# Tutorial:
# Sub-Nyquist artifacts and sampling moiré effects

**Isaac Amidror**

IC-LSP, EPFL
1015 Lausanne
Switzerland

ABSTRACT

Sampling moiré effects are well known in signal processing. They occur when a continuous periodic signal $g(x)$ is sampled using a sampling frequency $f_s$ that does not respect the Nyquist condition, and the signal frequency $f$ folds-over and gives a new, false low frequency in the sampled signal. However, some visible beating artifacts may also occur in the sampled signal when $g(x)$ is sampled using a sampling frequency $f_s$ which fully respects the Nyquist condition. We call these phenomena *sub-Nyquist artifacts*. Although these beating effects have already been reported in the literature, their detailed mathematical behaviour is not widely known. In this paper we study the behaviour of these phenomena and compare it with analogous results from the moiré theory. We show that both sampling moirés and sub-Nyquist artifacts obey the same basic mathematical rules, in spite of the differences between them. This leads us to a unified approach that explains all of these phenomena, and puts them under the same roof. In particular, it turns out that all of these phenomena occur when the signal frequency $f$ and the sampling frequency $f_s$ satisfy $f \approx (m/n)f_s$ with integer $m,n$, where $m/n$ is a reduced integer ratio; cases with $n = 1$ correspond to true sampling moiré effects.

**Keywords**: sampling, reconstruction, moiré effects, sub-Nyquist artifacts, sampling theorem

## 1. Introduction

According to the classical sampling theorem, all the information in a continuous signal $g(x)$ is preserved in its sampled version $g(x_k)$ if the sampling frequency $f_s$ respects the Nyquist condition, i.e. if $f_s$ is at least twice the highest frequency contained in $g(x)$.[1] When sampling a continuous periodic signal $g(x)$ using a sampling frequency $f_s$ that does not respect the Nyquist condition, various moiré or aliasing artifacts may appear in the resulting sampled signal $g(x_k)$ (see, for example, Fig. 1(i),(j)). Nevertheless, it is also known [1, pp. 222, 225], [2, p. 642], [3] that some beating artifacts (pseudo moirés) may appear in the sampled signal $g(x_k)$ when $g(x)$ is sampled with a sampling frequency $f_s$ which *does* respect the Nyquist condition, but is still very close to the Nyquist limit itself (Fig. 1(f)). [2]

Although these beating artifacts are not really new, they have rarely been treated in the literature, and they still remain poorly understood. This can be best attested by the

[1] Note that lower sampling frequencies may suffice for some special types of signals, but the classical sampling theorem addresses the general case.

[2] The case of Fig. 1(h), where aliasing already does occur, will be explained later (see also Remark 8).

numerous queries on this subject that are regularly posted in various forums or web sites by people involved in signal processing (see, for example, [4]; [5]; and in particular [6] who asserts having browsed through nearly 20 digital signal processing textbooks, but in vain).

These beating effects are often considered in the literature as reconstruction artifacts, but no deeper analysis is provided to explain their nature and properties. And sometimes they are even simply dismissed as "borderline artifacts" that occur when we get too close to the limits of the sampling theorem. However, it turns out that such beating effects are not limited to cases bordering on the Nyquist condition [7]. On the contrary, similar phenomena may also appear in many other cases that are by far *within* the Nyquist range. This happens, for example, when sampling a cosine signal $g(x) = \cos(2\pi fx)$ whose frequency $f$ is close to $\frac{1}{3}f_s$ or $\frac{1}{4}f_s$, i.e. far below half of the sampling frequency. These beating artifacts are clearly not aliasing effects, since they appear where the Nyquist condition is fully satisfied. As we will see in greater detail in Sec. 2, these beating artifacts also have several other intriguing properties. In particular, their periods (or frequencies) are not represented in the Fourier spectrum, although they are clearly visible in the sampled signal. But although these beating artifacts are not aliasing or moiré effects, in many ways their behaviour is very similar to that of true sampling moirés due to aliasing. This explains, indeed, why they are often called in the literature "pseudo moirés". We will call these artifacts hereinafter *sub-Nyquist artifacts*.

The fact that these moiré-like artifacts are not visible by our main moiré investigation tool, the Fourier theory, makes them more difficult to analyze, but therefore also more interesting and challenging. In the present paper we address this challenge, and show how we can investigate these phenomena and incorporate them, together with the classical sampling moiré effects, within a common unified framework, based on sampling theory and on the moiré theory.

Our work is structured as follows: We start in Sec. 2 with an overview of the basic notions and terms that will be used in the sequel. (Readers who already master parts of this material may simply skip the paragraphs in question.) In Sec. 3 we provide the initial setting, and illustrate the nature of the phenomena in question by means of some typical examples, using the simple case of the cosine signal $g(x) = \cos(2\pi fx)$. Then, in Sec. 4 we derive a fundamental theorem which clearly explains the mathematical nature of the sub-Nyquist artifacts, and we show that classical sampling moiré effects are in fact a particular case of this theorem. Thus, by treating both classical sampling moirés and sub-Nyquist artifacts with the same unified tools, we put them all under the same roof. Our theorem as presented in Sec. 4 is limited to the particular case of the cosine signal, but then, in Sec. 5, we extend it to any general periodic signals. In Sec. 6 we discuss the nature of the sub-Nyquist artifacts as reconstruction artifacts. In Sec. 7 we discuss the potential impact of sub-Nyquist artifacts in signal-processing applications. And finally, our conclusions are presented in Sec. 8.

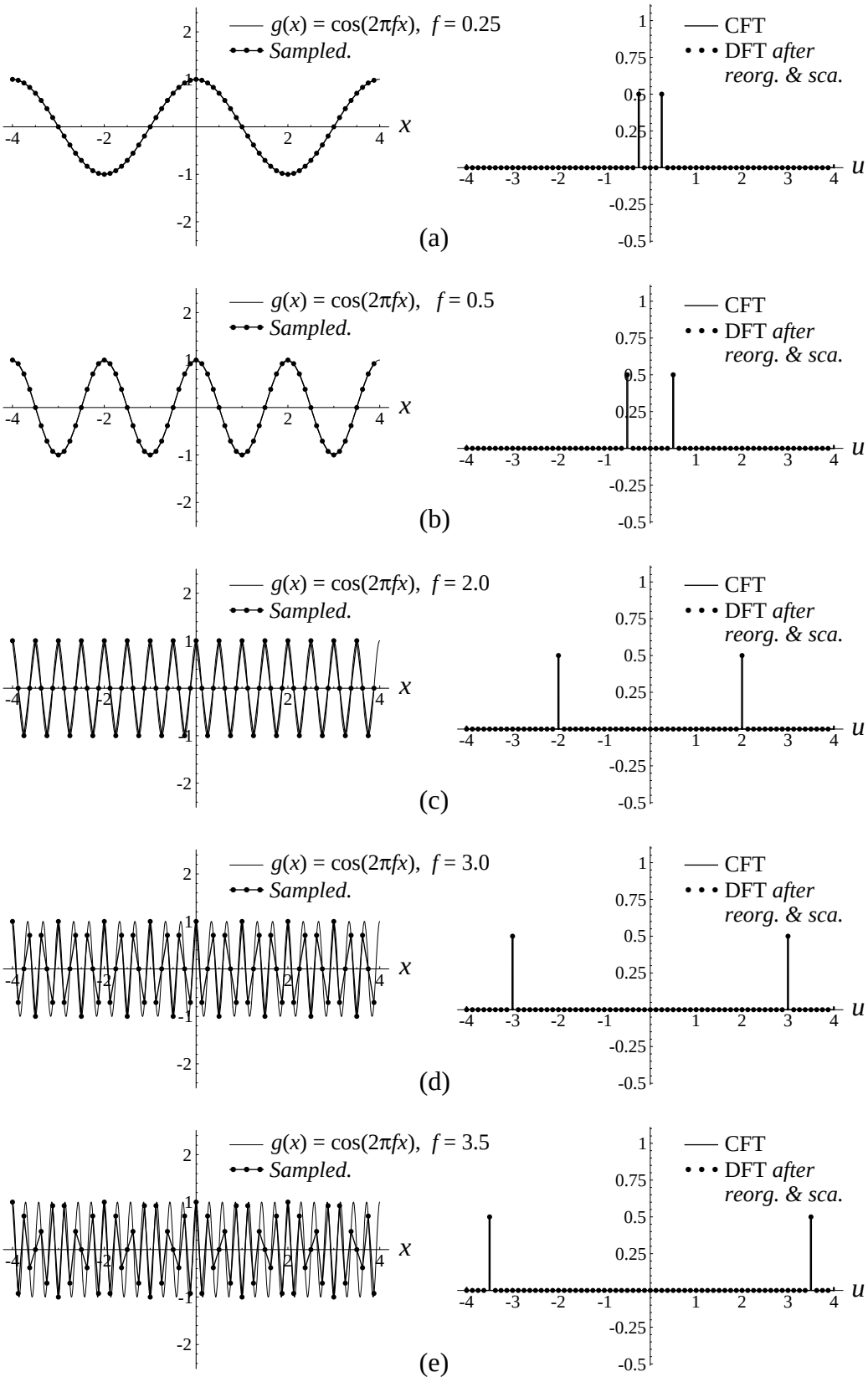
The material presented in this paper is situated on the border between sampling theory and the moiré theory, and it benefits from both points of view, which are often dual and complementary. Because this paper basically aims at a signal-processing audience,

which is not necessarily acquainted with the moiré theory, references are given whenever using concepts from the moiré theory. In addition, for the sake of completeness, Appendix A provides a brief review of the main moiré-theory notions and terms that are being used here.

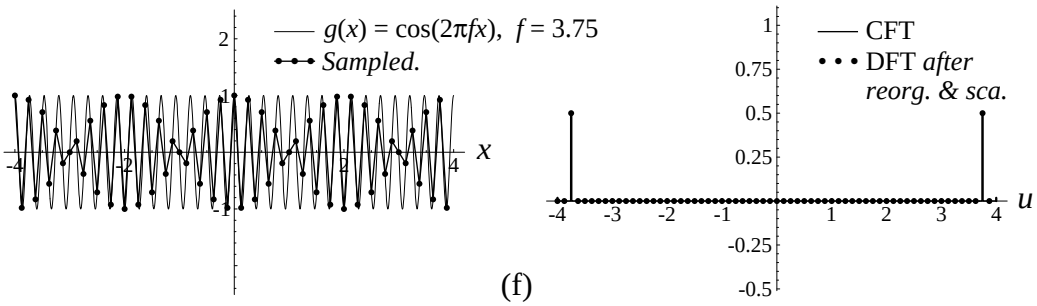
In addition to the figures which illustrate the discussions in the paper itself, two interactive Mathematica® applications are also provided in the supplementary material, along with their user's guide. Readers are encouraged to use these applications while reading this paper in order to better examine the cases under discussion, and for experimenting with any other cases they may wish to test. By manipulating the different parameters one may obtain a vivid graphic demonstration of the various sampling artifacts in question and of their dynamic behaviour, both in the signal domain and in the Fourier spectral domain. The source programs of these applications (which were also used to generate the figures of this paper) are provided, too. In addition, a glossary of the main terms which makes use of these applications for illustrating the various terms is also provided in the supplementary material.

