1,\Delta P_{x0})$. The procedure in [13] is a regular perturbation expansion in ϵ that does not assume that η_0 is small. Proceeding as they do, we write

$$\theta(\zeta,\epsilon) = \theta^0(\zeta) + \epsilon\theta^1(\zeta) + \epsilon^2\theta^2(\zeta) + O(\epsilon^3), \quad (3.184)$$

$$\eta(\zeta,\epsilon) = \eta^0(\zeta) + \epsilon\eta^1(\zeta) + \epsilon^2\eta^2(\zeta) + O(\epsilon^3). \quad (3.185)$$

We find

$$\eta^0(\zeta) = \eta_0, \quad (3.186)$$

$$\theta^0(\zeta) = 2\eta_0\zeta + \theta_0, \quad (3.187)$$

$$\eta^1(\zeta) = \frac{1}{2\eta_0}[\sin\theta_0 - \sin(2\eta_0\zeta + \theta_0)], \quad (3.188)$$

$$\theta^1(\zeta) = \frac{1}{\eta_0}\left\{\zeta\sin\theta_0 + \frac{1}{2\eta_0}[\cos(2\eta_0\zeta+\theta_0) - \cos\theta_0]\right\}, \quad (3.189)$$

$$\eta^2(\zeta) = \frac{1}{\eta_0}\int_0^\zeta dt\sin(2\eta_0 t + \theta_0) \times\left\{t\sin\theta_0 + \frac{1}{2\eta_0}[\cos(2\eta_0 t + \theta_0) - \cos\theta_0]\right\}. \quad (3.190)$$

It follows that $\overline{\eta^1(\zeta)}_{\theta_0} = 0$ and

$$\overline{\eta^2(\zeta)}_{\theta_0} = \frac{1}{2\eta_0}\int_0^\zeta\left(t\cos 2\eta_0 t - \frac{1}{2\eta_0}\sin 2\eta_0 t\right)dt. \quad (3.191)$$

We can rewrite (3.191) as

$$\overline{\eta^2(\zeta)}_{\theta_0} = \frac{\zeta^3}{4}\frac{d}{d\tau}\left(\frac{\sin\tau}{\tau}\right)^2, \qquad \tau := \eta_0\zeta, \quad (3.192)$$

and the gain becomes

$$G(\zeta,\varepsilon) = \epsilon^2\overline{\eta^2(\zeta)}_{\theta_0} = \varepsilon^4\mathcal{A}^2(1,\Delta P_{x0})\frac{1}{4}\zeta^3\frac{d}{d\tau}\left(\frac{\sin\tau}{\tau}\right)^2, \quad (3.193)$$

consistent with [13]. For η_0 small, which is required by our averaging approximation [since $\eta_0 = \varepsilon\chi_0$ and $\chi_0 = O(1)$], we obtain from (3.191) that

$$\overline{\eta^2(\zeta)}_{\theta_0} = \frac{1}{2\eta_0}\int_0^\zeta\left[-\frac{4}{3}\eta_0^2 t^3 + O(\eta_0 t)^4\right]dt \approx -\frac{1}{6}\eta_0\zeta^4. \quad (3.194)$$

It follows from (3.193) and (3.194) that

$$G(\zeta,\varepsilon) \approx -\epsilon^2\frac{1}{6}\eta_0\zeta^4 = -\frac{\varepsilon^5\zeta^4}{6}\mathcal{A}^2(1,\Delta P_{x0})\chi_0, \quad (3.195)$$

as in (3.181) with $a = 0$ and $k = 1$.

Thus, we see that (3.181) is consistent with the standard gain formula for $\tau = \eta_0\zeta$ small. The $O(\varepsilon^6)$ error in (3.181) can be made precise by estimating the remainder term in (3.178). However, we cannot justify the gain formula either in (3.181) or in (3.193) in the context of our Lorentz system in (2.19)–(2.22), because our NearR normal form approximation only gives an approximation to $O(\varepsilon)$. Thus, a justification of the gain formulas, based on our Lorentz system, would need to come from elsewhere, e.g., a numerical or higher order perturbation calculation based on (3.1) and (3.2).

IV. PROOF OF AVERAGING THEOREMS

In Sec. IV A we prove the NonR theorem, Theorem 1 of Sec. III E 1, and in Sec. IV B we prove the NearR theorem, Theorem 2 of Sec. III E 2.

A. Proof of Theorem 1 (averaging theorem in Δ-NonR case)

Here we compare solutions of the exact IVP (3.18) and (3.19):

$$\theta' = \varepsilon f_1(\chi, \zeta) + \varepsilon^2 g_1(\theta, \chi, \zeta, \varepsilon, \nu), \qquad \theta(0, \varepsilon) = \theta_0, \tag{4.1}$$

$$\chi' = \varepsilon f_2(\theta, \zeta, \nu) + \varepsilon^2 g_2(\theta, \chi, \zeta, \varepsilon, \nu), \qquad \chi(0, \varepsilon) = \chi_0, \tag{4.2}$$

where

$$f_1(\chi, \zeta) = \frac{2q(\zeta)\chi}{\bar{q}}, \tag{4.3}$$

$$\begin{aligned} f_2(\theta, \zeta, \nu) &= -K^2(\cos\zeta + \Delta P_{x0})\cos\{\nu[\theta - Q(\zeta)]\} \\ &= -\frac{K^2}{2} e^{i\nu\theta} \sum_{n\in\mathbb{Z}} \hat{\mathrm{jj}}(n; \nu, \Delta P_{x0}) e^{i(n-\nu)\zeta} + cc, \end{aligned} \tag{4.4}$$

with the normal form IVP of (3.39) and (3.40):

$$v_1' = \varepsilon \bar{f}_1(v_2), \qquad v_1(0, \varepsilon) = \theta_0, \tag{4.5}$$

$$v_2' = \varepsilon \bar{f}_2(v_1, \nu), \qquad v_2(0, \varepsilon) = \chi_0, \tag{4.6}$$

where

$$\bar{f}_1(v_2) = 2v_2, \qquad \bar{f}_2(v_1, \nu) = 0, \tag{4.7}$$

for $\nu \in [k + \Delta, k + 1 - \Delta]$.

Subtracting and integrating, we obtain from (3.127), (4.1), (4.2), (4.5), and (4.6) that

$$\begin{aligned} \theta(\zeta, \varepsilon) - v_1(\zeta, \varepsilon) &= \varepsilon \int_0^\zeta \{f_1[\chi(s, \varepsilon), s] - f_1[v_2(s, \varepsilon), s] + f_1[v_2(s, \varepsilon), s] - \bar{f}_1[v_2(s, \varepsilon)] + \varepsilon g_1[\theta(s, \varepsilon), \chi(s, \varepsilon), s, \varepsilon, \nu]\} ds \\ &= \varepsilon \int_0^\zeta \{f_1[\chi(s, \varepsilon), s] - f_1[v_2(s, \varepsilon), s] + \tilde{f}_1(\chi_0, s) + \varepsilon g_1[\theta(s, \varepsilon), \chi(s, \varepsilon), s, \varepsilon, \nu]\} ds, \end{aligned} \tag{4.8}$$

and

$$\begin{aligned} \chi(\zeta, \varepsilon) - v_2(\zeta, \varepsilon) &= \varepsilon \int_0^\zeta \{f_2[\theta(s, \varepsilon), s, \nu] - f_2[v_1(s, \varepsilon), s, \nu] + f_2[v_1(s, \varepsilon), s, \nu] + \varepsilon g_2[\theta(s, \varepsilon), \chi(s, \varepsilon), s, \varepsilon, \nu]\} ds \\ &= \varepsilon \int_0^\zeta \{f_2[\theta(s, \varepsilon), s, \nu] - f_2[v_1(s, \varepsilon), s, \nu] + \tilde{f}_2[v_1(s, \varepsilon), s, \nu] + \varepsilon g_2[\theta(s, \varepsilon), \chi(s, \varepsilon), s, \varepsilon, \nu]\} ds, \end{aligned} \tag{4.9}$$

for $\zeta \in I(\varepsilon, T) = [0, T/\varepsilon] \cap [0, \beta(\varepsilon))$. Important for our analysis below is that the points $[\theta(\zeta, \varepsilon), \chi(s, \varepsilon)]$ and $[v_1(s, \varepsilon), v_2(s, \varepsilon)]$ belong to the rectangle $W(\theta_0, \chi_0, d_1, d_2)$ for $\zeta \in I(\varepsilon, T)$. Note that we have added and subtracted $f_1[v_2(s, \varepsilon), s]$ in (4.8) and $f_2[v_1(s, \varepsilon), s, \nu]$ in (4.9), an idea introduced by Besjes [16] (see also [14]).

Taking absolute values, applying the Lipschitz condition on $W(\theta_0, \chi_0, d_1, d_2)$ and defining

$$e_1(s) := |\theta(s, \varepsilon) - v_1(s, \varepsilon)|, \tag{4.10}$$

$$e_2(s) := |\chi(s, \varepsilon) - v_2(s, \varepsilon)|, \tag{4.11}$$

gives, by (3.121), (3.122), (3.125), (3.126), (3.128), (4.8), and (4.9) for $\zeta \in I(\varepsilon, T)$,

$$\begin{aligned} 0 \le e_1(\zeta) &\le \varepsilon\left[L_1 \int_0^\zeta e_2(s) ds + \left|\int_0^\zeta \tilde{f}_1(\chi_0, s) ds\right| + \varepsilon \int_0^\zeta |g_1[\theta(s, \varepsilon), \chi(s, \varepsilon), s, \varepsilon, \nu]|\right] \\ &\le \varepsilon\left[L_1 \int_0^\zeta e_2(s) ds + B_1(\zeta) + TC_1\right] \le \varepsilon\left[L_1 \int_0^\zeta e_2(s) ds + B_{1,\infty}(T/\varepsilon) + TC_1\right] =: R_1(\zeta), \end{aligned} \tag{4.12}$$

$$\begin{aligned} 0 \le e_2(\zeta) &\le \varepsilon\left[L_2 \int_0^\zeta e_1(s) ds + \left|\int_0^\zeta \tilde{f}_2(2\chi_0 \varepsilon s + \theta_0, s, \nu) ds\right| + \varepsilon \int_0^\zeta |g_2[\theta(s, \varepsilon), \chi(s, \varepsilon), s, \varepsilon, \nu]|\right] \\ &\le \varepsilon\left[L_2 \int_0^\zeta e_1(s) ds + B_2(\zeta) + TC_2\right] \le \varepsilon\left[L_2 \int_0^\zeta e_1(s) ds + B_{2,\infty}(T/\varepsilon) + TC_2\right] =: R_2(\zeta), \end{aligned} \tag{4.13}$$

where we also used that $I(\varepsilon,T)\subset[0,T/\varepsilon]$ and where we have introduced the R_i as in the proof of the Gronwall inequality for a single integral inequality (the Gronwall inequality is discussed in many ODE books, see, e.g., page 36 in [31] and pages 310 and 317 in [33]). $\zeta\in I(\varepsilon,T)$.

Recall that L_1, L_2, C_1, C_2, B_1, B_2 are defined in items (6), (7), and (8) of the preamble to the theorem. For convenience we have suppressed the ε dependence of e_1 and e_2.

Before we proceed with the proof, several comments are in order.

1. We refer to the terms $B_1(\zeta)$, $B_2(\zeta)$ in (3.126) as Besjes terms since they were introduced by him in order to prove an averaging theorem without using a near-identity transformation; a simplification. Standard proofs use the near-identity transformation (see, e.g., [6,9,10]).

One may fear that the Besjes terms could grow as large as $O(1/\varepsilon)$ for $\zeta\in[0,T/\varepsilon]$, i.e., that $B_{i,\infty}(T/\varepsilon)=O(1/\varepsilon)$. However this does not happen here since, by (3.132), $\check{B}_1$, $\check{B}_2(T,\Delta)$ are upper bounds for $B_{i,\infty}(T/\varepsilon)$ and are ε independent. Two facts are mainly responsible for this: (a) the fact that for fixed v_1 and v_2 the integrands have zero mean, i.e., the quantities in (3.127) have zero mean in s, and (b) the fact that $v_1(s,\varepsilon)$ and $v_2(s,\varepsilon)$ are slowly varying.

2. We maintain the system form in (4.12) and (4.13). We could add these two inequalities and obtain an error estimate using a Gronwall inequality. That is, let $L_\infty=\max(L_1,L_2)$, $B_\infty=B_{1,\infty}+B_{2,\infty}$, $C_\infty=C_1+C_2$, then adding gives

$$0\le e_\infty(\zeta)\le\varepsilon\left[L_\infty\int_0^\zeta e_\infty(s)ds+B_\infty(T/\varepsilon)+C_\infty T\right],\tag{4.14}$$

where $e_\infty=e_1+e_2$. The Gronwall inequality gives $e_\infty(\zeta)\le\varepsilon[B_\infty(T/\varepsilon)+C_\infty T]\exp(\varepsilon L_\infty\zeta)$. However, our system approach gives better bounds.

3. We have a draft of a general paper on quasiperiodic averaging which uses the Besjes idea and deals with the small divisor problem (see [15]). However, the proof we are presenting here is simple, the small divisor problem is trivial and the error bounds are quite explicit. Thus, we feel it is good to give complete proofs here rather than appealing to a more general theory. Also it serves the pedagogical purpose of showing how an averaging theorem is proved in a simple context; here the context of (3.18), (3.19), (3.52), and (3.53). We have incorporated the Besjes idea in much of our previous averaging work, see [14,32,34–36].

We now proceed with the proof. It follows from (4.12) and (4.13) that

$$\begin{aligned}R_1'&=\varepsilon L_1e_2(\zeta)\le\varepsilon L_1R_2(\zeta),\\ R_1(0)&=\varepsilon[B_{1,\infty}(T/\varepsilon)+C_1T],\end{aligned}\tag{4.15}$$

$$\begin{aligned}R_2'&=\varepsilon L_2e_1(\zeta)\le\varepsilon L_2R_1(\zeta),\\ R_2(0)&=\varepsilon[B_{2,\infty}(T/\varepsilon)+C_2T],\end{aligned}\tag{4.16}$$

whence, by Appendix H for $\zeta\in I(\varepsilon,T)$,

$$R_1(\zeta)\le\varepsilon w_1(\varepsilon\zeta),\qquad R_2(\zeta)\le\varepsilon w_2(\varepsilon\zeta),\tag{4.17}$$

where

$$w_1'=L_1w_2,\qquad w_1(0)=B_{1,\infty}(T/\varepsilon)+C_1T,\tag{4.18}$$

$$w_2'=L_2w_1,\qquad w_2(0)=B_{2,\infty}(T/\varepsilon)+C_2T.\tag{4.19}$$

Note that in Appendix H we use the fact that R_1, R_2 are continuously differentiable.

Solving (4.18) and (4.19) we find

$$\begin{pmatrix}w_1(s)\\ w_2(s)\end{pmatrix}=\begin{pmatrix}\cosh(s\sqrt{L_1L_2}) & \sqrt{\frac{L_1}{L_2}}\sinh(s\sqrt{L_1L_2})\\ \sqrt{\frac{L_2}{L_1}}\sinh(s\sqrt{L_1L_2}) & \cosh(s\sqrt{L_1L_2})\end{pmatrix}\times\begin{pmatrix}B_{1,\infty}(T/\varepsilon)+C_1T\\ B_{2,\infty}(T/\varepsilon)+C_2T\end{pmatrix},\tag{4.20}$$

whence, by (4.12), (4.13), and (4.17),

$$\begin{aligned}e_1(\zeta)&\le\varepsilon w_1(\varepsilon\zeta)\le\varepsilon w_1(T)\\ &=\varepsilon\Big([B_{1,\infty}(T/\varepsilon)+C_1T]\cosh(T\sqrt{L_1L_2})\\ &\quad+[B_{2,\infty}(T/\varepsilon)+C_2T]\sqrt{\frac{L_1}{L_2}}\sinh(T\sqrt{L_1L_2})\Big),\end{aligned}\tag{4.21}$$

$$\begin{aligned}e_2(\zeta)&\le\varepsilon w_2(\varepsilon\zeta)\le\varepsilon w_2(T)\\ &=\varepsilon\Big([B_{1,\infty}(T/\varepsilon)+C_1T]\sqrt{\frac{L_2}{L_1}}\sinh(T\sqrt{L_1L_2})\\ &\quad+[B_{2,\infty}(T/\varepsilon)+C_2T]\cosh(T\sqrt{L_1L_2})\Big),\end{aligned}\tag{4.22}$$

for $\zeta\in I(\varepsilon,T)$, where, at the second inequalities, we have used the fact that w_1 and w_2 are increasing [the latter follows from (4.18)–(4.20)]. We thus have proven (3.130) and (3.131) in Theorem 1.