**Figure 1:** Sampling moiré effects due to aliasing and sub-Nyquist artifacts, as they occur in the sampling process. These sampling artifacts are illustrated here using a series of cosine functions with gradually increasing frequencies. Each row shows in the left-hand column a continuous signal $g(x) = \cos(2\pi fx)$ having a different frequency $f$, as well as its sampled version after being sampled with a sampling frequency of $f_S = 8.0$ (i.e. with a sampling interval of $\Delta x = 1/f_S = 1/8$). For the sake of clarity, consecutive samples (dots) are connected by straight line segments; the original, continuous cosines are drawn by thinner curves. The right-hand column shows the respective CFT (continuous Fourier transform) of the continuous signal $g(x)$, along with the DFT (discrete Fourier transform) of its sampled version, after having applied the required reorganizations and scaling. The only difference between the rows is in the frequency $f$ of the original signal $g(x)$: (a) $f = 0.25$; (b) $f = 0.5$; (c) $f = 2$; (d) $f = 3$; (e) $f = 3.5$; (f) $f = 3.75$; (g) $f = 4$, namely $f = \frac{1}{2}f_S$; (h) $f = 4.25$; (i) $f = 7.5$; (j) $f = 7.75$. In the first few rows no aliasing occurs, and the sampled signal looks more or less similar to the original signal. Row (g) shows the limit case where the cosine frequency $f$ equals half of the sampling frequency, $\frac{1}{2}f_S$ (which is also the maximum frequency that can be represented in the DFT spectrum). In rows (h)-(j) the cosine frequency $f$ already exceeds $\frac{1}{2}f_S$, meaning that aliasing occurs (note that the corresponding impulses exceed the boundaries of the DFT spectrum and re-enter from the opposite end). In particular, when the folded-over (aliased) impulses are sufficiently close to the spectrum origin (see rows (i),(j)) a strong low-frequency sampling moiré becomes more prominent and visible than the original cosine function itself. Note, however, that the low-frequency beating effects which appear in the signal domain in rows (e) or (f) are not true moirés: They occur when the frequency $f$ of the cosine signal being sampled is below the Nyquist limit $\frac{1}{2}f_S$, so that no aliasing may occur; and furthermore, no corresponding low-frequency impulses appear in the respective spectra. These beating effects are examples of sub-Nyquist artifacts. The case shown in row (h) will be explained in Remark 8.

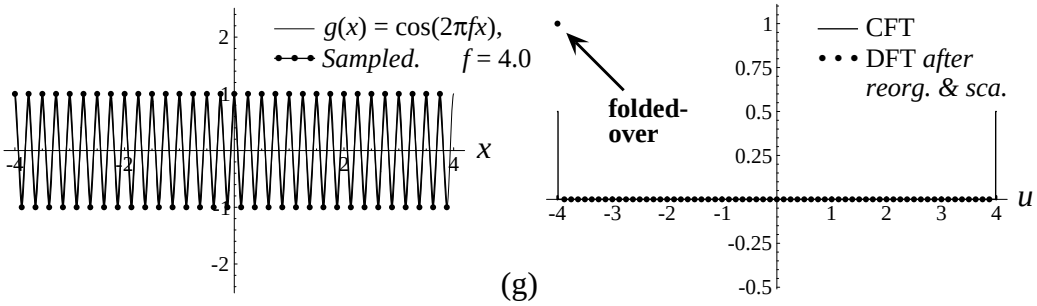
Signal domain Spectral domain



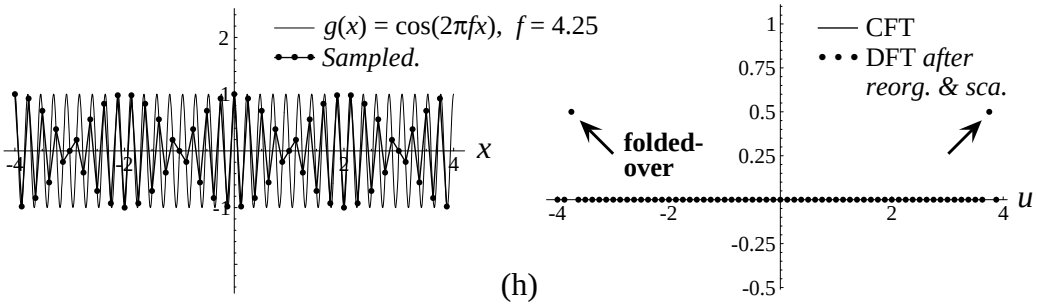
Signal domain Spectral domain



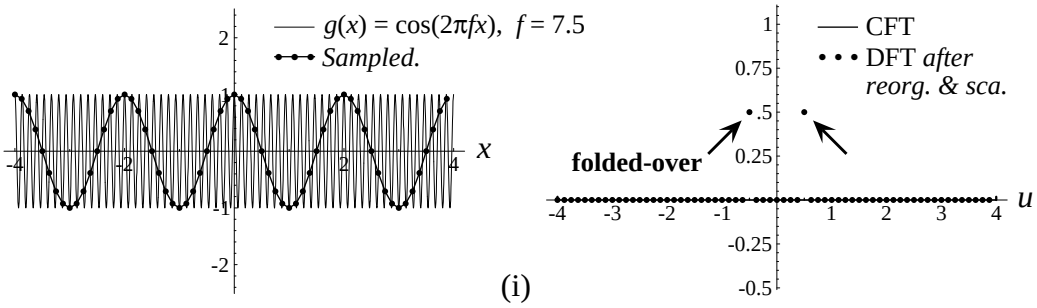

(f)



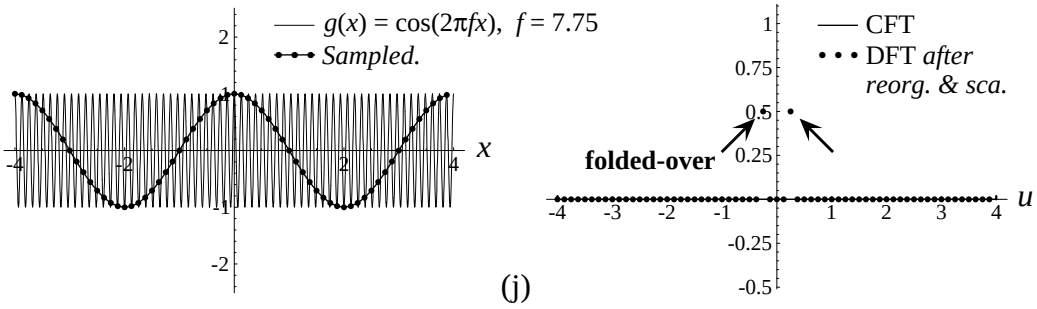

(g)




(h)




(i)




(j)

## 2. Background and basic notions

Let $g(x)$ be a continuous signal, and let $f_{\max}$ denote its maximum frequency (finite or infinite, depending on $g(x)$). As we know from the classical sampling theory, when $g(x)$ is sampled at a frequency $f_s$ which does not satisfy the Nyquist criterion, i.e. when $f_s < 2f_{\max}$ (and hence $f_{\max} > \frac{1}{2}f_s$), the signal's frequencies above $\frac{1}{2}f_s$ (and below $-\frac{1}{2}f_s$) fold-over into the frequency range $-\frac{1}{2}f_s...\frac{1}{2}f_s$. In other words, the resulting sampled signal $g(x_k)$ contains new false frequencies within this range, which do not exist in the original signal $g(x)$. This phenomenon is known as *aliasing*.

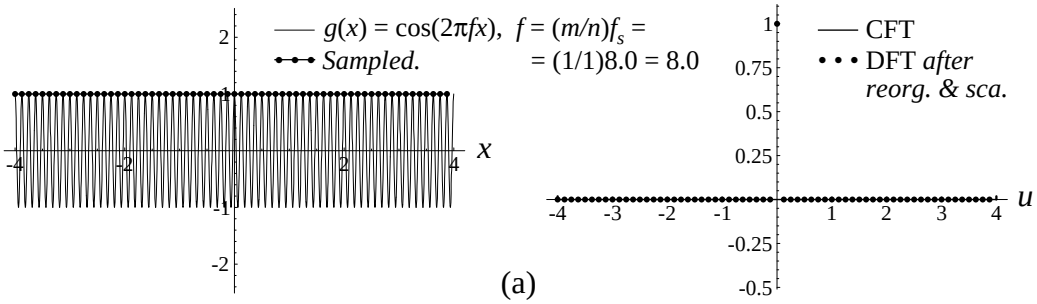
### 2.1 *Sampling moiré effects*

Aliasing artifacts may occur in the sampled signal whether $g(x)$ is periodic or not. But when $g(x)$ is periodic, so that its spectrum is impulsive, aliasing artifacts may become even more spectacular, due to their periodicity. From the point of view of the moiré theory (see Appendix A), the new folded-over low frequencies that we may get when sampling a periodic signal $g(x)$ correspond to a *sampling moiré effect*.

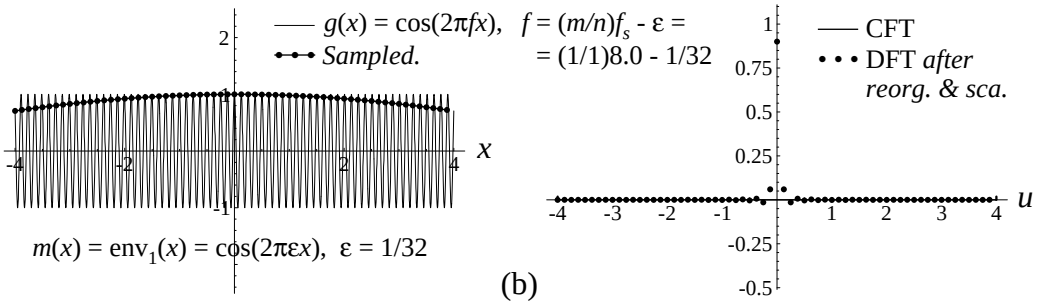
A sampling moiré effect is generated due to the interaction between the given periodic signal $g(x)$ and the periodic sampling process. It consists of a new false low-frequency $f_M$ which appears in the sampled signal $g(x_k)$ and in its spectrum, although it does not exist in the original signal $g(x)$. For example, suppose that $g(x)$ is the cosine signal $\cos(2\pi fx)$, whose spectrum consists of a single impulse pair located at $f$ and $-f$. When the frequency $f$ of our given signal is close to $f_s$, i.e. when $f_s - f \approx 0$, aliasing occurs, and the folded-over frequency falls at the point $f_M = f_s - f$, i.e. very close to the spectrum origin. The new false low frequency $f_M$ we therefore see in the sampled signal is a sampling-moiré effect. This moiré effect is shown in Fig. 2 for some values of $f$ close to $f_s$ (note that in all our figures we always use $f_s = 8$). As a second example, suppose that the frequency $f$ of our given cosine signal is close to $2f_s$, i.e. that $2f_s - f \approx 0$. Here, too, a false low frequency appears in the sampled signal, but this time due to an interaction between $f$ and $2f_s$. The folded-over frequency in this case is $f_M = 2f_s - f$, which is again a

**Figure 2:** The artifact that occurs when $f \approx (1/1)f_s$ (which is, in fact, a true first-order sampling moiré; see Figs. 1(i),(j)). Each row shows in the left-hand column the periodic signal $g(x) = \cos(2\pi fx)$ having frequency $f$, as well as its sampled version after being sampled with a sampling frequency of $f_s = 8.0$ (i.e. with a sampling interval of $\Delta x = 1/f_s = 1/8$). The right-hand column shows the respective CFT of the continuous signal $g(x)$, along with the DFT of its sampled version (after having applied the required reorganizations and scalings). The only difference between the 5 rows is in the frequency $f$ of the original signal $g(x)$: (a) $f = f_s$ (the singular state). (b) $f = f_s - 1/32$. (c) $f = f_s - 1/16$. (d) $f = f_s - 1/8$. (e) $f = f_s - 1/4$. The highly visible (1/1)-order artifact is generated because the sampled points $g(x_k)$ fall along a single low-frequency curve, which is simply a stretched version of $g(x)$.

Signal domain $\frac{m}{n} = \frac{1}{1}$ Spectral domain

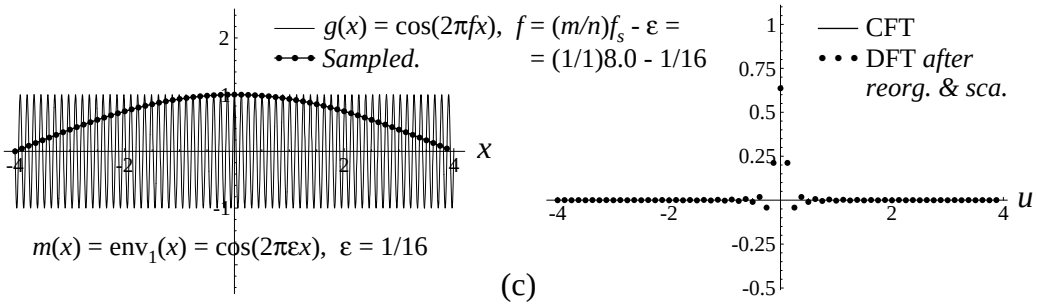



(a)



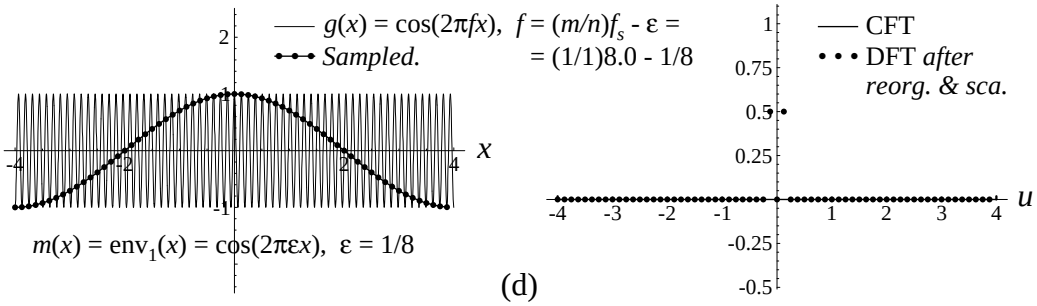

$m(x) = \text{env}_1(x) = \cos(2\pi\varepsilon x)$, $\varepsilon = 1/32$

(b)



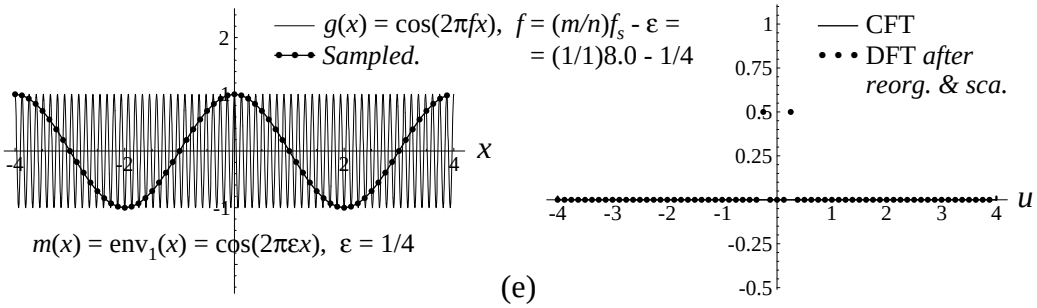

$m(x) = \text{env}_1(x) = \cos(2\pi\varepsilon x)$, $\varepsilon = 1/16$

(c)




$m(x) = \text{env}_1(x) = \cos(2\pi\varepsilon x)$, $\varepsilon = 1/8$

(d)




$m(x) = \text{env}_1(x) = \cos(2\pi\varepsilon x)$, $\varepsilon = 1/4$

(e)

very low frequency, located close to the spectrum origin. This effect, which is called a second-order sampling moiré, is shown in Fig. 3 for some values of $f$ close to $2f_s = 16$. It is important to note that this moiré effect is not the same as that of Fig. 2: When we gradually let $f$ move away from $2f_s$, as shown in the successive rows of Fig. 3 (and even further beyond that), the frequency of this second-order moiré gradually increases, until it finally becomes hardly visible. At some point another sampling moiré, say, the first-order moiré which is generated when $f$ gets closer to $f_s$, gradually starts gaining visibility, until finally it takes the upper hand and completely dominates the scene, when $f \approx f_s$ (Fig. 2).[3]

Similarly, a sampling moiré effect occurs in the sampled signal whenever $f$ approaches an integer multiple of $f_s$, i.e whenever $mf_s - f \approx 0$, $m = 1,2,\ldots$ In each of these cases a new low frequency $f_M = mf_s - f$ is generated in the spectrum of the sampled signal, close to the spectrum origin. Note that a sampling moiré effect is clearly visible both in the signal domain (as a false low-frequency signal) and in the spectral domain (as a folded-over false low-frequency spike); see Figs. 2 and 3. This is a key characteristic property of the moiré effects.

### 2.2 *Sub-Nyquist artifacts*

On the other hand, the *beating artifacts* on which we focus in the present work are quite different from aliasing or sampling moiré effects. They are generated in the sampled signal when the frequencies $f$ and $f_s$ satisfy $f \approx (m/n)f_s$ (with integer $m,n$ and $n > 1$).[4] As we can see in Figs. 4, 5 or 6, these beating artifacts are intriguing for several reasons:

(a) They may appear where the Nyquist condition is fully satisfied, so that no aliasing or sampling moiré artifacts should be present. For example, in Figs. 5 or 6 the frequency of the original function $g(x) = \cos(2\pi fx)$ is close to $\frac{1}{4}f_s$ or $\frac{1}{3}f_s$, respectively, i.e. far below half of the sampling frequency.

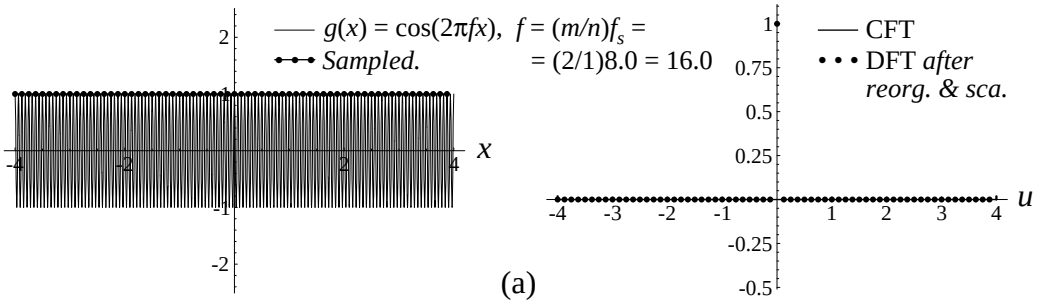
**Figure 3:** The artifact that occurs when $f \approx (2/1)f_s$ (which is, in fact, a true second-order sampling moiré). This figure is similar to Fig. 2, except for the signal-frequency $f$ being used in each row: (a) $f = 2f_s$ (the singular state). (b) $f = 2f_s - 1/32$. (c) $f = 2f_s - 1/16$. (d) $f = 2f_s - 1/8$. (e) $f = 2f_s - 1/4$. The highly visible (2/1)-order artifact is generated because the sampled points $g(x_k)$ fall along a single low-frequency curve, which is simply a stretched version of $g(x)$.

---

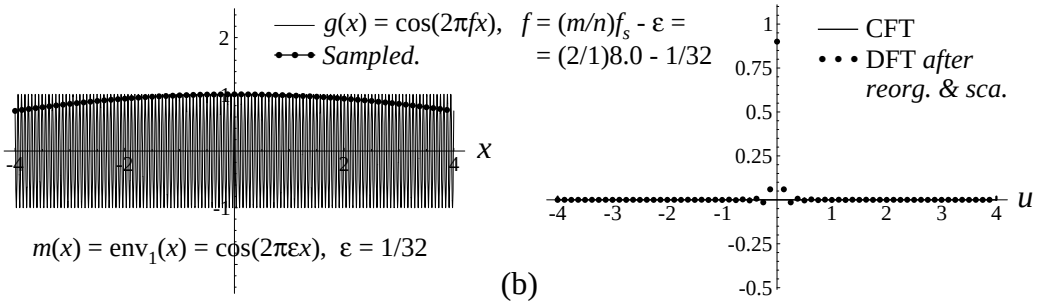
[3] Theoretically, a moiré effect "exists" even when its frequency gets far away from the spectrum origin; but practically we may (and do) ignore it when it is no longer visible and pertinent.

[4] Cases with $n = 1$ are simply *sampling moiré effects*; for example, the two sampling moiré effects discussed above correspond, respectively, to $m = n = 1$ and to $m = 2$, $n = 1$. We will return to this point later on, after having discussed the general $m/n$ case.