We note that $\check{B}_1$ and $\check{B}_{2,1}(T)$ are finite. Also, since the Fourier series of $\mathrm{jj}(\cdot;\nu,\Delta P_{x0})$ is absolutely convergent, we conclude from (3.136) that $\check{B}_{22}(T)$ is finite whence, by (3.134), $\check{B}_2(T,\Delta)$ is finite.

By restricting ε_0, and thus ε in (4.21) and (4.22), we can keep $[\theta(\zeta,\varepsilon),\chi(\zeta,\varepsilon)]$ away from the boundary of $W(\theta_0,\chi_0,d_1,d_2)$ for $\zeta\in I(\varepsilon,T)$. In this case T/ε must be less than $\beta(\varepsilon)$ thus $I(\varepsilon,T)=[0,T/\varepsilon]$.

To complete the proof we have to show (3.132) which is the heart of the proof. Thus we have to estimate B_1, B_2. From (2.44), (3.41), and (3.127) we obtain

$$\tilde{f}_1[v_2(s,\varepsilon),s] = 2\frac{q(s)-\bar{q}}{\bar{q}}v_2(s,\varepsilon) = \frac{2K^2}{\bar{q}}\Big[2\Delta P_{x0}\cos s + \frac{1}{2}\cos(2s)\Big]\chi_0,$$

and thus, by (3.126) and (3.133),

$$B_1(\zeta) = \frac{2K^2}{\bar{q}}\left|\int_0^\zeta\Big[2\Delta P_{x0}\cos s + \frac{1}{2}\cos(2s)\Big]\chi_0 ds\right| = \frac{2K^2|\chi_0|}{\bar{q}}\left|2\Delta P_{x0}\sin\zeta + \frac{1}{4}\sin(2\zeta)\right| \le \frac{2K^2|\chi_0|}{\bar{q}}\Big(2|\Delta P_{x0}| + \frac{1}{4}\Big) = \check{B}, \tag{4.23}$$

so that, by (3.128), $B_{1,\infty}(T/\varepsilon) \le \check{B}_1$. From (3.41), (3.127), and (4.4) we obtain

$$\tilde{f}_2[v_1(s,\varepsilon),s,\nu] = -\frac{K^2}{2}e^{i\nu[2\varepsilon\chi_0 s+\theta_0]}\sum_{n\in\mathbb{Z}}\hat{\text{jj}}(n;\nu,\Delta P_{x0})e^{i(n-\nu)s} + cc,$$

whence, by (3.126) and for $\zeta \in \mathbb{R}$,

$$\begin{aligned}B_2(\zeta) &= \frac{K^2}{2}\left|\int_0^\zeta e^{i\nu[2\varepsilon\chi_0 s+\theta_0]}\sum_{n\in\mathbb{Z}}\hat{\text{jj}}(n;\nu,\Delta P_{x0})e^{i(n-\nu)s}ds + cc\right| \\ &= \frac{K^2}{2}\left|\sum_{n\in\mathbb{Z}}\hat{\text{jj}}(n;\nu,\Delta P_{x0})\int_0^\zeta e^{i\nu[2\varepsilon\chi_0 s+\theta_0]}e^{i(n-\nu)s}ds + cc\right| \le K^2\sum_{n\in\mathbb{Z}}|\hat{\text{jj}}(n;\nu,\Delta P_{x0})|\left|\int_0^\zeta e^{i2\varepsilon\nu\chi_0 s}e^{i(n-\nu)s}ds\right|,\end{aligned} \tag{4.24}$$

where in the second equality we used the fact that the Fourier series of $\text{jj}(\cdot;\nu,\Delta P_{x0})$ is uniformly convergent. Integrating by parts gives, for $0 \le \zeta \le T/\varepsilon$,

$$\left|\int_0^\zeta e^{i2\varepsilon\nu\chi_0 s}e^{i(n-\nu)s}ds\right| = \left|\frac{e^{i(n-\nu+2\varepsilon\nu\chi_0)\zeta} - 1 - i2\varepsilon\nu\chi_0\int_0^\zeta e^{i(n-\nu+2\varepsilon\nu\chi_0)s}ds}{i(n-\nu)}\right| \le \frac{2+2\varepsilon\nu|\chi_0|\zeta}{|n-\nu|} \le \frac{2+2(k+1)|\chi_0|T}{|n-\nu|},$$

whence, by (4.24), for $0 \le \zeta \le T/\varepsilon$,

$$B_2(\zeta) \le 2K^2[1+(k+1)|\chi_0|T]\sum_{n\in\mathbb{Z}}\left|\frac{\hat{\text{jj}}(n;\nu,\Delta P_{x0})}{n-\nu}\right|. \tag{4.25}$$

The $n-\nu$ in the denominator is the so-called small divisor problem in this context. It is easily resolved in this Δ-NonR case. In fact, for ν Δ-NonR, i.e., $k+\Delta \le \nu \le k+1-\Delta$, we have

$$\begin{aligned}\sum_{n\in\mathbb{Z}}\left|\frac{\hat{\text{jj}}(n;\nu,\Delta P_{x0})}{n-\nu}\right| &= \frac{|\hat{\text{jj}}(k;\nu,\Delta P_{x0})|}{|k-\nu|} + \frac{|\hat{\text{jj}}(k+1;\nu,\Delta P_{x0})|}{|k+1-\nu|} + \sum_{n\in(\mathbb{Z}\setminus\{k,k+1\})}\frac{|\hat{\text{jj}}(n;\nu,\Delta P_{x0})|}{|n-\nu|} \\ &\le \frac{|\hat{\text{jj}}(k;\nu,\Delta P_{x0})|}{\Delta} + \frac{|\hat{\text{jj}}(k+1;\nu,\Delta P_{x0})|}{\Delta} + \sum_{n\in(\mathbb{Z}\setminus\{k,k+1\})}|\hat{\text{jj}}(n;\nu,\Delta P_{x0})|,\end{aligned}$$

whence, by (3.134)–(3.136) and (4.25),

$$\begin{aligned}B_2(\zeta) &\le 2K^2\{1+(k+1)|\chi_0|T\}\left\{\frac{|\hat{\text{jj}}(k;\nu,\Delta P_{x0})| + |\hat{\text{jj}}(k+1;\nu,\Delta P_{x0})|}{\Delta} + \sum_{n\in(\mathbb{Z}\setminus\{k,k+1\})}|\hat{\text{jj}}(n;\nu,\Delta P_{x0})|\right\} \\ &= \frac{1}{\Delta}\check{B}_{21}(T) + \check{B}_{22}(T) = \check{B}_2(T,\Delta),\end{aligned} \tag{4.26}$$

so that, by (3.128), $B_{2,\infty}(T/\varepsilon) \le \check{B}_2(T,\Delta)$.

This completes the proof.

B. Proof of Theorem 2 (averaging theorem in NearR case where $\nu = k + \varepsilon a$)

The proof goes analogously to the proof of Theorem 1 in Sec. IVA and so we omit some details.

Thus, we begin by comparing solutions of the exact IVP (3.52) and (3.53),

$$\theta' = \varepsilon f_1^R(\chi,\zeta) + \varepsilon^2 g_1^R(\theta,\chi,\zeta,\varepsilon,k,a), \qquad \theta(0,\varepsilon) = \theta_0, \tag{4.27}$$

$$\chi' = \varepsilon f_2^R(\theta,\varepsilon\zeta,\zeta,k,a) + \varepsilon^2 g_2^R(\theta,\chi,\zeta,\varepsilon,k,a), \qquad \chi(0,\varepsilon) = \chi_0, \tag{4.28}$$

where, by (3.51), (3.54), and (3.58),

$$f_1^R(\chi,\zeta)=\frac{2q(\zeta)\chi}{\bar{q}}, \tag{4.29}$$

$$f_2^R(\theta,\varepsilon\zeta,\zeta,k,a)=-\frac{K^2}{2}\exp(i[k\theta-a\varepsilon\zeta])\sum_{n\in\mathbb{Z}}\hat{\hat{\mathrm{jj}}}(n;k,\Delta P_{x0})e^{i\zeta[n-k]}+cc, \tag{4.30}$$

with the normal form IVP of (3.61) and (3.62),

$$v_1'=\varepsilon\bar{f}_1^R(v_2),\qquad v_1(0,\varepsilon)=\theta_0, \tag{4.31}$$

$$v_2'=\varepsilon\bar{f}_2^R(v_1,\varepsilon\zeta,k),\qquad v_2(0,\varepsilon)=\chi_0, \tag{4.32}$$

where

$$\bar{f}_1^R(v_2)=2v_2, \tag{4.33}$$

$$\bar{f}_2^R(v_1,\varepsilon\zeta,k)=-\frac{K^2}{2}\exp(i[kv_1-a\varepsilon\zeta])\hat{\hat{\mathrm{jj}}}(k;k,\Delta P_{x0})+cc. \tag{4.34}$$

Subtracting and integrating, we obtain from (3.148), (4.27), (4.28), (4.31), and (4.32) that

$$\begin{aligned}\theta(\zeta)-v_1(\zeta,\varepsilon)&=\varepsilon\int_0^\zeta\{f_1^R[\chi(s),s]-f_1^R[v_2(s,\varepsilon),s]+f_1^R[v_2(s,\varepsilon),s]-\bar{f}_1^R[v_2(s,\varepsilon)]+\varepsilon g_1^R[\theta(s),\chi(s),s,\varepsilon,k,a]\}ds\\&=\varepsilon\int_0^\zeta\{f_1^R[\chi(s),s]-f_1^R[v_2(s,\varepsilon),s]+\tilde{f}_1^R[v_2(s,\varepsilon),s]+\varepsilon g_1^R[\theta(s),\chi(s),s,\varepsilon,k,a]\}ds,\end{aligned} \tag{4.35}$$

and

$$\begin{aligned}\chi(\zeta)-v_2(\zeta,\varepsilon)=\varepsilon\int_0^\zeta&\{f_2^R[\theta(s),\varepsilon s,s,k,a]-f_2^R[v_1(s,\varepsilon),\varepsilon s,s,k,a]+f_2^R[v_1(s,\varepsilon),\varepsilon s,s,k,a]\\&-\bar{f}_2^R[v_1(s,\varepsilon),\varepsilon s,k]+\varepsilon g_2^R[\theta(s),\chi(s),s,\varepsilon,k,a]\}ds\\=\varepsilon\int_0^\zeta&\{f_2^R[\theta(s),\varepsilon s,s,k,a]-f_2^R[v_1(s,\varepsilon),\varepsilon s,s,k,a]+\tilde{f}_2^R[v_1(s,\varepsilon),\varepsilon s,s,k,a]\\&+\varepsilon g_2^R[\theta(s),\chi(s),s,\varepsilon,k,a]\}ds,\end{aligned} \tag{4.36}$$

for $\zeta\in I(\varepsilon,T)=[0,T/\varepsilon]\cap[0,\beta(\varepsilon))$. Taking absolute values, applying the Lipschitz condition and defining

$$e_1(s):=|\theta(s)-v_1(s,\varepsilon)|, \tag{4.37}$$

$$e_2(s):=|\chi(s)-v_2(s,\varepsilon)|, \tag{4.38}$$

gives, by (3.144)–(3.147), (3.149), (4.35), and (4.36) for $\zeta\in I(\varepsilon,T)$,

$$\begin{aligned}0\le e_1(\zeta)&\le\varepsilon\bigg[L_1^R\int_0^\zeta e_2(s)ds+\bigg|\int_0^\zeta\tilde{f}_1^R[v_2(s,\varepsilon),s]ds\bigg|+\varepsilon\int_0^\zeta|g_1^R[\theta(s),\chi(s),s,\varepsilon,k,a]|ds\bigg]\\&\le\varepsilon\bigg[L_1^R\int_0^\zeta e_2(s)ds+B_1^R(\zeta)+TC_1^R\bigg]\le\varepsilon\bigg[L_1^R\int_0^\zeta e_2(s)ds+B_{1,\infty}^R(T/\varepsilon)+TC_1^R\bigg],\end{aligned} \tag{4.39}$$

$$\begin{aligned}0\le e_2(\zeta)&\le\varepsilon\bigg[L_2^R\int_0^\zeta e_1(s)ds+\bigg|\int_0^\zeta\tilde{f}_2^R[v_1(s,\varepsilon),\varepsilon s,s,k,a]ds\bigg|+\varepsilon\int_0^\zeta|g_2^R[\theta(s),\chi(s),s,\varepsilon,k,a]|ds\bigg]\\&\le\varepsilon\bigg[L_2^R\int_0^\zeta e_1(s)ds+B_2^R(\zeta)+TC_2^R\bigg]\le\varepsilon\bigg[L_2^R\int_0^\zeta e_1(s)ds+B_{2,\infty}^R(T/\varepsilon)+TC_2^R\bigg],\end{aligned} \tag{4.40}$$

where we also used that $I(\varepsilon,T)\subset[0,T/\varepsilon]$. Recall that L_i^R, C_i^R, B_i^R are defined in items (6), (7), and (8) of the preamble to the theorem.

We are now in the same situation as in the proof of Theorem 1 since replacing L_i, C_i, B_i in (4.12) and (4.13) by L_i^R, C_i^R, B_i^R results in (4.39) and (4.40). Since, as shown in the proof of Theorem 1, (4.12) and (4.13) entail (4.21) and (4.22) we thus conclude here that (4.39) and (4.40) entail:

$$e_1(\zeta) \leq \varepsilon\left([B_{1,\infty}^R(T/\varepsilon) + C_1 T]\cosh\left(T\sqrt{L_1^R L_2^R}\right) + [B_{2,\infty}^R(T/\varepsilon) + C_2 T]\sqrt{\frac{L_1^R}{L_2^R}}\sinh\left(T\sqrt{L_1^R L_2^R}\right)\right), \tag{4.41}$$

$$e_2(\zeta) \leq \varepsilon\left([B_{1,\infty}^R(T/\varepsilon) + C_1 T]\sqrt{\frac{L_2^R}{L_1^R}}\sinh\left(T\sqrt{L_1^R L_2^R}\right) + [B_{2,\infty}^R(T/\varepsilon) + C_2 T]\cosh\left(T\sqrt{L_1^R L_2^R}\right)\right), \tag{4.42}$$

for $\zeta \in I(\varepsilon, T)$. We thus have proven (3.150) and (3.151).

Clearly, by (3.153), $\check{B}_1^R(T)$ is finite. Also, since $\mathrm{jj}(\cdot;\nu,\Delta P_{x0})$ is a C^∞ function, the series on the rhs of (3.154) converges whence $\check{B}_2^R(T)$ is also finite.

By restricting ε_0, and thus ε in (4.41) and (4.42), we can keep $[\theta(\zeta,\varepsilon), \chi(\zeta,\varepsilon)]$ away from the boundary of $W(\theta_0, \chi_0, d_1, d_2)$ for $\zeta \in I(\varepsilon, T)$. In this case T/ε must be less than $\beta(\varepsilon)$ thus $I(\varepsilon, T) = [0, T/\varepsilon]$.