Signal domain $\boxed{\frac{m}{n}=\frac{2}{1}}$ Spectral domain

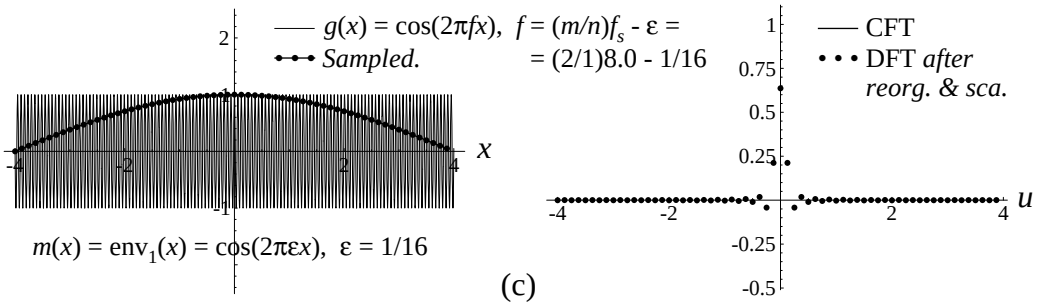



(a)



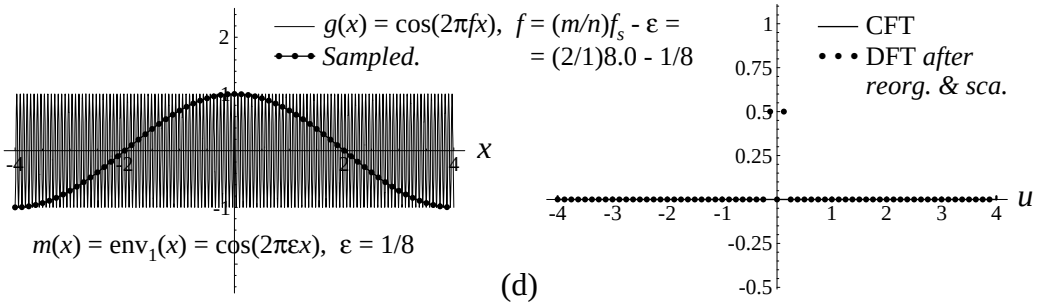

$m(x) = \text{env}_1(x) = \cos(2\pi\varepsilon x),\ \varepsilon = 1/32$

(b)



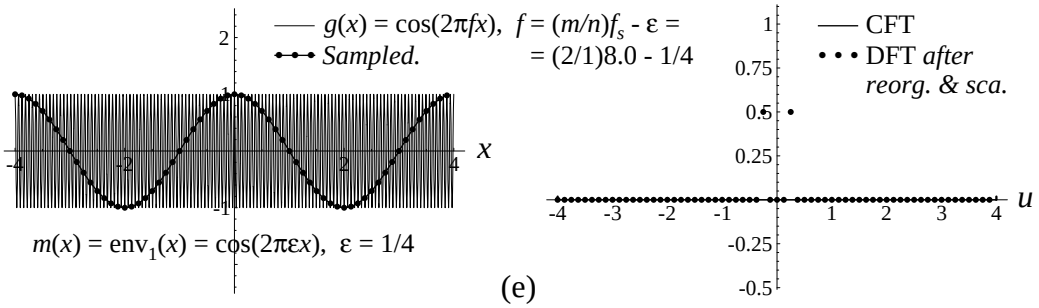

$m(x) = \text{env}_1(x) = \cos(2\pi\varepsilon x),\ \varepsilon = 1/16$

(c)




$m(x) = \text{env}_1(x) = \cos(2\pi\varepsilon x),\ \varepsilon = 1/8$

(d)




$m(x) = \text{env}_1(x) = \cos(2\pi\varepsilon x),\ \varepsilon = 1/4$

(e)

(b) Unlike in aliasing or moiré phenomena, the periods (or frequencies) of these beating artifacts are not represented in the Fourier spectrum, although they are clearly visible in the sampled signal. To see this, compare the spectra of our beating artifacts in any of Figs. 4-6 with the spectra of the sampling moiré effects of Figs. 2-3: While in Figs. 2-3 the low frequency we see in the signal domain is also represented in the spectral domain as a folded-over frequency, in Figs. 4-6 the low-frequency structures we see in the signal domain *do not* appear in the spectrum.[5]

(c) Furthermore, in the signal domain, the beating effect in question does not really correspond to a smooth low-frequency signal, but rather to a highly oscillating signal that is only *modulated* by low-frequency envelopes (compare the beating effect in Figs. 4-6 with the true sampling moiré effect in Figs. 2-3). Note that this may provide a clue to the understanding of point (b): The low frequencies we see in the signal domain belong to the *modulating envelopes* of the sampled signal, but they are not included in the sampled signal itself.

We call these beating effects *sub-Nyquist artifacts*, reflecting the fact that they may occur well below the Nyquist frequency. But as we will see later, these artifacts may also occur above the Nyquist frequency (see, for example, Fig. 1(h) and Remark 8).

Obviously, the first approach that comes to one's mind for investigating such phenomena is based on the Fourier theory, the main investigation tool being used in the moiré theory and in signal processing. But the fact that our beating artifacts are not visible in the spectrum of the sampled signal (see point (b) above) makes them more difficult to analyze.

In order to explore these "hidden" cases, we propose the following approach: We let the frequency $f$ of the continuous signal $g(x)$ sweep along the frequency axis, just as we do while running from row to row in Fig. 1, and we see what happens when the signal frequency $f$ approaches critical points such as multiples or integer fractions of the sampling frequency $f_s$. Some typical cases will be described in Sec. 3, in order to develop a better understanding of these phenomena and of their properties. Based on the insights gained in Sec. 3 we will proceed in Secs. 4-5 to the mathematical analysis of these phenomena.

Sweeping along the frequency axis reveals in the "hidden" areas, i.e. those that are not detected by classical Fourier methods, the same typical landscape as in the "visible" areas populated by true moiré effects: Each time our signal frequency $f$ approaches some critical frequency $(m/n)f_s$, where $f_s$ is the sampling frequency and $m$, $n$ are integers, we get in the sampled signal a visible artifact. The period of this artifact becomes infinitely large when $f$ reaches precisely the critical frequency. But then, as $f$ gradually moves away from this critical frequency (to either direction), the period becomes smaller and smaller until finally the artifact fades out and disappears.

---

[5] Of course, such frequencies can be made visible in the spectrum by applying to the beating signal in question a nonlinear operation such as envelope detection (rectification possibly followed by low-pass filtering [8, p. 216]). This practical "trick" can be used in applications to bypass the problem, but it cannot help us in the theoretical explanation of these phenomena.

For the sake of simplicity we start our exploration using the simple case of the cosine function; later on we will generalize our results to any general periodic signals.

## 3. Sampling a continuous cosinusoidal signal

Suppose we are given the continuous signal $g(x) = \cos(2\pi fx)$, i.e. a cosine function with frequency $f$, and that we sample it using the sampling frequency $f_s$. We denote the resulting sampled signal by $g(x_k) = \cos(2\pi fx_k)$, where $x_k = k\Delta x$, and $\Delta x = 1/f_s$ is the sampling step. How does the sampled signal behave when we gradually vary the cosine frequency $f$ or the sampling frequency $f_s$? Obviously, varying both $f$ and $f_s$ simultaneously in the same proportion (i.e. multiplying both of them by the same factor) simply scales $g(x)$ and its sampled version $g(x_k)$ together along the $x$ axis, and gives nothing new. Therefore, we can keep one of the frequencies $f$ or $f_s$ fixed, and explore what happens when we vary the other frequency alone. Both approaches (varying $f$ or varying $f_s$) are in fact equivalent, but for the sake of simplicity (technical considerations when applying DFT to our sampled signal) we prefer to keep the sampling frequency $f_s$ fixed, and to study what happens when we gradually vary the frequency $f$ of our given continuous function (as in Fig. 1). Note that in all our figures we always use the fixed sampling frequency $f_s = 8$; this choice is convenient for the DFT being used in our figures, but in principle, any other $f_s$ value could also be chosen, be it integer or not.

Let us now start varying the frequency $f$ along the frequency axis, and stop to explore a few typical examples of artifacts that we encounter on our way. Interested readers may do this exercise on their own using the provided interactive applications, and dynamically watch the various artifacts as they gradually appear and disappear while $f$ is slowly being varied.[6]

**Example 1** (*sampling moiré effect*):

Let us first explore what happens in the frequency range around $f \approx f_s$. This is shown in Fig. 1(i),(j), and in greater detail in Fig. 2. In this case we have:

$$f = f_s - \varepsilon \tag{1}$$

where $\varepsilon$ denotes a small positive value.[7] Since this case does not respect the Nyquist condition $f \leq \frac{1}{2}f_s$ (or $f_s \geq 2f$) required by the sampling theorem, aliasing occurs here. And indeed, looking at Fig. 2 we see that our sampled points fall on a low-frequency cosinusoidal curve whose frequency is precisely $\varepsilon$:

$$m(x) = \cos(2\pi\varepsilon x) \tag{2}$$

This is, in fact, a one-dimensional sampling moiré effect, which simply obeys the basic rules of the moiré theory (see, for example, [9, Sec. 4.2]):

---

[6] To slow down the movement of the "$f$" slider in the applications, hold down the Alt or Option key while dragging the slider with the mouse. The movement can be further slowed down by also holding down the Shift or Ctrl key, or both (see the user's guide in the supplementary material).

[7] The results obtained when taking $f = f_s + \varepsilon$ are similar, so we only need to consider one of the two cases; we will return to this point in greater detail later on, in Remark 3.

(a) Its frequency is $\varepsilon = f_s - f$;

(b) Its intensity profile $m(x)$ is an expanded version of the original continuous function $g(x)$, that is simply stretched (or magnified) along the $x$ axis by an expansion (magnification) rate of $f/\varepsilon$, in order to satisfy (a): [8]

$$m(x) = g(\tfrac{\varepsilon}{f}x)$$

This moiré effect is known in the moiré theory as a (1,-1)-moiré because it occurs when $mf_s + nf \approx 0$ with $m = 1$ and $n = -1$, i.e. $f \approx f_s$ (Appendix A provides a brief review of the main moiré-theory terms being used here). As we can see in row (a) of Fig. 2, when $f$ precisely equals $f_s$ (i.e. when $\varepsilon = 0$) we get the *singular state* of the (1,-1)-sampling moiré effect, where the moiré period $1/\varepsilon$ is infinitely large and therefore invisible. But when $f$ slightly moves away from the frequency $f_s$ (to either direction) the moiré "comes back from infinity", as shown in row (b), and becomes again visible with a long period $1/\varepsilon$ (small frequency $\varepsilon$). As $f$ moves farther away from the singular state $f = f_s$, i.e. as $|\varepsilon|$ gradually increases (see rows (c)-(e) in our figure), the period of the moiré effect becomes smaller and smaller, until it can no longer be detected. ■

**Example 2** (*sampling moiré effect*):

A similar phenomenon also happens when we let $f$ explore the frequency range around $f \approx 2f_s$. In this case the sampling moiré effect we obtain is the second-order (2,-1)-moiré, where our given cosine function has about two periods between consecutive sampling points (since $f \approx 2f_s$ means $1/f \approx 1/(2f_s)$ i.e. $p \approx \frac{1}{2}\Delta x$). In this case we have:

$$f = 2f_s - \varepsilon \tag{3}$$

As we can see in Fig. 3, here too, when $f$ precisely equals $2f_s$ (i.e. when $\varepsilon = 0$) we get the singular state of the moiré in question, where the moiré period $1/\varepsilon$ is infinitely large and therefore invisible (see row (a) of the figure). But when $f$ moves away from the singular frequency $2f_s$ (to either direction), i.e. as $|\varepsilon|$ gradually increases (see rows (b)-(e) in our figure), the period of the moiré effect becomes smaller and smaller, until it finally disappears. This is, indeed, the typical behaviour of any moiré effect in the neighbourhood of its singular frequency. Furthermore, looking at Fig. 3, we see that in the present case, too, all the sampled points $g(x_k)$ fall again on the same low-frequency cosinusoidal curve $m(x)$ as in the previous case:

$$m(x) = \cos(2\pi\varepsilon x) \tag{4}$$

Note that in terms of the sampling theory, this moiré effect, too, is simply a foldover artifact due to aliasing. The reason is that in both cases (Eqs. (1) and (3)) we do not respect the Nyquist condition (see, for instance, Example 5.4 in [10, Sec. 5.3]). ■

---

[8] According to a moiré theory result (see, for example, [9, Sec. 4.2 or Sec. 10.9]), the profile of this moiré is determined by the cyclic convolution of the profiles of the two periodic structures involved, the sampling comb III($x$) and the original continuous function $g(x)$, both being normalized to the same period. The profile of the resulting moiré is therefore equal to that of $g(x)$, since a cyclic convolution of $g(x)$ with an impulse comb gives back $g(x)$. This resulting profile is then stretched according to point (a).

Similar results are also obtained when we let $f$ explore the frequency range around $f \approx mf_s$ for any integer $m > 1$, giving an $m$-th order ($m$,-1)-sampling moiré. And indeed, when

$$f = mf_s - \varepsilon \tag{5}$$

our sampled points $g(x_k)$ will reside, once again, on the same low-frequency cosinusoidal curve (4) having the frequency $\varepsilon$.

So far we have seen what happens when the frequency $f$ of our continuous cosine $g(x)$ approaches integer multiples of the sampling frequency $f_s$. Let us now continue varying the frequency $f$, and study what happens when we arrive to fractional multiples of $f_s$.

**Example 3** (*sub-Nyquist artifact*):

Let us focus on the frequency range around $f \approx \frac{1}{2}f_s$ (i.e. $2f \approx f_s$), as shown in Fig. 1(e),(f) or in greater detail in Fig. 4, i.e. when:

$$f = \tfrac{1}{2}f_s - \varepsilon \tag{6}$$

Eq. (6) implies that we are located here in the "safe" side of the Nyquist frequency $\frac{1}{2}f_s$, meaning that the sampling condition stipulated by the sampling theorem, $f \leq \frac{1}{2}f_s$ (or $f_s \geq 2f$), is fully satisfied, and no aliasing or moiré effects should be present. And yet, looking at Fig. 4, we see that a low-frequency beating effect is visible in the sampled signal, due to a low-frequency cosinusoidal modulation. At the singular frequency itself, $f = \frac{1}{2}f_s$ (see row (a) of our figure), the modulating envelope of the beating effect has an infinitely long period, and is therefore invisible. But as $f$ moves away from this singular frequency, i.e. as $|\varepsilon|$ gradually increases (see rows (b)-(e) in our figure), the period of the beating effect becomes smaller and smaller, until it gradually becomes invisible and disappears. As we have seen above, this behaviour is typical to a moiré effect. However, our present beating effect is not a true sampling moiré but rather a *pseudo* moiré effect [11], or using our present terminology, a sub-Nyquist artifact: As we can see in Fig. 4, in our case no corresponding low-frequency impulses exist in the respective spectra (compare the low-frequency zone around the origin in the DFT spectra of Fig. 4 with the same zone in Figs. 2 and 3, where a true aliased sampling moiré does exist). Furthermore, even in terms of the signal domain, the beating effect in question does not really represent a low frequency signal (as in Figs. 2 and 3), but rather a highly oscillating signal that is only *modulated* by two interlaced low-frequency cosinusoidal envelopes. These two envelopes, which are highlighted in Fig. 4 by different colours, are expressed by:

$$\begin{aligned} \mathrm{env}_1(x) &= \cos(2\pi\varepsilon x) \\ \mathrm{env}_2(x) &= \cos(2\pi\varepsilon[x+a]) \end{aligned} \tag{7}$$

where the envelope frequency is $\varepsilon = \frac{1}{2}f_s - f$, and the shift $a$ equals half of the period $1/\varepsilon$ of the cosinusoidal envelope, i.e. $a = 1/(2\varepsilon)$. Because the frequency $\varepsilon$ is much lower than the frequency $f$ of the original cosine function $g(x)$ being sampled, this artifact may become quite conspicuous and distort our perception of the true nature of the original

signal. Interested readers may find a discussion on this particular sub-Nyquist artifact, both in the one- and two-dimensional cases, in [10, Sec. 8.6]. ■

The beating effect we have just described occurs when the frequency $f$ of the cosine function being sampled is located slightly below half of the sampling frequency, i.e. almost at the border of the "safe" frequency zone stipulated by the sampling theorem. This case could be therefore dismissed as a "borderline" artifact that occurs when we get too close to the limits of the theorem. However, it turns out that such sub-Nyquist artifacts may also appear when the cosine frequency $f$ is well below $\frac{1}{2}f_s$, as clearly illustrated by the following example.

**Example 4** (*sub-Nyquist artifact*):

Consider Fig. 5, which shows what happens when we arrive to the frequency range around $f \approx \frac{1}{4}f_s$ (i.e. $4f \approx f_s$), and in particular when:

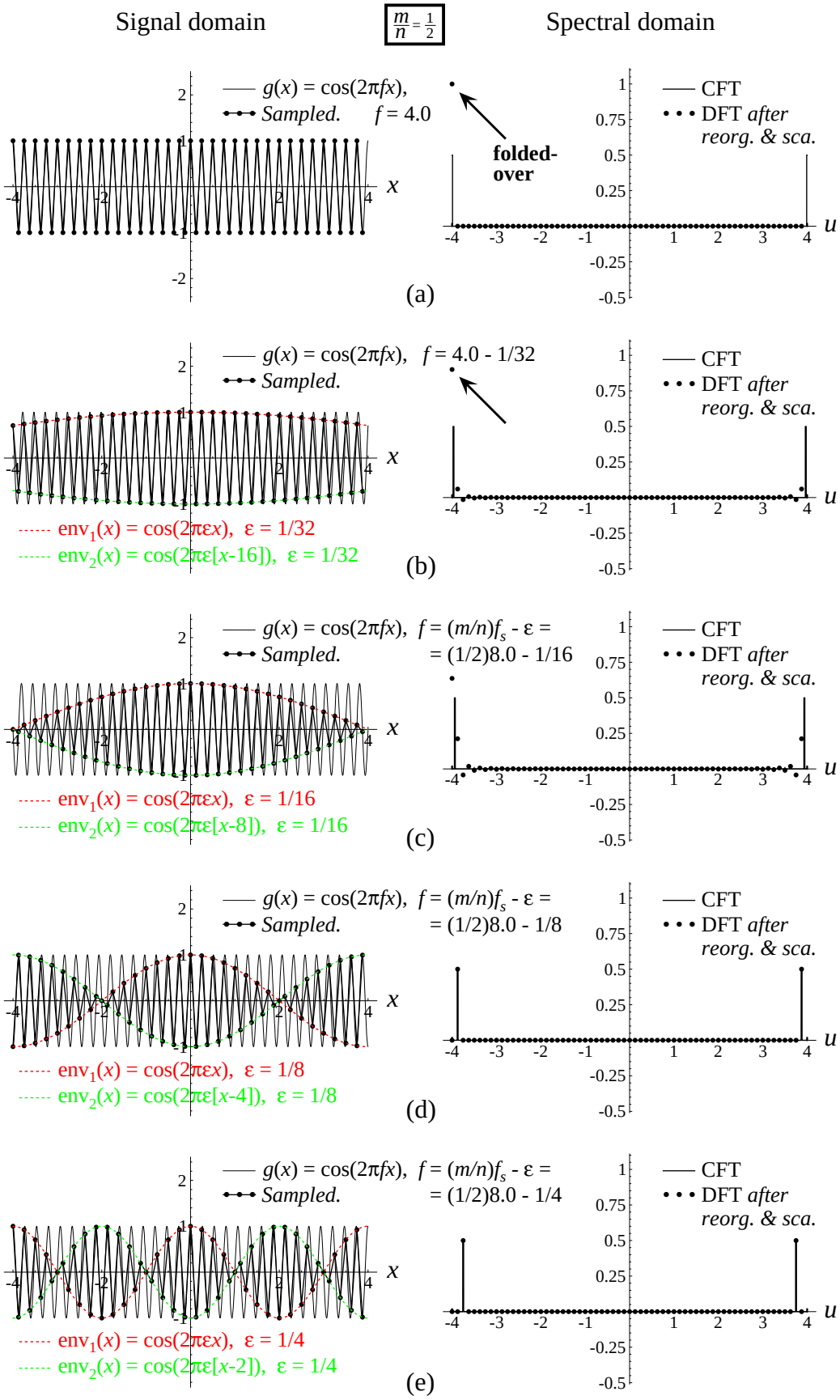
$$f = \tfrac{1}{4}f_s - \varepsilon \tag{8}$$

Row (a) of this figure shows the situation when the cosine frequency is exactly $f = \frac{1}{4}f_s = 2$, and rows (b)-(e) of the figure show what happens when the cosine frequency $f$ gradually moves away from this singular frequency, i.e. when $|\varepsilon|$ is gradually increased. As we can see, the resulting visual effect in the signal domain resembles the modulation effect we obtained in Fig. 4, where the cosine frequency $f$ was close to $\frac{1}{2}f_s = 4$, except that in Fig. 5 the beating effect no longer corresponds to *two* interlaced cosinusoidal envelopes, but rather to *four* interlaced cosinusoidal envelopes; this confers to the sampled signal a typical wavy appearance (ripple). The four interlaced envelopes obtained in this case, which are highlighted in Fig. 5 by four different colours or line styles, are expressed by:

$$\begin{aligned} \mathrm{env}_1(x) &= \cos(2\pi\varepsilon x) \\ \mathrm{env}_2(x) &= \cos(2\pi\varepsilon[x+a]) \\ \mathrm{env}_3(x) &= \cos(2\pi\varepsilon[x+2a]) \\ \mathrm{env}_4(x) &= \cos(2\pi\varepsilon[x+3a]) \end{aligned} \tag{9}$$

where the envelope frequency is $\varepsilon = = \frac{1}{4}f_s - f$, and the shift $a$ between adjacent envelopes equals 1/4 of the period $1/\varepsilon$ of the cosinusoidal envelope, i.e. $a = 1/(4\varepsilon)$. We will see the

**Figure 4:** The (1/2)-order sub-Nyquist artifact; see Figs. 1(e)-(h). This figure is similar to Figs. 2-3, except for the signal-frequency $f$ being used in each row: (a) $f = \frac{1}{2}f_s$ (the singular state). (b) $f = \frac{1}{2}f_s - 1/32$. (c) $f = \frac{1}{2}f_s - 1/16$. (d) $f = \frac{1}{2}f_s - 1/8$. (e) $f = \frac{1}{2}f_s - 1/4$. The highly visible (1/2)-order sub-Nyquist artifact is generated because consecutive points $g(x_k)$ of the sampled signal alternately jump from one of the $n = 2$ modulating envelopes to the other (each of the two modulating envelopes being simply a stretched and shifted version of $g(x)$). Note that these two interlaced modulating envelopes are highlighted in the figure in different colours.