To complete the proof we have to show (3.152). Thus, we have to estimate B_1^R, B_2^R and beginning with B_1^R we conclude from (2.44), (3.148), (4.29), and (4.33) that, for $\zeta \in \mathbb{R}$,

$$\tilde{f}_1^R[v_2(s,\varepsilon), s] = 2\frac{q(s)-\bar{q}}{\bar{q}}v_2(s,\varepsilon) = \frac{2K^2}{\bar{q}}\left[2\Delta P_{x0}\cos s + \frac{1}{2}\cos(2s)\right]v_2(s,\varepsilon),$$

whence, by (3.87), (3.147), (3.153), (4.32), and (4.34) for $0 \leq \zeta \leq T/\varepsilon$,

$$\begin{aligned}
B_1^R(\zeta) &= \frac{2K^2}{\bar{q}}\left|\int_0^\zeta\left[2\Delta P_{x0}\cos s + \frac{1}{2}\cos(2s)\right]v_2(s,\varepsilon)ds\right| \\
&= \frac{2K^2}{\bar{q}}\left|\left[2\Delta P_{x0}\sin\zeta + \frac{1}{4}\sin(2\zeta)\right]v_2(\zeta,\varepsilon) - \int_0^\zeta\left[2\Delta P_{x0}\sin s + \frac{1}{4}\sin(2s)\right]\frac{dv_2}{ds}(s,\varepsilon)ds\right| \\
&= \frac{2K^2}{\bar{q}}\left|\left[2\Delta P_{x0}\sin\zeta + \frac{1}{4}\sin(2\zeta)\right]v_2(\zeta,\varepsilon) + \varepsilon K^2\widehat{\mathrm{jj}}(k;k,\Delta P_{x0})\int_0^\zeta\left[2\Delta P_{x0}\sin s + \frac{1}{4}\sin(2s)\right]\cos[kv_1(s,\varepsilon) - \varepsilon as]ds\right| \\
&\leq \frac{2K^2}{\bar{q}}\left(\left[2|\Delta P_{x0}| + \frac{1}{4}\right]|v_2(\zeta,\varepsilon)| + \varepsilon K^2|\widehat{\mathrm{jj}}(k;k,\Delta P_{x0})|\left[2|\Delta P_{x0}| + \frac{1}{4}\right]\zeta\right) \\
&\leq \frac{2K^2}{\bar{q}}\left[2|\Delta P_{x0}| + \frac{1}{4}\right][|v_2(\zeta,\varepsilon)| + K^2\varepsilon\zeta|\widehat{\mathrm{jj}}(k;k,\Delta P_{x0})|] \\
&\leq \frac{2K^2}{\bar{q}}\left[2|\Delta P_{x0}| + \frac{1}{4}\right][\chi_\infty(\theta_0,\chi_0,k,a) + K^2 T|\widehat{\mathrm{jj}}(k;k,\Delta P_{x0})|] = \check{B}_1^R(T),
\end{aligned} \tag{4.43}$$

so that, by (3.149), $B_{1,\infty}^R(T/\varepsilon) \leq \check{B}_1^R(T)$ which proves (3.152) for $i = 1$. The key step here is the integration by parts at the second equality which makes explicit the slowly varying nature of v_2 by pulling out the explicit ε after the third equality.

To prove (3.152) for $i = 2$ we conclude from (3.148), (4.30), and (4.34) that, for $\zeta \in \mathbb{R}$,

$$\tilde{f}_2^R[v_1(s,\varepsilon), \varepsilon s, s, k, a] = -\frac{K^2}{2}e^{i[kv_1(s,\varepsilon)-\varepsilon as]}\sum_{n\in\mathbb{Z}\backslash\{k\}}\widehat{\mathrm{jj}}(n;k,\Delta P_{x0})e^{i(n-k)s} + cc,$$

whence, by (3.147) for $\zeta \in \mathbb{R}$,

$$\begin{aligned}
B_2^R(\zeta) &= \frac{K^2}{2}\left|\int_0^\zeta e^{i[kv_1(s,\varepsilon)-\varepsilon as]}\sum_{n\in\mathbb{Z}\backslash\{k\}}\widehat{\mathrm{jj}}(n;k,\Delta P_{x0})e^{i(n-k)s}ds + cc\right| \\
&\leq K^2\sum_{n\in\mathbb{Z}\backslash\{k\}}|\widehat{\mathrm{jj}}(n;k,\Delta P_{x0})|\left|\int_0^\zeta e^{i[kv_1(s,\varepsilon)-\varepsilon as]}e^{i(n-k)s}ds\right|,
\end{aligned} \tag{4.44}$$

where in the inequality we used the fact that the Fourier series of $\mathrm{jj}(\cdot; k, \Delta P_{x0})$ is uniformly convergent. Integrating by parts gives, by (3.87), (4.31), and (4.33) for $0 \le \zeta \le T/\varepsilon$,

$$\begin{aligned}
&\left|\int_0^\zeta e^{i[kv_1(s,\varepsilon)-\varepsilon as]}e^{i(n-k)s}ds\right| \\
&= \left|\frac{1}{i(n-k)}\left[e^{i[kv_1(\zeta,\varepsilon)-\varepsilon a\zeta]}e^{i(n-k)\zeta} - e^{ik\theta_0} - \int_0^\zeta i\left(k\frac{dv_1}{ds}(s,\varepsilon) - \varepsilon a\right)e^{i[kv_1(s,\varepsilon)-\varepsilon as]}e^{i(n-k)s}ds\right]\right| \\
&\le \frac{1}{|n-k|}\left[2 + \int_0^\zeta\left(k\left|\frac{dv_1}{ds}(s,\varepsilon)\right| + \varepsilon|a|\right)ds\right] \le \frac{1}{|n-k|}\left[2 + \varepsilon\int_0^\zeta[2k|v_2(s,\varepsilon)| + |a|)r]ds\right] \\
&\le \frac{1}{|n-k|}\{2 + \varepsilon\zeta[|a| + 2k\chi_\infty(\theta_0,\chi_0,k,a)]\} \le \frac{1}{|n-k|}\{2 + T[|a| + 2k\chi_\infty(\theta_0,\chi_0,k,a)]\},
\end{aligned}$$

whence, by (3.154) and (4.44) for $0 \le \zeta \le T/\varepsilon$,

$$B_2^R(\zeta) \le K^2\{2 + T[|a| + 2k\chi_\infty(\theta_0,\chi_0,k,a)]\}\sum_{n\in\mathbb{Z}\setminus\{k\}}\frac{|\widehat{\mathrm{jj}}(n;k,\Delta P_{x0})|}{|n-k|} = \check{B}_2^R(T), \tag{4.45}$$

so that, by (3.149), $B_{2,\infty}^R(T/\varepsilon) \le \check{B}_2^R(T)$. This completes the proof.

V. SUMMARY AND FUTURE WORK

We started with the 6D Lorentz equations for a planar undulator in (2.12)–(2.15) with time as the independent variable. In Sec. II B we introduced z as the independent variable and considered the IVP at $z = 0$ with $y_0 = p_{y0} = 0$. Solutions of this system are completely determined by the solutions of our basic 2D system (2.30) and (2.31) for α and γ. This basic 2D system is the starting point for the rest of the paper and the first step is to transform it into a form for first-order averaging; the subject of Sec. II C. We introduce $\zeta = k_u z$ as the new independent variable, and χ as a new dependent variable by $\gamma = \gamma_c(1 + \varepsilon\chi)$. Here we are thinking of electrons as part of an electron bunch with γ_c as a characteristic value of γ and ε as a measure of the energy spread so that χ is an $O(1)$ variable. We thus arrive at the system for $(\theta_{\mathrm{aux}}, \chi)$ given in (2.37) and (2.38) and we are interested, in this FEL application, in an asymptotic analysis for ε and $1/\gamma_c$ small. Expanding the vector field for (2.37) and (2.38) gives (2.45) and (2.46). Here θ_{aux} is not slowly varying and we thus introduce the generalized ponderomotive phase, θ, in (2.47) which leads to the slowly varying form of (2.49) and (2.50). Most importantly, we discover that in order for θ and χ to interact at first order we must have $\varepsilon = O(1/\gamma_c)$ and without loss of generality we take (1.5) as a result of (2.51). Finally we obtain (2.52) and (2.53) which is in a standard form for the MoA. Consequently, this will lead to a pendulum-type behavior which is central to FEL theory.

The MoA can be applied to (2.52) and (2.53) after an appropriate h is defined and the rest of the paper, in Secs. III and IV, focuses on the monochromatic case of (2.11).

Before continuing with the summary we note that in the collective case there is a continuous range of frequencies and so it is natural to ask, "What happens in the noncollective case considered in this paper if there is a continuous range of frequencies?" In this situation h can be modeled as

$$h(\alpha) = \int_{-\infty}^{\infty}\tilde{h}(\xi)\exp(-i\xi\alpha)d\xi. \tag{5.1}$$

For $\tilde{h}(\xi) = [\delta(\xi - \nu) + \delta(\xi + \nu)]/2$, where δ is the delta distribution, (5.1) gives $h(\alpha) = \cos(\nu\alpha)$ as in the monochromatic case of (2.11), and, as we have discussed in Sec. III, there are resonances for integer ν. However, we have found that for continuous $\tilde{h}$ the average of $(\cos\zeta + \Delta P_{x0})h[\theta - Q(\zeta)]$ is zero, i.e.,

$$\lim_{T\to\infty}\left[\frac{1}{T}\int_0^T(\cos\zeta + \Delta P_{x0})h[\theta - Q(\zeta)]d\zeta\right] = 0. \tag{5.2}$$

Thus the averaging normal form for (2.52) and (2.53) is just the NonR normal form of Sec. III C and thus a continuous $\tilde{h}(\xi)$, localized for example near the $\nu = 1$ (monochromatic) resonance, washes out the effect of that resonance in the first-order averaging normal form. This does not mean that there is no resonant behavior near $\nu = 1$ because we have not yet proved that the normal form (in this case the NonR normal form of Sec. III C) gives a good approximation, i.e., it may not be possible to prove an averaging theorem. We are pursuing this. However, even if an averaging theorem can be proven there might still be an effect in second-order averaging.

In Sec. III we begin by determining the $O(\varepsilon^2)$ terms of (2.52) and (2.53), using (2.54) and (2.55), which enter the error bounds. Thus we obtain (3.18)–(3.23) as our basic system for θ, χ. In Sec. III A we define a domain, $\mathcal{D}_0(\varepsilon_0) \subset \mathbb{R}^3$ such that g_1, g_2 are well defined and continuous on $\mathcal{D}_0(\varepsilon_0) \times (0, \varepsilon_0] \times [1/2, \infty)$. Moreover, the vector field in (3.18) and (3.19) is well defined and C^∞ on $\mathcal{D}_0(\varepsilon_0)$. Equations (3.18) and (3.19) are in a standard form for the MoA and for each ν a normal form is obtained by dropping the $O(\varepsilon^2)$ terms and averaging f_1, f_2 over ζ. However, the average of f_2 is not clear from (3.21) and it is convenient to expand it in a Fourier series which is given in (3.30)–(3.32). The average is then easily obtained in (3.34) and leads to the definition of the NonR, Δ-NonR, resonant and NearR cases. The NonR normal form equations are $\theta' = \varepsilon 2\chi$ and $\chi' = 0$ in (3.35) and the resonant normal form equations are given by (3.36). The NonR case is stated precisely in Sec. III C. Instead of focusing on the resonant case of (3.36) we consider in Sec. III D the more general NearR case where we study the dynamics in neighborhoods of the $\nu = k$ resonances. If the neighborhood is too small then the resonant normal form of (3.36) will be dominant thus the natural neighborhood to study with first-order averaging is $O(\varepsilon)$ and this is the content of Sec. III D. Replacing ν by $k + \varepsilon a$, our basic equations (3.18) and (3.19) are rewritten in (3.46) and (3.47). The function f_2 in (3.47) has two ε dependencies one of which contributes to the $O(\varepsilon^2)$ term and we are led to the basic NearR system (3.52)–(3.56). In Sec. III D 1 we observe that the g_i^R are well defined on $\mathcal{D}_0(\varepsilon_0) \times (0, \varepsilon_0] \times \mathbb{N} \times [-1/2, 1/2]$ and that $g_i^R(\cdot, k, \cdot)$ are continuous for every $k \in \mathbb{N}$. Moreover, the vector field in (3.52) and (3.53) is well defined and C^∞ on $\mathcal{D}_0(\varepsilon_0)$. In Sec. III D 2 the NearR normal form is presented in (3.61) and (3.62) and the solution behavior is illustrated for $a = 0$ and $a = 1/2$ in Figs. 2–4. The solution structure is conveniently illuminated, in terms of the simple pendulum system, in Sec. III D 3. The simple pendulum exhibits four types of behavior and these are exploited to discuss the structure of solutions of (3.61) and (3.62) in these four cases.

In Sec. III E we state the two averaging theorems which relate the Δ-NonR and NearR normal form approximations to the corresponding exact systems. Each theorem has a detailed preamble which sets up a compact statement of the theorem. The theorems establish the main results of the paper, namely, that the normal form solutions give an $O(\varepsilon)$ approximation to the exact solutions on long-time, $O(1/\varepsilon)$, intervals. In the Δ-NonR case, the ν interval can be made larger by making Δ smaller but this is at the expense of increasing the error as discussed in remark (1) of Sec. III E 3. As a result of the theorems we have good normal form approximations for $\nu \in [k + \Delta, k + 1 - \Delta]$ and $\nu \in [k - \varepsilon/2, k + \varepsilon/2]$. However, we point out there may be gaps between these two intervals where neither normal form applies.

The results of the theorems are applied in Sec. III F, where the normal form approximations are used to derive the approximate solutions of the Lorentz equations with z as the independent variable. In Sec. III G we discuss the small gain theory for $\nu = k + \varepsilon a$ based on our NearR normal form and compare it with the standard theory for $k = 1$, $a = 0$. However, we emphasize that we have not justified the low gain theory in the context of our NearR averaging theorem, as we mention at the end of Sec. III G.

Finally the proofs are given in Sec. IV. It can be seen that the proofs themselves are quite simple. The proofs are somewhat novel in that they do not use a near-identity transformation, due to the Besjes approach, and they use a system of differential inequalities in the calculation of the error bounds, rather than a Gronwall-type inequality, which leads to better error bounds. Therefore a solution of the system of differential inequalities is presented and verified in Appendix H. The first theorem, which is stated for the Δ-NonR case, is an example of a quasiperiodic averaging theorem with its concomitant small divisor problem. It is inherently interesting in that the small divisor problem arises in what must be the simplest possible way. We develop the general theory of quasiperiodic averaging in [15]. The second theorem, which is stated for the NearR case, is an example of periodic averaging which has a vast literature, however as mentioned above our approach here is novel.

While the proofs of Theorems 1 and 2 are simple the whole application of the MoA is not. There was considerable work to put the problem into the standard form and considerable effort to calculate the bounds on g_1, g_2 in Appendix C and g_1^R, g_2^R in Appendix E as well as their $\varepsilon = 0$ limits in Appendixes B and D. Since the $\varepsilon = 0$ singularities in the definitions (3.22), (3.23), (3.55), and (3.56) are removable, these functions could be extended to continuous functions on $\mathcal{D}_0(\varepsilon_0) \times [0, \varepsilon_0] \times [1/2, \infty)$ and $\mathcal{D}_0(\varepsilon_0) \times [0, \varepsilon_0] \times \mathbb{N} \times [-1/2, 1/2]$ respectively, but we chose not to do this. However, we note that g_1 and g_2 are rewritten without the singularity in (B6) and (B13) and g_1^R and g_2^R are rewritten without the singularity in Appendix D [see Eqs. (D1) and (D6)].

We now comment on future work. First of all it would be interesting to include the y dynamics using (2.8) as we do, but not assuming the zero initial conditions in y, thus treating the full 3D dynamics.

Second, it would be interesting to study the helical undulator as we have done here for the planar undulator, i.e., via first-order averaging.

Third, the work here sets the stage for a second-order averaging study of the NonR case in (3.18) and (3.19) using (3.43) and (3.44) and the NearR case in (3.52) and (3.53) using (3.59) and (3.60). In both cases we have systems of the form

$$\frac{dU}{dt} = \varepsilon F(U, t) + \varepsilon^2 G(U, t) + O(\varepsilon^3), \qquad (5.3)$$

with approximating normal form given by

$$\frac{dV}{dt} = \varepsilon \bar{F}(V) + \varepsilon^2 \hat{G}(V), \qquad (5.4)$$

where $\bar{F}$ is the t average of F and $\hat{G}$ is a linear combination of the t average of G and terms depending on F (see [32], Sec. 5, p. 610] for a construction of the normal form, i.e., $\hat{G}$, and an associated theorem and proof). Such a study would include a computation of the averages from (3.43), (3.44), (3.59), and (3.60) and then a phase plane analysis of this second-order normal form system including a comparison with our first-order normal form system. In addition averaging theorems could be proven which we anticipate will give an $O(\varepsilon^2)$ error on $[0, T/\varepsilon]$ as in [32]. Furthermore, it would be interesting to see what happens in the NonR case, e.g., is the energy deviation χ still conserved. We note that generically second-order averaging gives a better error estimate but the interval of validity remains the same (see [32] for situations where the time interval can be extended). Finally it would be interesting to know if, in the NearR case, there is a breakdown in the integrability of the NearR normal form due to separatrix splitting [37] with the concomitant chaotic behavior. This is a delicate issue, which cannot be studied with second-order averaging, since (5.4) is a second-order autonomous system and as such it cannot exhibit chaos as pointed out at the end of Sec. III D 3. This work could be a possible future project, however it does not appear to be interesting from the application point of view since collective effects are surely more important than noncollective effects at second order.

Fourth, we are therefore eager to move on to the collective case based in part on our understanding here. As a first step we are studying the consequence of (G1)–(G5). We have not seen this form of the solution of the 1D wave equation in the FEL literature although the first equality in (G3) is derived in many elementary partial differential equation books. In addition, we are pursuing the issue raised in the paragraph containing Eq. (5.1), concerning a continuous $\tilde{h}$.

VII. TABLE OF NOTATION

a	(3.45)
B_1, B_2	(3.126)
B_1^R, B_2^R	(3.147)
$\mathcal{D}(\varepsilon, \nu)$, $\mathcal{D}_0(\varepsilon_0)$	(3.10) and (3.12)
$\mathcal{E}$	(1.4)
f_1, f_2	(3.20) and (3.21)
f_1^R, f_2^R	(3.54) and (3.51)
g_1, g_2	(3.22) and (3.23)
g_1^R, g_2^R	(3.55) and (3.56)
h, H	(2.11)
jj, $\hat{\text{jj}}$	(3.27) and (3.29)
K	(1.1)
K_r	(1.3)
$\mathcal{A}$	(3.63)
MoA	Method of Averaging
NonR (nonresonant)	Sec. III B
NearR (near-to-resonant)	Sec. III B
$\mathbb{N}$	Set of positive integers
p_x, p_y, p_z (dependent variables)	(2.7)
P_x, P_z	(2.39) and (2.41)
q, $\bar{q}$, Q	(2.44), (3.4), and (2.48)
t (dependent and independent variable)	Sec. II A
v_1, v_2 (dependent variables)	(3.39), (3.40), (3.61), and (3.62)
$\hat{v}_1$, $\hat{v}_2$	(3.67)
W, W_R	(3.115) and (3.137)
x, y (dependent variables)	(2.6)
X, Y	(3.70)
z (dependent and independent variable)	(2.6)
$\mathbb{Z}$	Set of integers
$\check{\alpha}$, α	(2.10) and (2.23)
γ (dependent variable)	(2.4)
γ_c	(2.34)
Δ	III B
Δ − NonR (Δ − NonResonant)	III B
ΔP_{x0}	(2.40)
ε	(1.5)
ζ (independent variables)	(2.35)
η	(2.34)
θ_{aux}, θ (dependent variables)	(2.36) and (2.47)
Π_x, Π_z, $\Pi_{x,ub}$, $\Pi_{z,lb}$	(3.7), (3.8), (3.14), and (C22)
Υ_0, Υ_1	(3.6)
χ (dependent variable)	(2.34)
$\chi_{lb}(\varepsilon)$	(3.13)
Ω	(3.74)

ACKNOWLEDGMENTS

The work of J. A. E. and K. H. was supported by DOE under DE-FG-99ER41104 and the work of M. V. was supported by DESY. Matt Gooden played a significant role in the early stages of this work and was supported by a Teng summer fellowship at ANL and by NSF EMSW21-MCTP Grant No. DMS 0739417, at UNM. Discussions with M. Dohlus, H. S. Dumas, Z. Huang, K.-J. Kim, R. Lindberg, B. F. Roberts, R. Warnock, and I. Zagorodnov are gratefully acknowledged. A special thanks to R. Lindberg for several very helpful comments during the formulation of our approach and a special thanks to Z. Huang and K.-J. Kim for allowing us to sit in on their USPAS FEL course.

APPENDIX A: THE BESSEL EXPANSION

Here we derive the Bessel expansion (3.31) of $\text{jj}(\cdot; \nu, \Delta P_{x0})$. In fact by (3.27),

$$\begin{aligned}\text{jj}(\zeta; \nu, \Delta P_{x0}) &= (\cos\zeta + \Delta P_{x0}) \exp(-i\nu\Upsilon_0 \sin\zeta) \\ &\quad \times \exp(-i\nu\Upsilon_1 \sin 2\zeta) \\ &= \frac{1}{2}\text{jj}_1(\zeta) + \frac{1}{2}\text{jj}_{-1}(\zeta) + \Delta P_{x0}\text{jj}_0(\zeta), \end{aligned} \tag{A1}$$

where

$$\text{jj}_m(\zeta) := \exp(im\zeta)\exp(-i\nu[\Upsilon_0 \sin\zeta + \Upsilon_1 \sin 2\zeta]). \tag{A2}$$

Now

$$\begin{aligned}\exp(ix\sin\theta) &= \sum_{n\in\mathbb{Z}} J_n(x)\exp(in\theta), \\ J_{-n}(x) &= (-1)^n J_n(x),\end{aligned} \tag{A3}$$

whence, by (A2),

$$\mathrm{jj}_m(\zeta) = e^{im\zeta}e^{-i\nu\Upsilon_0\sin\zeta}e^{-i\nu\Upsilon_1\sin2\zeta} = e^{im\zeta}\Big[\sum_{k\in\mathbb{Z}}J_k(\nu\Upsilon_1)e^{-i2k\zeta}\Big]\Big[\sum_{l\in\mathbb{Z}}J_l(\nu\Upsilon_0)e^{-il\zeta}\Big]$$
$$= \sum_{k,l\in\mathbb{Z}}J_l(\nu\Upsilon_0)J_k(\nu\Upsilon_1)e^{i(m-l-2k)\zeta} = \sum_{n\in\mathbb{Z}}\Big(\sum_{k\in\mathbb{Z}}J_{m-n-2k}(\nu\Upsilon_0)J_k(\nu\Upsilon_1)\Big)e^{in\zeta}. \tag{A4}$$

Let

$$\mathcal{J}(n,m,\nu,\Upsilon_0,\Upsilon_1) := \sum_{k\in\mathbb{Z}}J_{m-n-2k}(\nu\Upsilon_0)J_k(\nu\Upsilon_1), \tag{A5}$$

then, by (A4),

$$\mathrm{jj}_m(\zeta) = \sum_{n\in\mathbb{Z}}\mathcal{J}(n,m,\nu,\Upsilon_0,\Upsilon_1)e^{in\zeta}, \tag{A6}$$

and thus, by (A1),

$$\mathrm{jj}(\zeta;\nu,\Delta P_{x0}) = \sum_{n\in\mathbb{Z}}\Big(\frac{1}{2}\mathcal{J}(n,1,\nu,\Upsilon_0,\Upsilon_1) + \frac{1}{2}\mathcal{J}(n,-1,\nu,\Upsilon_0,\Upsilon_1) + \Delta P_{x0}\mathcal{J}(n,0,\nu,\Upsilon_0,\Upsilon_1)\Big)e^{in\zeta}, \tag{A7}$$

whence, by (3.29),

$$\widehat{\mathrm{jj}}(n;\nu,\Delta P_{x0}) = \frac{1}{2}\mathcal{J}(n,1,\nu,\Upsilon_0,\Upsilon_1) + \frac{1}{2}\mathcal{J}(n,-1,\nu,\Upsilon_0,\Upsilon_1) + \Delta P_{x0}\mathcal{J}(n,0,\nu,\Upsilon_0,\Upsilon_1), \tag{A8}$$

so that indeed (3.31) holds.

It is useful for the discussion in Sec. III B to have the following special case. We have, by (A8),

$$\widehat{\mathrm{jj}}(k;k,0) = \frac{1}{2}[\mathcal{J}(k,1,k,0,\Upsilon_1) + \mathcal{J}(k,-1,k,0,\Upsilon_1)], \tag{A9}$$

where

$$\mathcal{J}(k,1,k,0,\Upsilon_1) = \sum_{k'\in\mathbb{Z}}J_{1-k-2k'}(0)J_{k'}(k\Upsilon_1) = \begin{cases} J_{(1-k)/2}(k\Upsilon_1) & \text{if } k \text{ odd} \\ 0 & \text{if } k \text{ even}, \end{cases} \tag{A10}$$

$$\mathcal{J}(k,-1,k,0,\Upsilon_1) = \sum_{k'\in\mathbb{Z}}J_{-1-k-2k'}(0)J_{k'}(k\Upsilon_1) = \begin{cases} J_{-(1+k)/2}(k\Upsilon_1) & \text{if } k \text{ odd} \\ 0 & \text{if } k \text{ even}. \end{cases} \tag{A11}$$

Thus from (A9) $\widehat{\mathrm{jj}}(k;k,0) = 0$ for k even and, for $k = 2n+1$ with $n\in\mathbb{Z}$,

$$\widehat{\mathrm{jj}}(2n+1;2n+1,0) = \frac{1}{2}\{J_{-n}[(2n+1)\Upsilon_1] + J_{-(n+1)}[(2n+1)\Upsilon_1]\} = \frac{1}{2}(-1)^n\{J_n[(2n+1)\Upsilon_1] - J_{n+1}[(2n+1)\Upsilon_1]\}. \tag{A12}$$

APPENDIX B: LIMIT OF g_1, g_2

In this Appendix we first rewrite the functions g_i into the convenient form (B6) and (B13) and use this to compute their limits as $\varepsilon \to 0+$. Furthermore the properties (B6) and (B13) will be used in Appendices C and D. Let therefore $\varepsilon \in (0,\varepsilon_0]$ with $\varepsilon_0 \in (0,1]$, let $\nu \in [1/2,\infty)$ and let $(\theta,\chi,\zeta) \in \mathcal{D}_0(\varepsilon_0)$.

We first consider g_1. Note that, by (2.44), (3.7), and (3.25),

$$1 + K^2\Pi_x^2(\theta,\zeta,\varepsilon,\nu)$$
$$= q(\zeta) + \frac{\varepsilon^2K^2\bar{q}}{2\nu}(\sin\{\nu[\theta - Q(\zeta)]\} - \sin(\nu\theta_0))[2(\cos\zeta + \Delta P_{x0}) + \frac{\varepsilon^2\bar{q}}{2\nu}(\sin\{\nu[\theta - Q(\zeta)]\} - \sin(\nu\theta_0))]. \tag{B1}$$

On the set $\{(\theta,\chi,\zeta,\varepsilon,\nu) \in [\mathcal{D}(\varepsilon,\nu)\times\mathbb{R}^2]: 0 < \varepsilon \le \varepsilon_0, \nu \ge 1/2\}$ we define the real valued function $\tilde{\Pi}_z$ by

$$\tilde{\Pi}_z(\theta,\chi,\zeta,\varepsilon,\nu) := \frac{1}{1+\varepsilon\chi}\Pi_z(\theta,\chi,\zeta,\varepsilon,\nu) = \sqrt{1-\frac{\varepsilon^2}{\mathcal{E}}(1+\varepsilon\chi)^{-2}[1+K^2\Pi_x^2(\theta,\zeta,\varepsilon,\nu)]}. \tag{B2}$$

We obtain from (3.22), (3.25), and (B2) that

$$\varepsilon^2 g_1(\theta,\chi,\zeta,\varepsilon,\nu) = \frac{2\mathcal{E}}{\varepsilon^2\bar{q}}\left(1-\frac{1}{\tilde{\Pi}_z(\theta,\chi,\zeta,\varepsilon,\nu)}\right)+\frac{q(\zeta)}{\bar{q}}(1-2\varepsilon\chi),$$

whence

$$\begin{aligned}\frac{1}{2\mathcal{E}}\bar{q}\tilde{\Pi}_z(\tilde{\Pi}_z+1)\varepsilon^4 g_1 &= \tilde{\Pi}_z^2-1+\frac{1}{2\mathcal{E}}q\tilde{\Pi}_z(\tilde{\Pi}_z+1)\varepsilon^2(1-2\varepsilon\chi)\\ &= \frac{1}{(1+\varepsilon\chi)^2}\left(-\frac{\varepsilon^2}{\mathcal{E}}(q+\varepsilon^2\kappa_1)+\frac{1}{2\mathcal{E}}q\tilde{\Pi}_z(\tilde{\Pi}_z+1)\varepsilon^2(1+\varepsilon\chi)^2(1-2\varepsilon\chi)\right),\end{aligned} \tag{B3}$$

where we used from (3.8), (B1), and (B2) the fact that

$$\tilde{\Pi}_z^2(\theta,\chi,\zeta,\varepsilon,\nu)-1 = -\frac{\varepsilon^2}{\mathcal{E}(1+\varepsilon\chi)^2}[q(\zeta)+\varepsilon^2\kappa_1(\theta,\zeta,\varepsilon,\nu)], \tag{B4}$$

with

$$\kappa_1(\theta,\zeta,\varepsilon,\nu) := \frac{K^2\bar{q}}{2\nu}(\sin\{\nu[\theta-Q(\zeta)]\}-\sin(\nu\theta_0))\left(2(\cos\zeta+\Delta P_{x0})+\frac{\varepsilon^2\bar{q}}{2\nu}(\sin\{\nu[\theta-Q(\zeta)]\}-\sin(\nu\theta_0))\right). \tag{B5}$$

Clearly, by (B3) and (B4),

$$\begin{aligned}\frac{1}{2\mathcal{E}}\bar{q}\tilde{\Pi}_z(\tilde{\Pi}_z+1)\varepsilon^4 g_1 &= -\frac{\varepsilon^2 q}{\mathcal{E}(1+\varepsilon\chi)^2}\left(1-\frac{1}{2}\tilde{\Pi}_z(\tilde{\Pi}_z+1)(1-3\varepsilon^2\chi^2-2\varepsilon^3\chi^3)\right)-\frac{\varepsilon^4\kappa_1}{\mathcal{E}(1+\varepsilon\chi)^2}\\ &= -\frac{\varepsilon^2 q}{\mathcal{E}(1+\varepsilon\chi)^2}\left(-\frac{1}{2}(\tilde{\Pi}_z-1)(\tilde{\Pi}_z+2)+\frac{1}{2}\tilde{\Pi}_z(\tilde{\Pi}_z+1)(3\varepsilon^2\chi^2+2\varepsilon^3\chi^3)\right)-\frac{\varepsilon^4\kappa_1}{\mathcal{E}(1+\varepsilon\chi)^2},\end{aligned}$$

whence

$$\begin{aligned}&\frac{1}{2\mathcal{E}}\bar{q}\tilde{\Pi}_z(\tilde{\Pi}_z+1)^2\varepsilon^4 g_1\\ &= -\frac{\varepsilon^2 q}{2\mathcal{E}(1+\varepsilon\chi)^2}[-(\tilde{\Pi}_z^2-1)(\tilde{\Pi}_z+2)+\varepsilon^2\tilde{\Pi}_z(\tilde{\Pi}_z+1)^2(3\chi^2+2\varepsilon\chi^3)]-\frac{\varepsilon^4\kappa_1}{\mathcal{E}(1+\varepsilon\chi)^2}(\tilde{\Pi}_z+1)\\ &= -\frac{\varepsilon^2 q}{2\mathcal{E}(1+\varepsilon\chi)^4}\left(\frac{\varepsilon^2}{\mathcal{E}}(q+\varepsilon^2\kappa_1)(\tilde{\Pi}_z+2)+\varepsilon^2\tilde{\Pi}_z(\tilde{\Pi}_z+1)^2(3\chi^2+2\varepsilon\chi^3)(1+\varepsilon\chi)^2\right)-\frac{\varepsilon^4\kappa_1}{\mathcal{E}(1+\varepsilon\chi)^2}(\tilde{\Pi}_z+1)\\ &= -\frac{\varepsilon^2 q}{2\mathcal{E}(1+\varepsilon\chi)^4}\left(\frac{\varepsilon^2}{\mathcal{E}}q(\tilde{\Pi}_z+2)+\varepsilon^2\tilde{\Pi}_z(\tilde{\Pi}_z+1)^2(3\chi^2+2\varepsilon\chi^3)(1+\varepsilon\chi)^2\right)\\ &\quad -\frac{\varepsilon^6 q(\tilde{\Pi}_z+2)\kappa_1}{2\mathcal{E}^2(1+\varepsilon\chi)^4}-\frac{\varepsilon^4\kappa_1}{\mathcal{E}(1+\varepsilon\chi)^2}(\tilde{\Pi}_z+1)\\ &= -\frac{\varepsilon^2 q}{2\mathcal{E}(1+\varepsilon\chi)^4}\left(\frac{\varepsilon^2}{\mathcal{E}}q(\tilde{\Pi}_z+2)+\varepsilon^2\tilde{\Pi}_z(\tilde{\Pi}_z+1)^2(3\chi^2+2\varepsilon\chi^3)(1+\varepsilon\chi)^2\right)\\ &\quad -\frac{\varepsilon^4\kappa_1}{2\mathcal{E}(1+\varepsilon\chi)^4}\left(2(1+\varepsilon\chi)^2(\tilde{\Pi}_z+1)+\frac{\varepsilon^2}{\mathcal{E}}q(\tilde{\Pi}_z+2)\right),\end{aligned}$$

so that

$$\begin{aligned}\bar{q}\tilde{\Pi}_z(\tilde{\Pi}_z+1)^2 g_1 &= -\frac{q}{(1+\varepsilon\chi)^4}\left(\frac{q}{\mathcal{E}}(\tilde{\Pi}_z+2)+\tilde{\Pi}_z(\tilde{\Pi}_z+1)^2(3\chi^2+2\varepsilon\chi^3)(1+\varepsilon\chi)^2\right)\\ &\quad -\frac{\kappa_1}{(1+\varepsilon\chi)^4}\left(2(1+\varepsilon\chi)^2(\tilde{\Pi}_z+1)+\frac{\varepsilon^2 q}{\mathcal{E}}(\tilde{\Pi}_z+2)\right),\end{aligned}$$

i.e.,

$$g_1(\theta, \chi, \zeta, \varepsilon, \nu) = -\frac{q}{\bar{q}\tilde{\Pi}_z(\tilde{\Pi}_z + 1)^2(1+\varepsilon\chi)^4}\left(\frac{q}{\mathcal{E}}(\tilde{\Pi}_z + 2) + \tilde{\Pi}_z(\tilde{\Pi}_z + 1)^2(3\chi^2 + 2\varepsilon\chi^3)(1+\varepsilon\chi)^2\right)$$
$$- \frac{\kappa_1}{\bar{q}\tilde{\Pi}_z(\tilde{\Pi}_z + 1)^2(1+\varepsilon\chi)^4}\left(2(1+\varepsilon\chi)^2(\tilde{\Pi}_z + 1) + \frac{\varepsilon^2 q}{\mathcal{E}}(\tilde{\Pi}_z + 2)\right). \tag{B6}$$