Signal domain
m/n = 1/2
Spectral domain
g(x) = cos(2πfx),
Sampled.
f = 4.0
folded-over
CFT
DFT after reorg. & sca.
x
u
(a)
g(x) = cos(2πfx), f = 4.0 - 1/32
Sampled.
CFT
DFT after reorg. & sca.
env1(x) = cos(2πεx), ε = 1/32
env2(x) = cos(2πε[x-16]), ε = 1/32
(b)
g(x) = cos(2πfx), f = (m/n)fs - ε =
= (1/2)8.0 - 1/16
Sampled.
CFT
DFT after reorg. & sca.
env1(x) = cos(2πεx), ε = 1/16
env2(x) = cos(2πε[x-8]), ε = 1/16
(c)
g(x) = cos(2πfx), f = (m/n)fs - ε =
= (1/2)8.0 - 1/8
Sampled.
CFT
DFT after reorg. & sca.
env1(x) = cos(2πεx), ε = 1/8
env2(x) = cos(2πε[x-4]), ε = 1/8
(d)
g(x) = cos(2πfx), f = (m/n)fs - ε =
= (1/2)8.0 - 1/4
Sampled.
CFT
DFT after reorg. & sca.
env1(x) = cos(2πεx), ε = 1/4
env2(x) = cos(2πε[x-2]), ε = 1/4
(e)

formal derivation of these expressions in Sec. 4. Note that in this case, too, since the frequency $\varepsilon$ is much lower than the frequency $f$ of the original signal $g(x)$ being sampled, this artifact may become quite conspicuous. But once again, this beating effect in the sampled signal is misleading, since it is only a sampling artifact and it does not reflect the true behaviour of our original continuous-world signal $g(x)$. Furthermore, its low frequency is not represented in the DFT spectra, meaning that it is not a true sampling moiré effect. ■

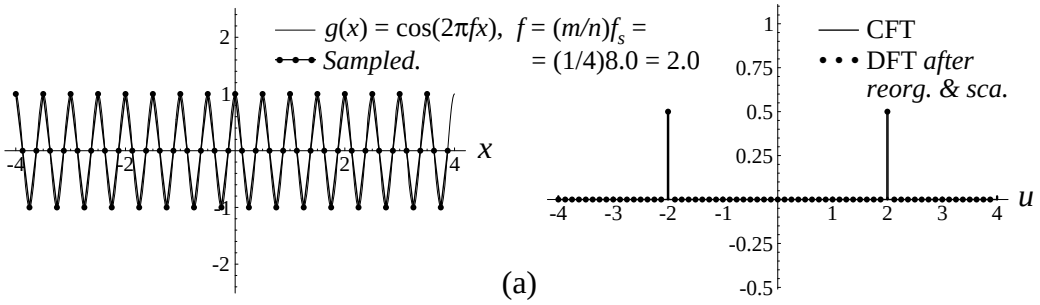
Pursuing our sweep along the frequency axis it turns out that similar sub-Nyquist artifacts may occur in many other cases, too, even well below the Nyquist frequency limit set up by the sampling theorem. This happens whenever our cosine frequency $f$ is located around a rational fraction $m/n$ of the sampling frequency $f_s$, i.e. whenever $f \approx (m/n)f_s$. We will henceforth call this phenomenon the ($m/n$)-*order sub-Nyquist artifact*. This artifact manifests itself in a curious way — it forces all the sampled values $g(x_k)$ of our given continuous signal $g(x)$ to fall on $n$ interlaced modulating envelopes:

$$\begin{aligned} \mathrm{env}_1(x) &= \cos(2\pi\varepsilon x) \\ \mathrm{env}_2(x) &= \cos(2\pi\varepsilon[x+a]) \\ &\ldots \\ \mathrm{env}_n(x) &= \cos(2\pi\varepsilon[x+(n-1)a]) \end{aligned} \tag{10}$$

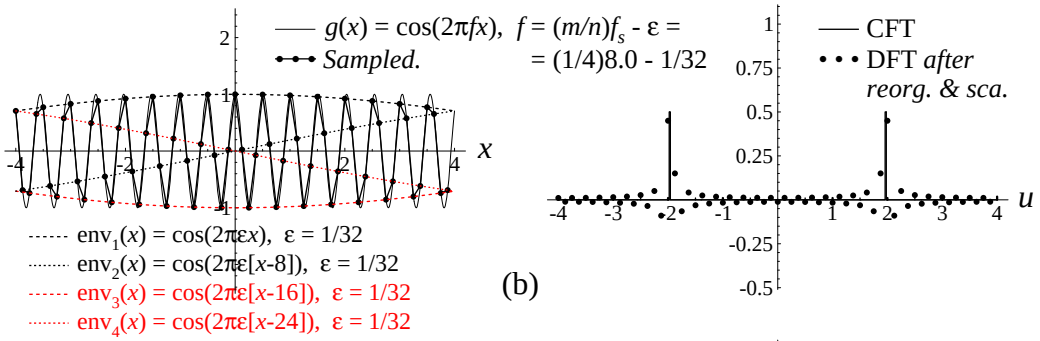
where the frequency of each envelope is $\varepsilon = (m/n)f_s - f$, and the shift $a$ between adjacent envelopes equals $m/n$-th of the cosinusoidal envelope's period $1/\varepsilon$, i.e. $a = m/(n\varepsilon)$. (By "rational fraction" we actually mean here a reduced integer fraction, i.e. a ratio of two mutually prime integers $m$ and $n$; for example, the (2/4)-order sub-Nyquist artifact is identical to the (1/2)-order sub-Nyquist artifact and should not be considered as a distinct case.) Note that in any ($m/n$)-order sub-Nyquist artifact the frequency $f_{\mathrm{env}} = \varepsilon$ (and period $p_{\mathrm{env}} = 1/\varepsilon$) of the interlaced modulating envelopes only depend on the distance $\varepsilon$ of the frequency $f$ of the given continuous-world cosine from the corresponding critical frequency $(m/n)f_s$. For this reason, the envelope frequencies (and periods) in Figs. 2-6 are row-wise identical, since in all these figures we always use the same value of $\varepsilon$ in the same rows. The only difference between these figures is in the number of interlaced envelopes, and in the shift $a$ between each two adjacent envelopes.

**Figure 5:** The (1/4)-order sub-Nyquist artifact. This figure is similar to Figs. 2-4, except for the signal-frequency $f$ being used in each row: (a) $f = \frac{1}{4}f_s$ (the singular state). (b) $f = \frac{1}{4}f_s - 1/32$. (c) $f = \frac{1}{4}f_s - 1/16$. (d) $f = \frac{1}{4}f_s - 1/8$. (e) $f = \frac{1}{4}f_s - 1/4$. The highly visible (1/4)-order sub-Nyquist artifact is generated because consecutive points $g(x_k)$ of the sampled signal alternately jump from one of the $n = 4$ modulating envelopes to the others (each of the four modulating envelopes being simply a stretched and shifted version of $g(x)$). Note that these four interlaced modulating envelopes are highlighted in the figure in different colours or line styles.

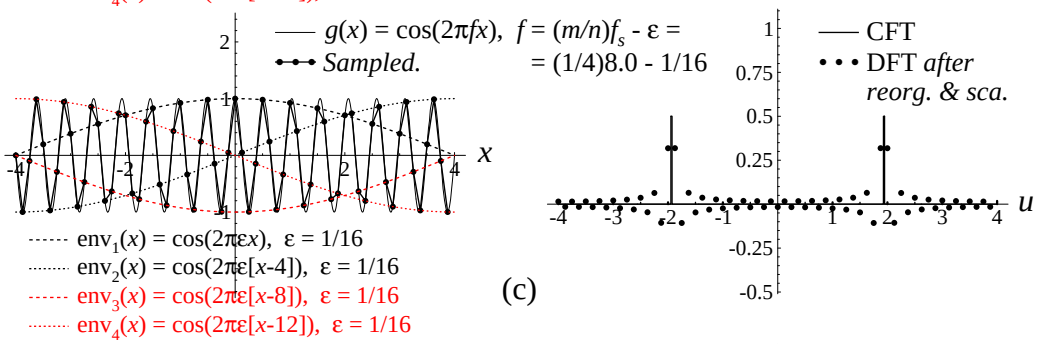
Signal domain $\boxed{\frac{m}{n} = \frac{1}{4}}$ Spectral domain



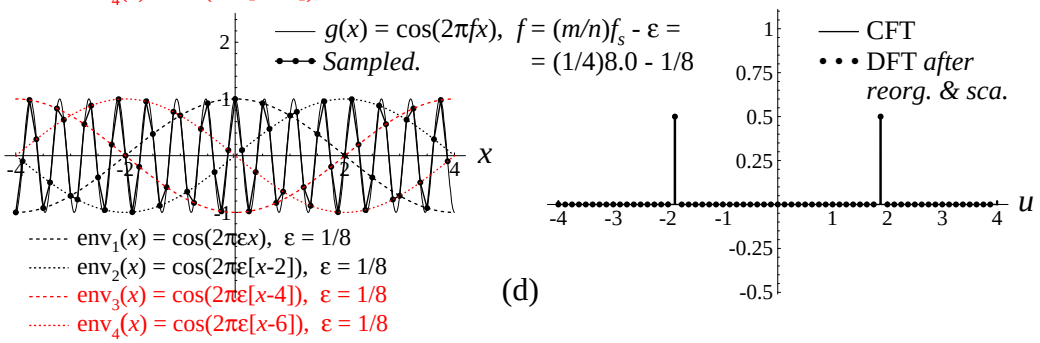

(a)