Clearly, by (3.8) and (B5),

$$\lim_{\varepsilon\to 0+}[\tilde{\Pi}_z(\theta, \chi, \zeta, \varepsilon, \nu)] = 1, \tag{B7}$$

$$\lim_{\varepsilon\to 0+}[\kappa_1(\chi, \zeta, \varepsilon, \nu)] = \frac{K^2\bar{q}}{\nu}(\sin\{\nu[\theta - Q(\zeta)]\} - \sin(\nu\theta_0))(\cos\zeta + \Delta P_{x0}), \tag{B8}$$

whence, by (B6),

$$\lim_{\varepsilon\to 0+}[g_1(\theta, \chi, \zeta, \varepsilon, \nu)] = -\frac{q(\zeta)}{4\bar{q}}\left(\frac{3}{\mathcal{E}}q(\zeta) + 12\chi^2\right) - \frac{K^2}{\nu}(\sin\{\nu[\theta - Q(\zeta)]\} - \sin(\nu\theta_0))(\cos\zeta + \Delta P_{x0}). \tag{B9}$$

We note that the formula for g_1 in (B6) does not have a singularity at $\varepsilon = 0$ as mentioned at the end of Sec. III A. This could be used as the definition of g_1 including the point $\varepsilon = 0$; we chose not to as the formula is a bit complex.

We now consider g_2 and we obtain from (3.23) and (B2) that

$$\varepsilon^2 g_2(\theta, \chi, \zeta, \varepsilon, \nu) = \varepsilon K^2\cos\{\nu[\theta - Q(\zeta)]\}\left(\cos\zeta + \Delta P_{x0} - \frac{\Pi_x(\theta, \zeta, \varepsilon, \nu)}{\tilde{\Pi}_z(\theta, \chi, \zeta, \varepsilon, \nu)}\right),$$

whence

$$\tilde{\Pi}_z(1+\varepsilon\chi)\varepsilon g_2 = K^2\cos\{\nu[\theta - Q(\zeta)]\}[(1+\varepsilon\chi)\tilde{\Pi}_z(\cos\zeta + \Delta P_{x0}) - \Pi_x]$$
$$= K^2\cos\{\nu[\theta - Q(\zeta)]\}\{(\cos\zeta + \Delta P_{x0})[(1+\varepsilon\chi)\tilde{\Pi}_z - 1] - \varepsilon^2\kappa_2\}, \tag{B10}$$

where we used from (3.7) the fact that

$$\Pi_x(\theta, \zeta, \varepsilon, \nu) = \cos\zeta + \Delta P_{x0} + \varepsilon^2\kappa_2(\theta, \zeta, \nu), \tag{B11}$$

with

$$\kappa_2(\theta, \zeta, \nu) := \frac{\bar{q}}{2\nu}(\sin\{\nu[\theta - Q(\zeta)]\} - \sin(\nu\theta_0)). \tag{B12}$$

Clearly, by (B10),

$$\tilde{\Pi}_z(1+\varepsilon\chi)\varepsilon g_2 = K^2\cos\{\nu[\theta - Q(\zeta)]\}\{(\cos\zeta + \Delta P_{x0})[\tilde{\Pi}_z - 1 + \varepsilon\chi\tilde{\Pi}_z] - \varepsilon^2\kappa_2\},$$

whence, by (B4),

$$(\tilde{\Pi}_z + 1)\tilde{\Pi}_z(1+\varepsilon\chi)\varepsilon g_2$$
$$= K^2\cos\{\nu[\theta - Q(\zeta)]\}\{(\cos\zeta + \Delta P_{x0})[\tilde{\Pi}_z^2 - 1 + \varepsilon\chi\tilde{\Pi}_z(\tilde{\Pi}_z + 1)] - \varepsilon^2\kappa_2(\tilde{\Pi}_z + 1)\}$$
$$= K^2\cos\{\nu[\theta - Q(\zeta)]\}\left((\cos\zeta + \Delta P_{x0})\left[-\frac{\varepsilon^2}{\mathcal{E}(1+\varepsilon\chi)^2}(q + \varepsilon^2\kappa_1) + \varepsilon\chi\tilde{\Pi}_z(\tilde{\Pi}_z + 1)\right] - \varepsilon^2\kappa_2(\tilde{\Pi}_z + 1)\right),$$

so that

$$\tilde{\Pi}_z(\tilde{\Pi}_z + 1)(1+\varepsilon\chi)^3\varepsilon g_2$$
$$= K^2\cos\{\nu[\theta - Q(\zeta)]\}\left((\cos\zeta + \Delta P_{x0})\left[-\frac{\varepsilon^2}{\mathcal{E}}(q + \varepsilon^2\kappa_1) + \varepsilon\chi\tilde{\Pi}_z(\tilde{\Pi}_z + 1)(1+\varepsilon\chi)^2\right] - \varepsilon^2\kappa_2(\tilde{\Pi}_z + 1)(1+\varepsilon\chi)^2\right),$$

which entails that

$$\tilde{\Pi}_z(\tilde{\Pi}_z+1)(1+\varepsilon\chi)^3 g_2 = K^2\cos\{\nu[\theta-Q(\zeta)]\}\Big((\cos\zeta+\Delta P_{x0})\Big[-\frac{\varepsilon}{\mathcal{E}}(q+\varepsilon^2\kappa_1)+\chi\tilde{\Pi}_z(\tilde{\Pi}_z+1)(1+\varepsilon\chi)^2\Big]-\varepsilon\kappa_2(\tilde{\Pi}_z+1)(1+\varepsilon\chi)^2\Big),$$

i.e.,

$$g_2(\theta,\chi,\zeta,\varepsilon,\nu) = \frac{K^2\cos\{\nu[\theta-Q(\zeta)]\}}{\tilde{\Pi}_z(\tilde{\Pi}_z+1)(1+\varepsilon\chi)^3}\Big((\cos\zeta+\Delta P_{x0})\Big[-\frac{\varepsilon}{\mathcal{E}}[q(\zeta)+\varepsilon^2\kappa_1]+\chi\tilde{\Pi}_z(\tilde{\Pi}_z+1)(1+\varepsilon\chi)^2\Big]-\varepsilon\kappa_2(\tilde{\Pi}_z+1)(1+\varepsilon\chi)^2\Big). \tag{B13}$$

Clearly, by (B7) and (B13),

$$\lim_{\varepsilon\to 0+}[g_2(\theta,\chi,\zeta,\varepsilon,\nu)] = \chi K^2\cos\{\nu[\theta-Q(\zeta)]\}(\cos\zeta+\Delta P_{x0}). \tag{B14}$$

We note that the formula for g_2 in (B13) does not have a singularity at $\varepsilon = 0$ as mentioned at the end of Sec. III A. This could be used as the definition of g_2 including the point $\varepsilon = 0$; we chose not to as the formula is a bit complex.

APPENDIX C: BOUNDS ON g_1, g_2

Let $\varepsilon \in (0, \varepsilon_0]$ with $\varepsilon_0 \in (0, 1]$, let $\nu \in [1/2, \infty)$ and let $\chi_0 > \chi_{lb}(\varepsilon_0)$. Let also (3.118) hold, i.e., $\chi_{lb}(\varepsilon_0) < 0$ as in Theorem 1 [see item (4) of the setup list for Theorem 1]. We also assume that

$$(\theta,\chi,\zeta) \in \mathbb{R}\times[\chi_0-d_2,\chi_0+d_2]\times\mathbb{R}, \tag{C1}$$

where

$$0 < d_2 < \chi_0 - \chi_{lb}(\varepsilon_0). \tag{C2}$$

Note that, by (3.12), (C1), and (C2),

$$(\theta,\chi,\zeta) \in (\mathbb{R}\times[\chi_0-d_2,\chi_0+d_2]\times\mathbb{R}) \subset \mathcal{D}_0(\varepsilon_0) \subset \mathcal{D}(\varepsilon,\nu). \tag{C3}$$

In this Appendix we derive the bounds (C26) and (C29) of g_1 and g_2. We thus show in this Appendix that the properties (C26) and (C29) hold in the situation of Theorem 1 [see item (7) of the setup of Theorem 1]. Moreover, the properties (C26) and (C29) will be used in Appendix E. Note that our assumptions in this Appendix allow us to apply the results of Appendix B.

We first consider g_1 and we obtain from (B6)

$$|g_1| = \Big|-\frac{q}{\bar{q}\tilde{\Pi}_z(\tilde{\Pi}_z+1)^2(1+\varepsilon\chi)^4}\Big(\frac{q}{\mathcal{E}}(\tilde{\Pi}_z+2)+\tilde{\Pi}_z(\tilde{\Pi}_z+1)^2(3\chi^2+2\varepsilon\chi^3)(1+\varepsilon\chi)^2\Big) - \frac{\kappa_1}{\bar{q}\tilde{\Pi}_z(\tilde{\Pi}_z+1)^2(1+\varepsilon\chi)^4}\Big(2(1+\varepsilon\chi)^2(\tilde{\Pi}_z+1)+\frac{\varepsilon^2 q}{\mathcal{E}}(\tilde{\Pi}_z+2)\Big)\Big|. \tag{C4}$$

It follows from (2.44), (3.4), (3.11), (B2), and (C3) that

$$q>0,\qquad \bar{q}>0,\qquad 1+\varepsilon\chi>0,\qquad 0<\tilde{\Pi}_z<1,\qquad 3\chi^2+2\varepsilon\chi^3=\chi^2+2\chi^2(1+\varepsilon\chi)\ge 0, \tag{C5}$$

whence, by (C4),

$$|g_1| \leq \frac{q}{\bar{q}\tilde{\Pi}_z(\tilde{\Pi}_z+1)^2(1+\varepsilon\chi)^4}\left(\frac{q}{\mathcal{E}}(\tilde{\Pi}_z+2)+\tilde{\Pi}_z(\tilde{\Pi}_z+1)^2(3\chi^2+2\varepsilon\chi^3)(1+\varepsilon\chi)^2\right)$$
$$+\frac{|\kappa_1|}{\bar{q}\tilde{\Pi}_z(\tilde{\Pi}_z+1)^2(1+\varepsilon\chi)^4}\left(2(1+\varepsilon\chi)^2(\tilde{\Pi}_z+1)+\frac{\varepsilon^2 q}{\mathcal{E}}(\tilde{\Pi}_z+2)\right)$$
$$=\frac{q}{\bar{q}(1+\varepsilon\chi)^2}\left(\frac{q(\tilde{\Pi}_z+2)}{\mathcal{E}\tilde{\Pi}_z(\tilde{\Pi}_z+1)^2(1+\varepsilon\chi)^2}+3\chi^2+2\varepsilon\chi^3\right)+\frac{|\kappa_1|}{\bar{q}\tilde{\Pi}_z(\tilde{\Pi}_z+1)^2(1+\varepsilon\chi)^2}\left(2(\tilde{\Pi}_z+1)+\frac{\varepsilon^2 q(\tilde{\Pi}_z+2)}{\mathcal{E}(1+\varepsilon\chi)^2}\right). \tag{C6}$$

Note also that, by (3.8), (3.15), and (B2),

$$\tilde{\Pi}_z^2(\theta,\chi,\zeta,\varepsilon,\nu)=1-\frac{\varepsilon^2}{\mathcal{E}}\frac{1+K^2\Pi_x^2(\theta,\zeta,\varepsilon,\nu)}{(1+\varepsilon\chi)^2}\geq 1-\frac{\varepsilon^2}{\mathcal{E}}\frac{1+K^2\Pi_{x,ub}^2(\varepsilon)}{(1+\varepsilon\chi)^2}. \tag{C7}$$

Moreover $\varepsilon^2/(1+\varepsilon\chi)^2$ and $1+K^2\Pi_{x,ub}^2(\varepsilon,\nu)$ are increasing wrt ε whence, by (C7),

$$\tilde{\Pi}_z^2(\theta,\chi,\zeta,\varepsilon,\nu)\geq 1-\frac{\varepsilon_0^2}{\mathcal{E}}\frac{1+K^2\Pi_{x,ub}^2(\varepsilon_0)}{(1+\varepsilon_0\chi)^2}. \tag{C8}$$

Since $0<\varepsilon\leq\varepsilon_0$ we have, by (C1),

$$1+\varepsilon\chi\geq 1+\varepsilon(\chi_0-d_2)\geq 1+\inf_{\varepsilon\in(0,\varepsilon_0]}[\varepsilon(\chi_0-d_2)]=1+\min[0,\varepsilon_0(\chi_0-d_2)]=:\kappa_3(\chi_0,\varepsilon_0,d_2). \tag{C9}$$

Note that, by (3.13) and (C2),

$$1+\varepsilon_0(\chi_0-d_2)>1+\varepsilon_0\chi_{lb}(\varepsilon_0)>0, \tag{C10}$$

whence, by (C9),

$$\kappa_3(\chi_0,\varepsilon_0,d_2)>0, \tag{C11}$$

so that, for $n\in\mathbb{N}$ and by (C9),

$$\frac{1}{(1+\varepsilon\chi)^n}\leq\frac{1}{\kappa_3^n(\chi_0,\varepsilon_0,d_2)}. \tag{C12}$$

It follows from (C8) and (C12),

$$\tilde{\Pi}_z^2(\theta,\chi,\zeta,\varepsilon,\nu)\geq\check{\Pi}_{z,lb}(\varepsilon_0), \tag{C13}$$

where

$$\check{\Pi}_{z,lb}(\varepsilon):=1-\varepsilon^2\frac{1+K^2\Pi_{x,ub}^2(\varepsilon)}{\mathcal{E}\kappa_3^2(\chi_0,\varepsilon,d_2)}. \tag{C14}$$

To show that $\check{\Pi}_{z,lb}(\varepsilon_0)>0$ we compute, by using (3.13),

$$\varepsilon_0^2\frac{1+K^2\Pi_{x,ub}^2(\varepsilon_0)}{\mathcal{E}\kappa_3^2(\chi_0,\varepsilon_0,d_2)}=\left(\frac{1+\varepsilon_0\chi_{lb}(\varepsilon_0)}{\kappa_3(\chi_0,\varepsilon_0,d_2)}\right)^2. \tag{C15}$$

If $\chi_0\leq 0$ then, by (C9) and (C10),

$$\kappa_3(\chi_0,\varepsilon_0,d_2)=1+\varepsilon_0(\chi_0-d_2)>1+\varepsilon_0\chi_{lb}(\varepsilon_0)>0, \tag{C16}$$

whence

$$0<\frac{1+\varepsilon_0\chi_{lb}(\varepsilon_0)}{\kappa_3(\chi_0,\varepsilon_0,d_2)}<1, \tag{C17}$$

so that, by (C15),

$$\varepsilon_0^2 \frac{1 + K^2 \Pi_{x,ub}^2(\varepsilon_0)}{\mathcal{E}\kappa_3^2(\chi_0, \varepsilon_0, d_2)} < 1. \tag{C18}$$

If $\chi_0 > 0$ then, by (3.13), (3.118), and (C9),

$$\kappa_3(\chi_0, \varepsilon_0, d_2) = 1 > 1 + \varepsilon_0 \chi_{lb}(\varepsilon_0) > 0, \tag{C19}$$

whence again (C17) holds which entails (C18) by (C15). Having thus proven (C18) we conclude from (C14) that

$$\check{\Pi}_{z,lb}(\varepsilon_0) > 0, \tag{C20}$$

whence, by (C5) and (C13),

$$\tilde{\Pi}_z(\theta, \chi, \zeta, \varepsilon, \nu) > \tilde{\Pi}_{z,lb}(\varepsilon_0), \tag{C21}$$

where

$$\tilde{\Pi}_{z,lb}(\varepsilon) := \sqrt{\check{\Pi}_{z,lb}(\varepsilon)} = \sqrt{1 - \varepsilon^2 \frac{1 + K^2 \Pi_{x,ub}^2(\varepsilon)}{\mathcal{E}\kappa_3^2(\chi_0, \varepsilon, d_2)}}. \tag{C22}$$

Of course since $\tilde{\Pi}_z, \tilde{\Pi}_{z,lb} > 0$ we conclude from (C21) that

$$\frac{1}{\tilde{\Pi}_z(\theta, \chi, \zeta, \varepsilon, \nu)} < \frac{1}{\tilde{\Pi}_{z,lb}(\varepsilon_0)}. \tag{C23}$$

Inserting (C5), (C12), and (C23) into (C6) yields to

$$|g_1| \le \frac{q}{\bar{q}\kappa_3^2(\chi_0, \varepsilon_0, d_2)} \left(\frac{3q}{\mathcal{E}\tilde{\Pi}_{z,lb}(\varepsilon_0)\kappa_3^2(\chi_0, \varepsilon_0, d_2)} + 3\chi^2 + 2\varepsilon_0|\chi|^3 \right) + \frac{|\kappa_1|}{\bar{q}\tilde{\Pi}_{z,lb}(\varepsilon_0)\kappa_3^2(\chi_0, \varepsilon_0, d_2)} \left(4 + \frac{3\varepsilon_0^2 q}{\mathcal{E}\kappa_3^2(\chi_0, \varepsilon_0, d_2)} \right). \tag{C24}$$

Furthermore, by (2.44), (B5), (C1), and (C5),

$$\begin{aligned} |\chi| &= |\chi - \chi_0 + \chi_0| \le |\chi - \chi_0| + |\chi_0| < d_2 + |\chi_0|, \\ |\kappa_1(\theta, \zeta, \varepsilon, \nu)| &\le \frac{K^2 \bar{q}}{\nu} \left(2 + 2|\Delta P_{x0}| + \frac{\varepsilon^2 \bar{q}}{\nu} \right) \le \frac{K^2 \bar{q}}{\nu} \left(2 + 2|\Delta P_{x0}| + \frac{\varepsilon_0^2 \bar{q}}{\nu} \right), \\ q(\zeta) &\le 1 + K^2(1 + |\Delta P_{x0}|)^2 =: q_{ub}. \end{aligned} \tag{C25}$$

Inserting (C25) into (C24) yields to

$$\begin{aligned} |g_1(\theta, \chi, \zeta, \varepsilon, \nu)| &\le \frac{q_{ub}}{\bar{q}\kappa_3^2(\chi_0, \varepsilon_0, d_2)} \left(\frac{3q_{ub}}{\mathcal{E}\tilde{\Pi}_{z,lb}(\varepsilon_0)\kappa_3^2(\chi_0, \varepsilon_0, d_2)} + 3(d_2 + |\chi_0|)^2 + 2\varepsilon_0(d_2 + |\chi_0|)^3 \right) \\ &\quad + \frac{K^2}{\nu\tilde{\Pi}_{z,lb}(\varepsilon_0)\kappa_3^2(\chi_0, \varepsilon_0, d_2)} \left(2 + 2|\Delta P_{x0}| + \frac{\varepsilon_0^2 \bar{q}}{\nu} \right) \left(4 + \frac{3\varepsilon_0^2 q_{ub}}{\mathcal{E}\kappa_3^2(\chi_0, \varepsilon_0, d_2)} \right) \\ &=: C_1(\chi_0, \varepsilon_0, \nu, d_2). \end{aligned} \tag{C26}$$

We now consider g_2 and we obtain from (B13) and (C5)

$$\begin{aligned} |g_2| &\le \frac{K^2}{\tilde{\Pi}_z(\tilde{\Pi}_z + 1)(1 + \varepsilon\chi)^3} \left((1 + |\Delta P_{x0}|) \left[\frac{\varepsilon_0}{\mathcal{E}}(q + \varepsilon_0^2|\kappa_1|) + |\chi|\tilde{\Pi}_z(\tilde{\Pi}_z + 1)(1 + \varepsilon\chi)^2 \right] + \varepsilon_0|\kappa_2|(\tilde{\Pi}_z + 1)(1 + \varepsilon\chi)^2 \right) \\ &= K^2 \left(\frac{\varepsilon_0(1 + |\Delta P_{x0}|)}{\mathcal{E}\tilde{\Pi}_z(\tilde{\Pi}_z + 1)(1 + \varepsilon\chi)^3}(q + \varepsilon_0^2|\kappa_1|) + \frac{|\chi|(1 + |\Delta P_{x0}|)}{1 + \varepsilon\chi} + \frac{\varepsilon_0|\kappa_2|}{\tilde{\Pi}_z(1 + \varepsilon\chi)} \right). \end{aligned} \tag{C27}$$

Note that, by (B12) and (C5),

$$|\kappa_2(\theta, \zeta, \nu)| \le \frac{\bar{q}}{\nu}. \tag{C28}$$

Inserting (C5), (C12), (C23), (C25), and (C28) into (C27) yields to

$$|g_2(\theta,\chi,\zeta,\varepsilon,\nu)| \le K^2\bigg(\frac{\varepsilon_0(1+|\Delta P_{x0}|)}{\mathcal{E}\tilde{\Pi}_{z,lb}(\varepsilon_0)\kappa_3^3(\chi_0,\varepsilon_0,d_2)}\bigg[q_{ub}+\varepsilon_0^2\frac{K^2\bar{q}}{\nu}\bigg(2+2|\Delta P_{x0}|+\frac{\varepsilon_0^2\bar{q}}{\nu}\bigg)\bigg] + \frac{(d_2+|\chi_0|)(1+|\Delta P_{x0}|)}{\kappa_3(\chi_0,\varepsilon_0,d_2)} + \frac{\varepsilon_0\bar{q}}{\nu\tilde{\Pi}_{z,lb}(\varepsilon_0)\kappa_3(\chi_0,\varepsilon_0,d_2)}\bigg) =: C_2(\chi_0,\varepsilon_0,\nu,d_2), \tag{C29}$$

where κ_3, $\tilde{\Pi}_{z,lb}$, q_{ub} are given by (C9), (C22), and (C25).

APPENDIX D: LIMIT OF g_1^R, g_2^R

In this Appendix we first rewrite the functions g_i^R into the convenient form (D1), (D4), (D6), and (D8) and use this to compute their limits as $\varepsilon \to 0+$. Furthermore the properties (D1), (D4), (D6), and (D8) will be used in Appendix E. Let therefore $\varepsilon \in (0,\varepsilon_0]$ with $\varepsilon_0 \in (0,1]$ and $k \in \mathbb{N}$, $a \in [-1/2, 1/2]$ and let $(\theta,\chi,\zeta) \in \mathcal{D}_0(\varepsilon_0)$. Note that our assumptions in this Appendix allow us to apply the results of Appendix B.

We first consider g_1 and we obtain from (3.55) and (B6) that

$$\begin{aligned} g_1^R(\theta,\chi,\zeta,\varepsilon,k,a) &= g_1(\theta,\chi,\zeta,\varepsilon,k+\varepsilon a) \\ &= -\frac{q}{\bar{q}\tilde{\Pi}_z(\tilde{\Pi}_z+1)^2(1+\varepsilon\chi)^4}\bigg(\frac{q}{\mathcal{E}}(\tilde{\Pi}_z+2)+\tilde{\Pi}_z(\tilde{\Pi}_z+1)^2(3\chi^2+2\varepsilon\chi^3)(1+\varepsilon\chi)^2\bigg) \\ &\quad -\frac{\kappa_1}{\bar{q}\tilde{\Pi}_z(\tilde{\Pi}_z+1)^2(1+\varepsilon\chi)^4}\bigg(2(1+\varepsilon\chi)^2(\tilde{\Pi}_z+1)+\frac{\varepsilon^2 q}{\mathcal{E}}(\tilde{\Pi}_z+2)\bigg), \end{aligned} \tag{D1}$$

where $\tilde{\Pi}_z = \tilde{\Pi}_z(\theta,\chi,\zeta,\varepsilon,k+\varepsilon a)$ and $\kappa_1 = \kappa_1(\theta,\zeta,\varepsilon,k+\varepsilon a)$ whence, by (B6) and (B9),

$$\begin{aligned} \lim_{\varepsilon\to 0+}[g_1^R(\theta,\chi,\zeta,\varepsilon,k,a)] &= \lim_{\varepsilon\to 0+}[g_1(\theta,\chi,\zeta,\varepsilon,k)] \\ &= -\frac{q(\zeta)}{4\bar{q}}\bigg(\frac{3}{\mathcal{E}}q(\zeta)+12\chi^2\bigg)-\frac{K^2}{k}(\sin\{k[\theta-Q(\zeta)]\}-\sin(k\theta_0))(\cos\zeta+\Delta P_{x0}). \end{aligned} \tag{D2}$$

We now consider g_2^R and we first use (3.57) and (3.58) to write (3.56) as

$$\begin{aligned} g_2^R(\theta,\chi,\zeta,\varepsilon,k,a) &= g_2(\theta,\chi,\zeta,\varepsilon,k+\varepsilon a) - \frac{K^2}{\varepsilon}(\cos\zeta+\Delta P_{x0})(\cos\{(k+\varepsilon a)[\theta-\zeta-\Upsilon_0\sin\zeta-\Upsilon_1\sin 2\zeta]\} \\ &\quad -\cos(k[\theta-\zeta-\Upsilon_0\sin\zeta-\Upsilon_1\sin 2\zeta]-\varepsilon a\zeta)). \end{aligned} \tag{D3}$$

It follows from (D3) that

$$g_2^R(\theta,\chi,\zeta,\varepsilon,k,a) = g_{2,1}^R(\theta,\chi,\zeta,\varepsilon,k,a) + g_{2,2}^R(\theta,\chi,\zeta,\varepsilon,k,a), \tag{D4}$$

where

$$g_{2,1}^R(\theta,\chi,\zeta,\varepsilon,k,a) := g_2(\theta,\chi,\zeta,\varepsilon,k+\varepsilon a), \tag{D5}$$

$$\begin{aligned} g_{2,2}^R(\theta,\chi,\zeta,\varepsilon,k,a) &:= -\frac{K^2}{\varepsilon}(\cos\zeta+\Delta P_{x0})[\cos(\kappa_4+\kappa_5)-\cos(\kappa_4)] \\ &= \frac{2K^2}{\varepsilon}(\cos\zeta+\Delta P_{x0})\sin(\kappa_5/2)\sin\bigg(\frac{1}{2}[2\kappa_4+\kappa_5]\bigg) \\ &= \frac{K^2}{\varepsilon}(\cos\zeta+\Delta P_{x0})\kappa_5\,\mathrm{sinc}(\kappa_5/2)\sin\bigg(\frac{1}{2}[2\kappa_4+\kappa_5]\bigg) \\ &= K^2 a(\theta-\Upsilon_0\sin\zeta-\Upsilon_1\sin 2\zeta)(\cos\zeta+\Delta P_{x0})\mathrm{sinc}(\kappa_5/2)\sin\bigg(\frac{1}{2}[2\kappa_4+\kappa_5]\bigg), \end{aligned} \tag{D6}$$

with

$$\kappa_4(\theta,\zeta,\varepsilon,k,a) := k(\theta-\zeta-\Upsilon_0\sin\zeta-\Upsilon_1\sin 2\zeta)-\varepsilon a\zeta,\quad \kappa_5(\theta,\zeta,\varepsilon,a) := \varepsilon a(\theta-\Upsilon_0\sin\zeta-\Upsilon_1\sin 2\zeta). \tag{D7}$$

We obtain from (B13) and (D5)

$$
\begin{aligned}
g^R_{2,1}(\theta,\chi,\zeta,\varepsilon,k,a) &= g_2(\theta,\chi,\zeta,\varepsilon,k+\varepsilon a)\\
&= \frac{K^2}{\tilde{\Pi}_z(\tilde{\Pi}_z+1)(1+\varepsilon\chi)^3}\cos\{\nu[\theta-Q(\zeta)]\}\\
&\quad\times\Big((\cos\zeta+\Delta P_{x0})\Big[-\frac{\varepsilon}{\mathcal{E}}(q+\varepsilon^2\kappa_1)+\chi\tilde{\Pi}_z(\tilde{\Pi}_z+1)(1+\varepsilon\chi)^2\Big]-\varepsilon\kappa_2(\tilde{\Pi}_z+1)(1+\varepsilon\chi)^2\Big),
\end{aligned}
\tag{D8}
$$

where $\tilde{\Pi}_z=\tilde{\Pi}_z(\theta,\chi,\zeta,\varepsilon,k+\varepsilon a)$ and $\kappa_2=\kappa_2(\theta,\zeta,k+\varepsilon a)$ whence, by (B13) and (B14),

$$\lim_{\varepsilon\to0+}[g^R_{2,1}(\theta,\chi,\zeta,\varepsilon,k,a)]=\lim_{\varepsilon\to0+}[g_2(\theta,\chi,\zeta,\varepsilon,k)]=\chi K^2\cos\{k[\theta-Q(\zeta)]\}(\cos\zeta+\Delta P_{x0}). \tag{D9}$$

Clearly, by (D7),

$$\lim_{\varepsilon\to0+}\{\mathrm{sinc}[\kappa_5(\theta,\zeta,\varepsilon,a)/2]\}=1,\qquad \lim_{\varepsilon\to0+}\{\sin(\frac{1}{2}[2\kappa_4+\kappa_5])\}=\lim_{\varepsilon\to0+}[\sin(\kappa_4)]=\sin(k[\theta-\zeta-\Upsilon_0\sin\zeta-\Upsilon_1\sin2\zeta]), \tag{D10}$$

whence, by (D6),

$$\lim_{\varepsilon\to0+}[g^R_{2,2}(\theta,\chi,\zeta,\varepsilon,k,a)]=K^2a(\theta-\Upsilon_0\sin\zeta-\Upsilon_1\sin2\zeta)\sin(k[\theta-\zeta-\Upsilon_0\sin\zeta-\Upsilon_1\sin2\zeta])(\cos\zeta+\Delta P_{x0}), \tag{D11}$$

so that, by (D4) and (D9),

$$
\begin{aligned}
\lim_{\varepsilon\to0+}[g^R_2(\theta,\chi,\zeta,\varepsilon,k,a)] &= \chi K^2\cos\{k[\theta-Q(\zeta)]\}(\cos\zeta+\Delta P_{x0})\\
&\quad+K^2a(\theta-\Upsilon_0\sin\zeta-\Upsilon_1\sin2\zeta)\sin(k[\theta-\zeta-\Upsilon_0\sin\zeta-\Upsilon_1\sin2\zeta])(\cos\zeta+\Delta P_{x0}).
\end{aligned}
\tag{D12}
$$

APPENDIX E: BOUNDS ON g^R_1, g^R_2

Let $\varepsilon\in(0,\varepsilon_0]$ with $\varepsilon_0\in(0,1]$ and let $k\in\mathbb{N}$, $a\in[-1/2,1/2]$. Let also $\theta_0\in\mathbb{R}$ and $\chi_0>\chi_{lb}(\varepsilon_0)$. Moreover let (3.118) hold, i.e., $\chi_{lb}(\varepsilon_0)<0$ [see also item (4) of the setup list for Theorem 2]. Furthermore we assume that

$$(\theta,\chi,\zeta)\in[\theta_0-d_1,\theta_0+d_1]\times[\chi_0-d_2,\chi_0+d_2]\times\mathbb{R}, \tag{E1}$$

where χ_0, d_1, d_2 satisfy

$$0<d_1,\qquad 0<d_2<\chi_0-\chi_{lb}(\varepsilon_0). \tag{E2}$$

In this Appendix we derive the bounds (E6) and (E14) of g^R_1 and g^R_2. We thus show in this Appendix that the properties (E6) and (E14) hold in the situation of Theorem 2 [see item (7) of the setup of Theorem 2]. Since all assumptions of this Appendix are also satisfied in Appendix C and D, we can apply the results of those Appendices. Note that our assumptions in this Appendix allow us to apply the results of Appendices C and D.