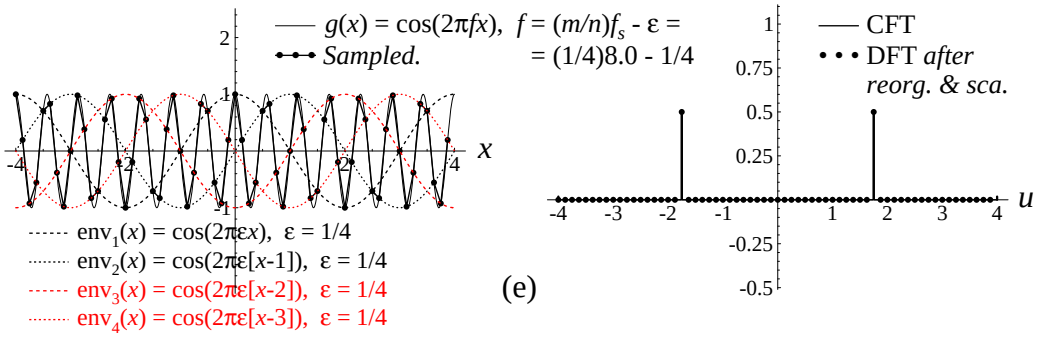

(b)




(c)




(d)




(e)

Interestingly, the true ($m$,-1) moiré cases shown in Figs. 2 and 3 also obey the very same rule, but this time with $n = 1$ envelopes, so that no interlacing occurs. In fact, according to our new notation, these true moiré cases can be considered as ($m$/1)-order sub-Nyquist artifacts. This confirms, once again, the common nature of sampling moiré effects and sub-Nyquist artifacts. (See Remark 10 in Appendix C on the relationship between the ($m$,-$n$) notation that is used for moiré effects and the ($m$/$n$) notation that we use for sub-Nyquist artifacts.)

As we can see, the ($m$/$n$)-order sub-Nyquist artifact is indeed a generalization of the "borderline" sub-Nyquist artifact that occurs when the frequency $f$ of our continuous cosine function is close to $(1/2)f_s$ (see Eq. (6) above). According to our present notations, this "borderline" case is simply the (1/2)-order sub-Nyquist artifact. Obviously, the most interesting ($m$/$n$)-order sub-Nyquist artifacts are those with $m < n/2$, for which $(m/n)f_s$ is lower than $(1/2)f_s$, i.e. below half of the sampling frequency (like the (1/3)-order, (2/5)-order, etc.). But other cases such as the (2/3)-order, (3/2)-order, (3/5)-order, etc. may also be considered; see Remark 8 in Appendix C.

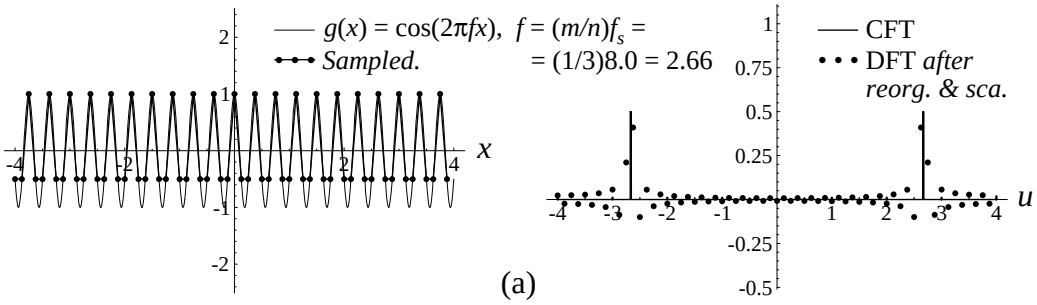
So how can we explain these artifacts, that are "hidden" to the Fourier spectra? Why do these visible phenomena occur although $f$ is located in the "safe" side of the Nyquist frequency $\frac{1}{2}f_s$, meaning that the sampling condition stipulated by the sampling theorem, $f \leq \frac{1}{2}f_s$, is fully satisfied, and no aliasing or moiré effects should be present?[9] We will see the answer in the following sections, first for the particular case of the cosine function $g(x) = \cos(2\pi fx)$, in Sec. 4, and then for the general case with any periodic function $g(x)$, in Sec. 5.

**Remark 1:** When the value of $n$ is odd, like in the (1/3)-order sub-Nyquist artifact, we get an odd number $n$ of interlaced envelopes (see Fig. 6). Consequently, the modulation effect we obtain here is not vertically symmetric as in Figs. 4 and 5 (its maxima and minima, i.e. its top ripples and bottom ripples, are not synchronized but rather intermittent). We will henceforth call the synchronized type of ripples "even ripples", and the non-synchronized type "odd ripples". ■

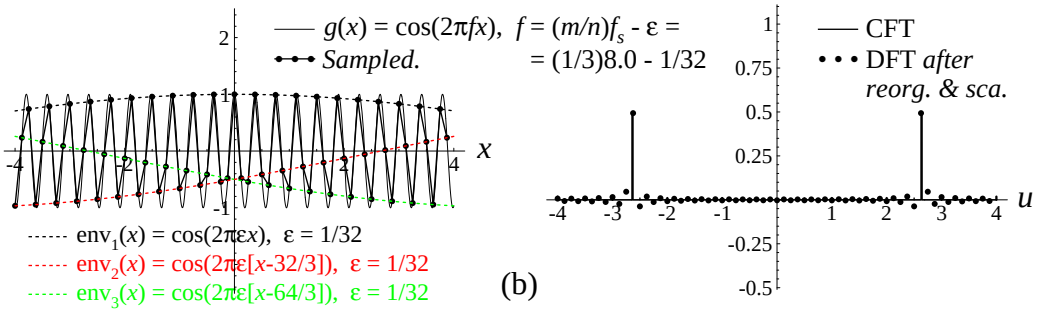
**Figure 6:** The (1/3)-order sub-Nyquist artifact (belonging to the odd type; see Remark 1). This figure is similar to Figs. 2-5, except for the signal-frequency $f$ being used in each row: (a) $f = \frac{1}{3}f_s$ (the singular state). (b) $f = \frac{1}{3}f_s - 1/32$. (c) $f = \frac{1}{3}f_s - 1/16$. (d) $f = \frac{1}{3}f_s - 1/8$. (e) $f = \frac{1}{3}f_s - 1/4$. The highly visible (1/3)-order sub-Nyquist artifact is generated because consecutive points $g(x_k)$ of the sampled signal alternately jump from one of the $n = 3$ modulating envelopes to the others (each of the three modulating envelopes being simply a stretched and shifted version of $g(x)$). Note that these three interlaced modulating envelopes are highlighted in the figure in different colours.

[9] In other words, reformulating this question using the dual moiré-theory terminology, why do these artifacts occur in spite of the lack of higher-order harmonics in the original continuous function $g(x)$?

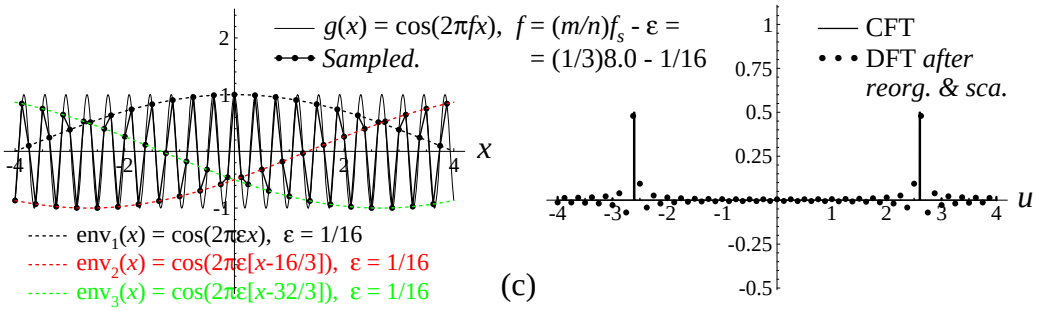
Signal domain $\boxed{\frac{m}{n} = \frac{1}{3}}$ Spectral domain



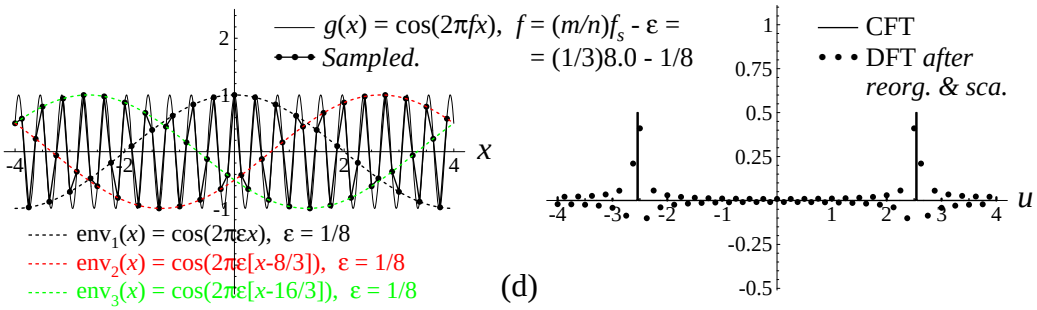

(a)



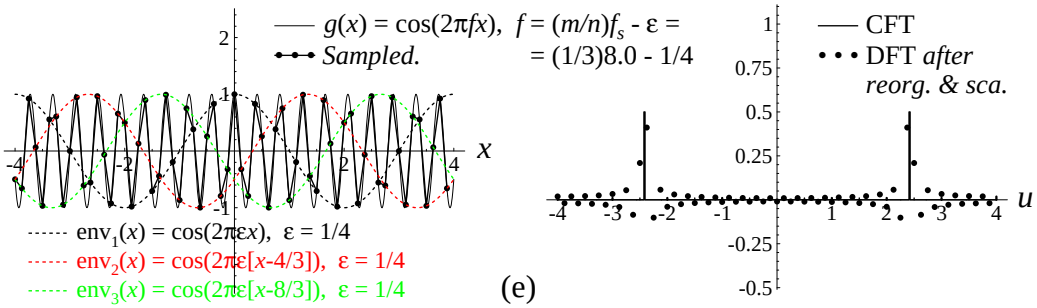

(b)




(c)




(d)




(e)

## 4. Derivation of the interlaced modulation envelopes

Let us first explain the phenomena which occur in the ($m$/$n$)-order sub-Nyquist artifact in the simplest case, when we are sampling a cosine function $g(x)$ with frequency $f$ and period $p = 1/f$:

$$g(x) = \cos(2\pi f x) \tag{11}$$

When we sample $g(x)$ at the sampling frequency $f_s$, i.e. using a sampling step of $\Delta x = 1/f_s$, we obtain the sampled signal:

$$g(x_k) = \cos(2\pi f x_k) \tag{12}$$

with $x_k = k\Delta x = k/f_s$.