We first consider g^R_1 and we obtain from (3.55) that

$$|g^R_1(\theta,\chi,\zeta,\varepsilon,k,a)|=|g_1(\theta,\chi,\zeta,\varepsilon,k+\varepsilon a)|, \tag{E3}$$

whence, by (C26),

$$|g^R_1(\theta,\chi,\zeta,\varepsilon,k,a)|\le C_1(\chi_0,\varepsilon_0,k+\varepsilon a,d_2), \tag{E4}$$

where C_1 is given by (C26). Note that, by (C26), $C_1(\chi_0,\varepsilon_0,\nu,d_2)$ is decreasing wrt ν whence

$$C_1(\chi_0,\varepsilon_0,k+\varepsilon a,d_2)\le C_1(\chi_0,\varepsilon_0,k-1/2,d_2)=:C^R_1(\chi_0,\varepsilon_0,k,d_2), \tag{E5}$$

so that, by (E4),

$$|g^R_1(\theta,\chi,\zeta,\varepsilon,k,a)|\le C^R_1(\chi_0,\varepsilon_0,k,d_2), \tag{E6}$$

where C^R_1 is given by (E5).

We now consider g_2^R and we obtain from (D4) that

$$|g_2^R(\theta, \chi, \zeta, \varepsilon, k, a)| \leq |g_{2,1}^R(\theta, \chi, \zeta, \varepsilon, k, a)| + |g_{2,2}^R(\theta, \chi, \zeta, \varepsilon, k, a)|. \tag{E7}$$

Note that, by (C29) and (D5),

$$|g_{2,1}^R(\theta, \chi, \zeta, \varepsilon, k, a)| = |g_2(\theta, \chi, \zeta, \varepsilon, k + \varepsilon a)| \leq C_2(\chi_0, \varepsilon_0, k + \varepsilon a, d_2), \tag{E8}$$

where C_2 is given by (C29). Note that, by (C29), $C_2(\chi_0, \varepsilon_0, \nu, d_2)$ is decreasing wrt ν whence

$$C_2(\chi_0, \varepsilon_0, k + \varepsilon a, d_2) \leq C_2(\chi_0, \varepsilon_0, k - 1/2, d_2) =: C_{2,1}^R(\chi_0, \varepsilon_0, k, d_2), \tag{E9}$$

so that, by (E8),

$$|g_{2,1}^R(\theta, \chi, \zeta, \varepsilon, k, a)| \leq C_{2,1}^R(\chi_0, \varepsilon_0, k, d_2), \tag{E10}$$

where $C_{2,1}^R$ is given by (E9). We also have, by (D6),

$$\begin{aligned} |g_{2,2}^R(\theta, \chi, \zeta, \varepsilon, k, a)| &= \left| K^2 a(\theta - \Upsilon_0 \sin\zeta - \Upsilon_1 \sin 2\zeta)(\cos\zeta + \Delta P_{x0}) \mathrm{sinc}(\kappa_5/2) \sin\left(\frac{1}{2}[2\kappa_4 + \kappa_5]\right) \right| \\ &\leq K^2 |a| |\theta - \Upsilon_0 \sin\zeta - \Upsilon_1 \sin 2\zeta| (1 + |\Delta P_{x0}|). \end{aligned} \tag{E11}$$

Of course, by (E1),

$$|\theta - \Upsilon_0 \sin\zeta - \Upsilon_1 \sin 2\zeta| \leq |\theta| + |\Upsilon_0| + |\Upsilon_1| \leq |\theta_0| + d_1 + |\Upsilon_0| + |\Upsilon_1|, \tag{E12}$$

whence, by (E11),

$$|g_{2,2}^R(\theta, \chi, \zeta, \varepsilon, k, a)| \leq K^2 |a| (1 + |\Delta P_{x0}|)(|\theta_0| + d_1 + |\Upsilon_0| + |\Upsilon_1|) =: C_{2,2}^R(\theta_0, a, d_1). \tag{E13}$$

We conclude from (E7), (E10), and (E13) that

$$|g_2^R(\theta, \chi, \zeta, \varepsilon, k, a)| \leq C_{2,1}^R(\chi_0, \varepsilon_0, k, d_2) + C_{2,2}^R(\theta_0, a, d_1) =: C_2^R(\theta_0, \chi_0, \varepsilon_0, k, a, d_1, d_2), \tag{E14}$$

where $C_{2,1}^R$ is given by (E9) and $C_{2,2}^R$ is given by (E13).

APPENDIX F: DERIVATIVES FOR LOW GAIN PROBLEM

We here derive (F6) which is needed in Sec. III G. By (3.177) we have

$$\begin{aligned} &v_1'(\cdot, 1) = 2v_2(\cdot, 1), \qquad v_1(0, 1) = \theta_0, \\ &v_2'(\tau, 1) = -\mathcal{A}(k, \Delta P_{x0}) \cos[k v_1(\tau, 1) - a\tau] = -\frac{\mathcal{A}(k, \Delta P_{x0})}{2} \exp[u(\tau)] + cc, \qquad v_2(0, 1) = \chi_0, \end{aligned} \tag{F1}$$

where

$$u(\tau) := i[k v_1(\tau, 1) - a\tau]. \tag{F2}$$

It follows from (F1) that

$$\begin{aligned} v_2''(\tau, 1) &= \mathcal{A}(k, \Delta P_{x0})[k v_1'(\tau, 1) - a] \sin[k v_1(\tau, 1) - a\tau] = \mathcal{A}(k, \Delta P_{x0})[2k v_2(\tau, 1) - a] \sin[k v_1(\tau, 1) - a\tau] \\ &= -\frac{\mathcal{A}(k, \Delta P_{x0})}{2} \exp[u(\tau)] u'(\tau) + cc, \\ v_2'''(\cdot, 1) &= -\frac{\mathcal{A}(k, \Delta P_{x0})}{2} \exp(u)[u'' + (u')^2] + cc, \qquad v_2''''(\cdot, 1) = -\frac{\mathcal{A}(k, \Delta P_{x0})}{2} \exp(u)[u''' + 3u'u'' + (u')^3] + cc, \end{aligned} \tag{F3}$$

and from (F1)–(F3) that

$$u'(\tau) = i[kv_1'(\cdot, 1) - a] = i[2kv_2(\cdot, 1) - a], \qquad u''(\tau) = i2kv_2'(\tau, 1) = -i2k\mathcal{A}(k, \Delta P_{x0})\cos(kv_1(\tau, 1) - a\tau),$$
$$u'''(\tau) = i2kv_2''(\tau, 1) = i2k\mathcal{A}(k, \Delta P_{x0})[2kv_2(\tau, 1) - a]\sin[kv_1(\tau, 1) - a\tau]. \tag{F4}$$

We conclude from (F1), (F2), and (F4) that

$$\begin{aligned} u(0) &= ikv_1(0, 1) = ik\theta_0, \qquad u'(0) = i[2kv_2(0, 1) - a] = i[2k\chi_0 - a], \\ u''(0) &= -i2k\mathcal{A}(k, \Delta P_{x0})\cos[kv_1(0, 1)] \\ &= -i2k\mathcal{A}(k, \Delta P_{x0})\cos(k\theta_0), \\ u'''(0) &= i2k\mathcal{A}(k, \Delta P_{x0})[2kv_2(0, 1) - a]\sin[kv_1(0, 1)] \\ &= i2k\mathcal{A}(k, \Delta P_{x0})(2k\chi_0 - a)\sin(k\theta_0), \end{aligned} \tag{F5}$$

whence, by (F1) and (F3),

$$\begin{aligned} v_2'(0, 1) &= -\mathcal{A}(k, \Delta P_{x0})\cos[kv_1(0, 1)] = -\mathcal{A}(k, \Delta P_{x0})\cos(k\theta_0), \\ v_2''(0, 1) &= \mathcal{A}(k, \Delta P_{x0})[2kv_2(0, 1) - a]\sin[kv_1(0, 1)] \\ &= \mathcal{A}(k, \Delta P_{x0})(2k\chi_0 - a)\sin(k\theta_0), \\ v_2'''(0, 1) &= -\frac{\mathcal{A}(k, \Delta P_{x0})}{2}\exp[u(0)]\{u''(0) + [u'(0)]^2\} + cc \\ &= -\frac{\mathcal{A}(k, \Delta P_{x0})}{2}\exp(ik\theta_0)\{-i2k\mathcal{A}(k, \Delta P_{x0})\cos(k\theta_0) - [2k\chi_0 - a]^2\} + cc \\ &= -\mathcal{A}(k, \Delta P_{x0})\{2k\mathcal{A}(k, \Delta P_{x0})\sin(k\theta_0)\cos(k\theta_0) - [2k\chi_0 - a]^2\cos(k\theta_0)\} \\ &= \mathcal{A}(k, \Delta P_{x0})\{-k\mathcal{A}(k, \Delta P_{x0})\sin(2k\theta_0) + [2k\chi_0 - a]^2\cos(k\theta_0)\}, \\ v_2''''(0, 1) &= -\frac{\mathcal{A}(k, \Delta P_{x0})}{2}\exp[u(0)]\{u'''(0) + 3u'(0)u''(0) + [u'(0)]^3\} + cc \\ &= -\frac{\mathcal{A}(k, \Delta P_{x0})}{2}\exp(ik\theta_0)\{i2k\mathcal{A}(k, \Delta P_{x0})(2k\chi_0 - a)\sin(k\theta_0) \\ &\quad + 6k\mathcal{A}(k, \Delta P_{x0})[2k\chi_0 - a]\cos(k\theta_0) - i[2k\chi_0 - a]^3\} + cc \\ &= -\frac{\mathcal{A}(k, \Delta P_{x0})}{2}\{-4k\mathcal{A}(k, \Delta P_{x0})(2k\chi_0 - a)\sin^2(k\theta_0) \\ &\quad + 12k\mathcal{A}(k, \Delta P_{x0})[2k\chi_0 - a]\cos^2(k\theta_0) + 2[2k\chi_0 - a]^3\sin(k\theta_0)\} \\ &= \mathcal{A}(k, \Delta P_{x0})\{2k\mathcal{A}(k, \Delta P_{x0})(2k\chi_0 - a)\sin^2(k\theta_0) - 6k\mathcal{A}(k, \Delta P_{x0})[2k\chi_0 - a]\cos^2(k\theta_0) \\ &\quad - [2k\chi_0 - a]^3\sin(k\theta_0)\}. \end{aligned} \tag{F6}$$

APPENDIX G: ESTIMATE OF E_r/cB_u

In this Appendix we aim to estimate the magnitude of the electric field. The basic field equation is

$$\left(\frac{\partial^2}{\partial t^2} - c^2\frac{\partial^2}{\partial z^2}\right)E_x(z, t) = -cZ_{vac}\frac{\partial j}{\partial t}(z, t), \tag{G1}$$

where $Z_{\rm vac} = 1/c\epsilon_0$ is the free space impedance and

$$j(z, t) := -\frac{ecK}{\Sigma_\perp}\cos(k_u z)\sum_{n=1}^{N}\frac{1}{\gamma_n(t)}\delta[z - z_n(t)] \approx -\frac{ecKN}{\gamma_c\Sigma_\perp}\cos(k_u z)\frac{1}{N}\sum_{n=1}^{N}\delta[z - z_n(t)], \tag{G2}$$

with $\Sigma_\perp$ being the transverse emittance, see [13,38]. We proceed in two ways. In the first we solve (G1) and (G2) directly and in the second we use Fourier transforms.

The unique solution of the homogeneous IVP at $t = 0$ is

$$E_x(z,t) = -\frac{Z_{\text{vac}}}{2}\int_0^t ds\int_{z-ct+cs}^{z+ct-cs} dy\frac{\partial j}{\partial s}(y,s) = -\frac{Z_{\text{vac}}}{2}[U_-(z,t)+U_+(z,t)], \tag{G3}$$

where

$$U_-(z,t) := \int_{z-ct}^{z} dy\Big[j\Big(y,t+\frac{1}{c}(y-z)\Big) - j(y,0)\Big], \tag{G4}$$

$$U_+(z,t) := \int_{z}^{z+ct} dy\Big[j\Big(y,t-\frac{1}{c}(y-z)\Big) - j(y,0)\Big]. \tag{G5}$$

The first equality in (G3) is often obtained using Duhamel's principle and d'Alembert's formula and the second equality is obtained after changing the order of integration. To obtain our estimate we consider $z_n(t) = \beta_c ct + z_n(0)$ which is quite crude (but may suffice for a rough estimate) and where the nonnegative β_c is determined by $\beta_c^2 = (\gamma_c^2 - 1)/\gamma_c^2$. We obtain [39] $U_+ \ll U_-$ and

$$U_-(z,t) \approx -\frac{2ecK\gamma_c N}{\Sigma_\perp}\frac{1}{N}\sum_{n=1}^{N} I_n(z,t)\cos\{2k_u\gamma_c^2[z-ct-z_n(0)]\}, \tag{G6}$$

where

$$I_n(z,t) := \begin{cases} 1 & \text{if } z_n(t) < z < z_n(0)+ct \\ 0 & \text{if otherwise.}\end{cases} \tag{G7}$$

So if all the particles contributed at z, which they do not, then $U_-(z,t) = O(\frac{2ecK\gamma_c N}{\Sigma_\perp})$ and $E_{r1} = \frac{Z_{vac}ecK\gamma_c N}{\Sigma_\perp}$ would be a typical value of the field E_x at (z,t).

We now give a second estimate, E_{r2}, of E_r. Following [38] which is based on [13] we Fourier transform (G1) by defining

$$\hat{E}_x(z,\omega) := \frac{1}{2\pi}\int_{-\infty}^{\infty} dsE_x\Big(z,\frac{z}{c}-\frac{s}{ck_r}\Big)\exp(-i\omega s). \tag{G8}$$

The Fourier inversion theorem gives

$$E_x(z,t) = \int_{-\infty}^{\infty} d\omega\hat{E}_x(z,\omega)\exp(i\omega k_r[z-ct]). \tag{G9}$$

We define $\hat{j}(z,\omega)$ in the same way as $\hat{E}_x(z,\omega)$ whence, in the slowly varying approximation, (G1) reduces to

$$\frac{\partial\hat{E}_x}{\partial z}(z,\omega) = -\frac{Z_{\text{vac}}}{2}\hat{j}(z,\omega), \tag{G10}$$

and from (G2) we obtain

$$\hat{j}(z,\omega) = -\frac{ecKNk_r}{2\pi\beta_c\gamma_c\Sigma_\perp}\check{j}(z,\omega), \tag{G11}$$

where

$$\check{j}(z,\omega) := \cos(k_u z)\exp(-i\omega k_r z)\frac{1}{N}\sum_{n=1}^{N}\exp[i\omega ck_r T_n(z)]. \tag{G12}$$

Here the function T_n is the inverse of the function z_n. To obtain our estimate we note that $|\check{j}|$ is bounded by 1 and replace it by 1 which is quite crude but may suffice for a rough estimate. Inserting this into (G10) and integrating we obtain

$$\hat{E}_x(z,\omega) = O\Big(\frac{Z_{\text{vac}}}{2}\frac{ecKNk_r}{2\pi\beta_c\gamma_c\Sigma_\perp}\frac{1}{k_u}k_u z\Big), \tag{G13}$$

and, for $k_u z = O(1)$,

$$\hat{E}_x = O(E_{r2}), \qquad E_{r2} := \frac{Z_{\text{vac}}}{4\pi}\frac{ecKN}{\Sigma_\perp}\frac{k_r}{k_u\beta_c\gamma_c^2}\gamma_c. \tag{G14}$$

We now have, recalling that $K = 3.7$ in LCLS,

$$\frac{E_{r1}}{E_{r2}} = 4\pi\frac{k_u\gamma_c^2}{k_r} = 4\pi/K_r = 2\pi\left(1+\frac{K^2}{2}\right) \approx 2\pi[1+(3.7)^2/2] \approx 49, \tag{G15}$$

and we calculate E_{r2}/cB_u. From (G14),

$$\frac{E_{r2}}{cB_u} = \frac{Z_{\text{vac}}c}{4\pi}\frac{eK}{cB_u}\frac{k_r}{k_u\gamma_c^2}\gamma_c\frac{N}{\Sigma_\perp}. \tag{G16}$$

Now $K/cB_u = e/mc^2k_u$ and $k_r/k_u\gamma_c^2 = 2(1+K^2/2)^{-1}$ therefore

$$\frac{E_{r2}}{cB_u} = \frac{Z_{\text{vac}}c}{4\pi}\frac{e^2}{mc^2}\frac{1}{k_u}\frac{2}{(1+K^2/2)}\gamma_c\frac{N}{\Sigma_\perp} = r_e\frac{1}{k_u}\frac{2}{(1+K^2/2)}\gamma_c\frac{N}{\Sigma_\perp}, \tag{G17}$$

where r_e denotes the classical electron radius. Furthermore,

$$r_e \approx 2.82\times 10^{-15}\ \text{m}, \qquad \frac{1}{k_u} = \frac{3\ \text{cm}}{2\pi}, \qquad \frac{2}{(1+K^2/2)} \approx 0.255, \qquad \gamma_c = 10^4,$$

and so

$$\frac{E_{r2}}{cB_u} \approx 0.034\times 10^{-12}\ \text{m}^2\frac{N}{\Sigma_\perp} \approx 34, \qquad \frac{E_{r1}}{cB_u} = \frac{E_{r2}}{cB_u}\frac{E_{r1}}{E_{r2}} \approx 34\times 49 \approx 1700,$$

for $N = 10^9$ and $\Sigma_\perp = 1\ \text{mm}^2$.

APPENDIX H: IVP FOR A SYSTEM OF DIFFERENTIAL INEQUALITIES

Here we present and verify a solution of the IVP for a system of differential inequalities which is used in Secs. IV A and IV B. Consider the IVP for

$$R_1'(\zeta) \le a_1R_2(\zeta), \tag{H1}$$

$$R_2'(\zeta) \le a_2R_1(\zeta), \tag{H2}$$

where $a_1, a_2 > 0$ and R_1, R_2 are continuously differentiable. We want to show, for $\zeta \ge 0$, that

$$R_1(\zeta) \le r_1(\zeta), \qquad R_2(\zeta) \le r_2(\zeta), \tag{H3}$$

where

$$r_1' = a_1r_2, \qquad r_1(0) = R_1(0), \tag{H4}$$

$$r_2' = a_2r_1, \qquad r_2(0) = R_2(0). \tag{H5}$$

We do this in two ways. First we define $\hat{r}_j(\zeta) := R_j(\zeta) - r_j(\zeta)$ for $j = 1, 2$, $\zeta \ge 0$ whence, by (H1), (H2), (H4), and (H5),

$$\hat{r}_1'(\zeta) \le a_1\hat{r}_2(\zeta), \qquad \hat{r}_2'(\zeta) \le a_2\hat{r}_1(\zeta), \qquad \hat{r}_1(0) = \hat{r}_2(0) = 0. \tag{H6}$$

Clearly we have to show that, for $j = 1, 2$, $\zeta \ge 0$,

$$\hat{r}_j(\zeta) \le 0. \tag{H7}$$

It follows from (H6) that

$$\hat{r}_1'(\zeta) \le a_1\int_0^\zeta ds\hat{r}_2'(s) \le a_1a_2\int_0^\zeta ds\hat{r}_1(s), \qquad \hat{r}_2'(\zeta) \le a_2\int_0^\zeta ds\hat{r}_1'(s) \le a_1a_2\int_0^\zeta ds\hat{r}_2(s),$$

i.e.,

$$\hat{r}_j'(\zeta) \le a_0^2 \int_0^\zeta ds \hat{r}_j(s), \tag{H8}$$

where $a_0 := \sqrt{a_1 a_2}$. It follows from (H8) and by partial integration that

$$\exp(-a_0\zeta)\hat{r}_j(\zeta) + a_0 \int_0^\zeta ds \exp(-a_0 s)\hat{r}_j(s) = \int_0^\zeta ds \exp(-a_0 s)\hat{r}_j'(s) \le a_0^2 \int_0^\zeta ds \exp(-a_0 s) \int_0^s d\tilde{s}\hat{r}_j(\tilde{s})$$
$$= -a_0 \exp(-a_0\zeta) \int_0^\zeta ds \hat{r}_j(s) + a_0 \int_0^\zeta ds \exp(-a_0 s)\hat{r}_j(s), \tag{H9}$$

which entails

$$\hat{r}_j(\zeta) \le -a_0 \int_0^\zeta ds \hat{r}_j(s). \tag{H10}$$

Abbreviating

$$\check{r}_j(\zeta) := \int_0^{\zeta/a_0} ds \hat{r}_j(s), \tag{H11}$$

we obtain from (H10)

$$\check{r}_j'(\zeta) = \frac{1}{a_0}\hat{r}_j(\zeta/a_0) \le -\int_0^{\zeta/a_0} ds \hat{r}_j(s) = -\check{r}_j(\zeta), \tag{H12}$$

whence

$$0 \ge \exp(\zeta)[\check{r}_j(\zeta) + \check{r}_j'(\zeta)] = [\exp(\zeta)\check{r}_j(\zeta)]', \tag{H13}$$

so that $\exp(\zeta)\check{r}_j(\zeta)$ is decreasing wrt ζ which entails, by (H11), that

$$0 = \exp(0)\check{r}_j(0) \ge \exp(\zeta)\check{r}_j(\zeta), \tag{H14}$$

i.e.,

$$\check{r}_j(\zeta) \le 0. \tag{H15}$$

We conclude from (H8), (H11), and (H15) that

$$\hat{r}_j'(\zeta) \le a_0^2 \int_0^\zeta ds \hat{r}_j(s) = \check{r}_j(a_0\zeta) \le 0, \tag{H16}$$

whence $\hat{r}_j(\zeta)$ is decreasing wrt ζ so that (H7) follows from (H6).

The result in (H3) is a special case of a much more general theorem on pages 112–113 of [33]. That proof simplifies in the special case here and we present it for the interested reader. The proof proceeds by cleverly introducing a comparison function $\mathbf{h}$. Here

$$\mathbf{h}(\zeta) = \begin{pmatrix} h_1(\zeta) \\ h_2(\zeta) \end{pmatrix} := a_4 \exp(2a_3\zeta)\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \tag{H17}$$

where $a_3 := \max(a_1, a_2)$, $a_4 > 0$. Then

$$h_1' = 2a_3 h_1 = 2a_3 h_2 > a_1 h_2, \tag{H18}$$

$$h_2' = 2a_3 h_2 = 2a_3 h_1 > a_2 h_1, \tag{H19}$$

and we have, by (H6),

$$\hat{r}_1' - a_1\hat{r}_2 \le 0 < h_1' - a_1 h_2, \tag{H20}$$

$$\hat{r}_2' - a_2\hat{r}_1 \le 0 < h_2' - a_2 h_1. \tag{H21}$$

We now show that, for $j = 1, 2$, $\zeta \ge 0$,

$$\hat{r}_j(\zeta) \le h_j(\zeta). \tag{H22}$$

Suppose that (H22) is wrong then there exists a smallest $\zeta_0 > 0$ such that an index j_0 exists with

$$\hat{r}_{j_0}(\zeta_0) = h_{j_0}(\zeta_0), \tag{H23}$$

where we used that, by (H6) and (H17) and for $j = 1, 2$,

$$\hat{r}_j(0) = 0 < a_4 = h_j(0). \tag{H24}$$

Clearly, for $j = 1, 2$, $0 \le \zeta < \zeta_0$,

$$\hat{r}_j(\zeta) < h_j(\zeta). \tag{H25}$$

Without loss of generality we take $j_0 = 1$ whence, for $0 \le \zeta \le \zeta_0$,

$$\hat{r}_2(\zeta) \le h_2(\zeta). \tag{H26}$$

It follows from (H25) that at the first intersection

$$\hat{r}_1'(\zeta_0) \ge h_1'(\zeta_0). \tag{H27}$$

But by (H20) and (H26)

$$\hat{r}_1'(\zeta_0) - h_1'(\zeta_0) < a_1[\hat{r}_2(\zeta_0) - h_2(\zeta_0)] \le 0, \tag{H28}$$

which is a contradiction.

[1] P. Baxevanis, R. D. Ruth, and Z. Huang, Phys. Rev. ST Accel. Beams **16**, 010705 (2013).
[2] http://www-ssrl.slac.stanford.edu/lcls/lcls_parms.html.
[3] http://flash.desy.de/accelerator/.
[4] Note that in Sec. II, where the field will not be monochromatic, we will have the seven parameters ΔP_{x0}, γ_c, ε, B_u, k_u, E_r, k_r, a general function h and the single constraint of (1.5).
[5] K.-J. Kim (private communication).
[6] J. Murdock, *Classics in Applied Mathematics*, (SIAM, Philadelphia, PA, 1999), Vol. 27, first published by Wiley in 1991.
[7] K. Nozaki and Y. Oono, Phys. Rev. E **63**, 046101 (2001).
[8] The literature on Hamiltonian perturbation theory (HPT) is vast. Basically it involves canonical transformations to simplify the Hamiltonian in leading order of a perturbation parameter. The transformations are often defined in terms

of Lie generating functions (e.g., Lie series or transformations) which is more direct than the use of mixed generating functions (although not necessarily better). Two highlights of HPT are the statements and proofs of the Nekhoroshev and KAM theorems. The Nekhoroshev theorem can be viewed as the *ultimate* averaging theorem. An interesting discussion of these theorems, with a focus on the KAM case, can be found in the book ''The KAM Story: A Friendly Introduction to the History, Content, and Significance of Classical Kolmogorov-Arnold-Moser Theory'', by H. Scott Dumas to be published by World Scientific (see also Chapter 7 of [9]). The Lie method is briefly discussed in Sec. 3.11. A Hamiltonian, Lie transformation approach to a more general version of the problem we consider in this paper is being pursued by R. R. Lindberg. His approach is influenced by the early work of Littlejohn on the so-called guiding-center motion. Lindberg's article is entitled ''A derivation of the three-dimensional free-electron particle equations based upon Lie transformation techniques'' and a recent review of the guiding-center problem is presented in ''Hamiltonian theory of guiding-center motion,'' J. R. Cary and A. B. Brizard, Rev. Mod. Phys. **81**, 693 (2009).

[9] P. Lochak and C. Meunier, *Applied Mathematical Sciences* 72 (Springer-Verlag, New York 1988), translated by H. S. Dumas.

[10] J. A. Sanders, F. Verhulst, and J. Murdock, *Applied Mathematical Sciences* (Springer, New York, 2007), Vol. 59, 2nd ed.

[11] J. Murdock, Normal forms. Scholarpedia, 1, 1902 (2006).

[12] J. A. Sanders, Averaging. Scholarpedia, 1, 1760 (2006).

[13] K.-J. Kim, Z. Huang, and R. Lindberg (unpublished).

[14] J. A. Ellison, A. W. Saenz, and H. S. Dumas, J. Diff. Equ. **84**, 383 (1990).

[15] H. S. Dumas, J. A. Ellison, and K. Heinemann (unpublished).

[16] B. J. Besjes, J. Me'canique **8**, 357 (1969).

[17] P. Schmüser, M. Dohlus, and J. Rossbach, *Springer Tracts in Modern Physics* (Springer-Verlag, Berlin, 2008), Vol. 229.

[18] J. B. Murphy and C. Pellegrini, in *Laser Handbook*, edited by W. B. Colson, C. Pellegrini, and R. Renieri (North-Holland, Amsterdam, 1990), Vol. 6, pp. 9–70.

[19] E. L. Saldin, E. A.Schneidmiller, and M. V. Yurkov, (Springer-Verlag, Berlin, 2000).

[20] James A. Clarke, *Oxford Series on Synchrotron Radiation* (Oxford University Press, New York, 2004).

[21] This transformation to the slowly varying θ in (2.47) works nicely because ζ (equivalently z) is the independent variable. If we had stayed with t as the independent variable this step would not work.

[22] A periodic function has one basic frequency whereas a quasiperiodic function has a finite number of basic frequencies (tunes) denoted by $\nu = (\nu_1, \nu_2, \ldots, \nu_k)$. In our paper ''Quasiperiodic spin-orbit motion and spin tunes in storage rings'' [40], we give the following definition: A function $f\colon \mathbb{R} \to \mathbb{R}$ is said to be *quasiperiodic* with tune vector ν in $\mathbb{R}^k$ if a continuous and 2π-periodic function $F\colon \mathbb{R}^k \to \mathbb{R}$ exists such that $f(t) = F(\nu t)$. Clearly f_2 in (3.24) is quasiperiodic in ζ in this sense with tune vector $(1, \nu)$. We also use this definition in [15], however there is no agreement in the literature concerning the definition. In KAM related studies it often refers to linear flow on the k torus and the definition by A. M. Samoilenko [Quasiperiodic Oscillations. Scholarpedia 2, 1783 (2007)] is similar to ours but requires the ν vector to be nonresonant in the sense of [40].

[23] Z. Huang and K.-J. Kim, Phys. Rev. ST Accel. Beams **10**, 034801 (2007).

[24] We remind the reader that phase plane orbits of smooth autonomous systems cannot cross for if they did it would contradict uniqueness. This makes the PPP ideal for the study of such systems and it makes apparent the qualitative behavior of all solutions. All that is missing is the time it takes to move from one point on an orbit to another. Conversely, the phase plane is not useful for studying the *family* of orbits of the nonautonomous system (3.61) and (3.62) because orbits can cross. However, looking at a small subset of orbits can be useful.

[25] The well-known phase plane portrait for the pendulum equation can be found in many places, e.g., see Fig. 6.4, p. 248 of [26].

[26] F. Brauer and J. A. Nohel, *The Qualitative Theory of Ordinary Differential Equations: An Introduction* (Dover Publications, New York, 1989.

[27] The potential plane is simply a plot of the potential $1 - \cos X$ versus X placed above the phase plane portrait. This is nicely illustrated in [26], Figs. 6.5 and 6.6 on p. 249. See also Fig. 95, p. 142 of [29].

[28] See, e.g., [26], Sec. 6.2, for a derivation of the period as a function of amplitude as well as a proof that the integral is well defined as an improper integral (note that the integrand is singular). The formula for the pendulum equation is given explicitly on p. 244.

[29] V. I. Arnold, *Ordinary Differential Equations* (Springer-Verlag, Berlin, 1991), 3rd ed., translated by R. Cooke.

[30] There are several good sources for the theory of second-order autonomous systems. Here we mention Chapters 15 and 16 in E. A. Coddington and N. Levinson, *The Theory of Ordinary Differential Equations* (McGraw-Hill, New York, 1955); Chapters 4 and 5 in W. Hurewicz, *Lectures on Ordinary Differential Equations* (MIT Press, Cambridge, MA, 1958); Chapters 5 and 6 in R. A. Struble, *Nonlinear Differential Equations* (McGraw-Hill, New York, 1962); and Chapter IV in G. Sansone and R. Conte, *Non-Linear Differential Equations* (The Macmillian Company, New York, 1964).

[31] J. K. Hale, *Ordinary Differential Equations* (Krieger Publishing Company, Malabar, Florida, 1980).

[32] J. A. Ellison and H. J. Shih, in *Accelerator Physics Lectures at the Superconducting Super Collider*, AIP Conf. Proc. No. 326, edited by Y. Yan and M. Syphers (AIP, New York, 1995), pp. 590–632.

[33] W. Walter, *Ordinary Differential Equations* (Springer-Verlag, New York 1998).

[34] H. S. Dumas, J. A. Ellison, A. W. Saenz, Ann. Phys. (N.Y.) **209**, 97 (1991).

[35] H. S. Dumas, J. A. Ellison, F. Golse, Physica (Amsterdam) **146**, 341 (2000).

[36] H. S. Dumas, J. A. Ellison, M. Vogt, SIAM J. Appl. Dyn. Syst. **3**, 409 (2004).

[37] The literature on separatrix splitting is large, our work is contained in J. A. Ellison, M. Kummer, and A. W. Saenz, J. Dyn. Differ. Equ. **5**, 241 (1993); M. Kummer, J. A. Ellison, and A. W. Saenz, in *Asymptotics Beyond All Orders*, NATO Advanced Study Institutes, Ser. B, Vol. 284, edited by S. Tanveer, H. Segur, and H. Levine (Plenum Press, New York, 1992), p. 197.
[38] R. Warnock, One-Dimensional FEL Equations, Notes, 2013.
[39] J. A. Ellison and K. Heinemann (unpublished).
[40] D. P. Barber, J. A. Ellison, and K. Heinemann, Phys. Rev. ST Accel. Beams **7**, 124002 (2004).