Suppose, first, that the frequency of the original continuous cosine $g(x)$ is exactly

$$f = \tfrac{m}{n} f_s \tag{13}$$

In this case $p = \tfrac{n}{m}\Delta x$, meaning that we have exactly $\tfrac{n}{m}$ samples in each period $p$ of the given cosine, and the sampling step $\Delta x$ is $\tfrac{m}{n}$ of the cosine period $p$. The sampled signal we obtain in this case is:

$$g(x_k) = \cos(2\pi f x_k) = \cos(2\pi\, [\tfrac{m}{n} f_s]\, k/f_s) = \cos(2\pi k \tfrac{m}{n}) \tag{14}$$

This corresponds, indeed, to the singular state of the ($m$/$n$)-order sub-Nyquist artifact, which is systematically plotted in Figs. 2 and on in row (a).

Now, suppose that the frequency of the given cosine $g(x)$ is not exactly $f = \tfrac{m}{n} f_s$, but rather:

$$f = \tfrac{m}{n} f_s + \varepsilon \tag{15}$$

(where $\varepsilon$ may be either negative, as we did so far, or positive). The sampling frequency $f_s$ remains unchanged, so that we still have $x_k = k\Delta x = k/f_s$. The sampled signal is, in this case:

$$\begin{aligned} g(x_k) &= \cos(2\pi f x_k) = \cos(2\pi\, [\tfrac{m}{n} f_s + \varepsilon]\, k/f_s) \\ &= \cos(2\pi k \tfrac{m}{n} + 2\pi\varepsilon\, k/f_s) \\ &= \cos(2\pi x_k \varepsilon + 2\pi\phi) \qquad \text{with } \phi = k\tfrac{m}{n} \\ &= \cos(2\pi\varepsilon\, [x_k + \tfrac{\phi}{\varepsilon}]) \end{aligned} \tag{16}$$

As we can see, this is a sampled cosine having frequency $\varepsilon$ (i.e. period $1/\varepsilon$) and a nominal shift of $\tfrac{\phi}{\varepsilon}$, i.e. a *relative shift* of $\phi$ periods (see the definition of the various phase-related terms and notations, including $\phi$ and $\varphi$, in Appendix B).

This means that:

(1) For integer $\phi = k\frac{m}{n}$, namely for $k = 0, n, 2n, 3n, \ldots$ we have:

$$g(x_k) = \cos(2\pi x_k \varepsilon + 0) = \cos(2\pi\varepsilon\,[x_k + 0])$$

This cosine has the phase $\varphi = 0$, i.e. a shift of 0 or of an integer multiple of its period $\frac{1}{\varepsilon}$.

(2) For $k = 1, n+1, 2n+1, 3n+1, \ldots$ we have:

$$g(x_k) = \cos(2\pi x_k \varepsilon + \tfrac{m}{n}2\pi) = \cos(2\pi\varepsilon\,[x_k + \tfrac{m}{n}\tfrac{1}{\varepsilon}])$$

This cosine has the phase $\varphi = \frac{m}{n}2\pi$, i.e. a shift of $\frac{m}{n}$ times the period $\frac{1}{\varepsilon}$.

(3) For $k = 2, n+2, 2n+2, 3n+2, \ldots$ we have:

$$g(x_k) = \cos(2\pi x_k \varepsilon + 2\tfrac{m}{n}2\pi) = \cos(2\pi\varepsilon\,[x_k + 2\tfrac{m}{n}\tfrac{1}{\varepsilon}])$$

This cosine has the phase $\varphi = 2\frac{m}{n}2\pi$, i.e. a shift of $2\frac{m}{n}$ times the period $\frac{1}{\varepsilon}$.

. . .

($n$) For $k = n-1, 2n-1, 3n-1, 4n-1, \ldots$ we have:

$$g(x_k) = \cos(2\pi x_k \varepsilon + (n-1)\tfrac{m}{n}2\pi) = \cos(2\pi\varepsilon\,[x_k + (n-1)\tfrac{m}{n}\tfrac{1}{\varepsilon}])$$

This cosine has the phase $\varphi = (n-1)\frac{m}{n}2\pi$, i.e. a shift of $(n-1)\frac{m}{n}$ times the period $\frac{1}{\varepsilon}$.

This means that the successive sampled points of our original cosine function, $g(x_k) = \cos(2\pi f x_k)$, $k = 0, 1, 2,\ldots$ fall intermittently on one of $n$ cosinusoidal curves (that we call *envelopes*), which have all the same frequency $\varepsilon$ and period $1/\varepsilon$, and which only differ from each other in their phase. More precisely, these $n$ envelopes only differ from each other by successive relative shifts of $\frac{m}{n}$ periods $1/\varepsilon$, i.e. by successive shifts of $a = \frac{m}{n\varepsilon}$. This corresponds, indeed, to the situation we see in Figs. 2 and on in rows (b)-(e).

We have proved, therefore, the following result:

**Theorem 1**: Suppose we are given a continuous cosine function $g(x) = \cos(2\pi f x)$ having frequency $f$ and period $p = 1/f$, and that we sample this function at the sampling frequency $f_s$, i.e. with a sampling step of $\Delta x = 1/f_s$. If the frequency $f$ of our given cosine $g(x)$ differs by $\varepsilon$ from the singular frequency $\frac{m}{n}f_s$ for some integers $m$ and $n$:

$$f = \tfrac{m}{n}f_s + \varepsilon$$

(where $\varepsilon$ may be positive or negative), then the successive sampled points of our original cosine function, $g(x_k) = \cos(2\pi f x_k)$, $k = 0, 1, 2,\ldots$ fall intermittently on one of $n$ interlaced cosinusoidal envelopes, which have all the same frequency $f_{\text{env}} = \varepsilon$ and period $p_{\text{env}} = 1/\varepsilon$, and which only differ from each other in their phase. Any two successive envelopes are simply displaced from each other by $\frac{m}{n}$ of their period $p_{\text{env}}$, i.e. by a shift of $a = \frac{m}{n\varepsilon}$. ■

This theorem explains the ($m$/$n$)-order sub-Nyquist artifacts in the case of a cosinusoidal function. As already mentioned earlier, the rational fraction $m/n$ should be understood here as a reduced integer fraction, i.e. as a ratio of two mutually prime integers $m$ and $n$.

It is interesting to note that this theorem remains valid even in cases with $n = 1$. In such cases all the sampled points $g(x_k)$ fall on a single envelope, which corresponds to a *true* moiré effect: there are no interlaced envelopes, and the sampled points no longer jump intermittently from one curve to another as they do in a pseudo-moiré effect.

## 5. Sub-Nyquist artifacts in general periodic functions

In the previous sections we have only considered sub-Nyquist artifacts which occur when sampling a cosinusoidal function.

However, the same results can be also demonstrated, exactly in the same manner, for *any* periodic function $g(x)$, by considering the Fourier series development of $g(x)$. The detailed demonstration is provided in Appendix D. We thus obtain the following generalization of Theorem 1:[10]

**Theorem 2**: Suppose we are given a continuous periodic function $g(x)$ having frequency $f$ and period $p = 1/f$, and that we sample this function at the sampling frequency $f_s$, i.e. with a sampling step of $\Delta x = 1/f_s$. If the frequency $f$ of our given function $g(x)$ differs by $\varepsilon$ from the singular frequency $\frac{m}{n}f_s$ for some integers $m$ and $n$:

$$f = \frac{m}{n}f_s + \varepsilon$$

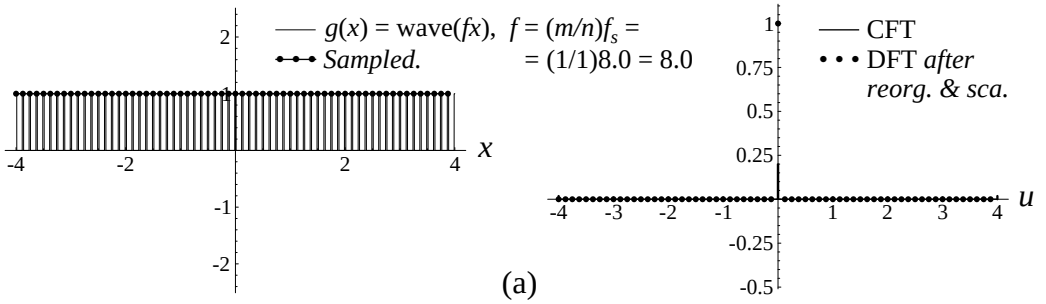
(where $\varepsilon$ may be positive or negative), then the successive sampled points of our original function, $g(x_k)$, $k = 0, 1, 2,...$ fall intermittently on one of $n$ interlaced low-frequency envelopes, which are simply expanded (stretched) versions of $g(x)$ having the frequency $f_{\text{env}} = \varepsilon$ and period $p_{\text{env}} = 1/\varepsilon$, and which only differ from each other in their phase. Any two successive envelopes are displaced from each other by $\frac{m}{n}$ of their period $p_{\text{env}}$, i.e. by a shift of $a = \frac{m}{n\varepsilon}$. ■

Once again, this theorem remains valid in cases with $n = 1$, too; in such cases all the sampled points $g(x_k)$ fall on a single envelope, which corresponds to a *true* moiré effect.

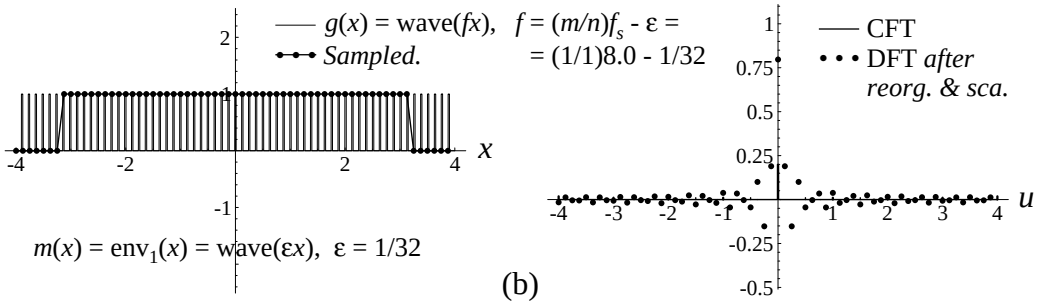
**Figure 7:** This figure is similar to Fig. 2, except that the original continuous function being sampled is the periodic square wave $g(x) = \text{wave}(fx)$ with frequency $f$ (and an opening ratio of $\tau/p = 1/5$, meaning that the 1-valued part of the wave occupies 1/5 of its period $p = 1/f$). The highly visible (1/1)-order artifact is generated because the sampled points $g(x_k)$ fall along a single low-frequency curve, which is simply a stretched version of $g(x)$.

[10] Interestingly, the fact that $g(x)$ may also contain higher harmonics of its basic frequency $f$, and that these harmonics may cause aliasing, does not affect the validity of the result. This point is explained by Remark 9 in Appendix C.

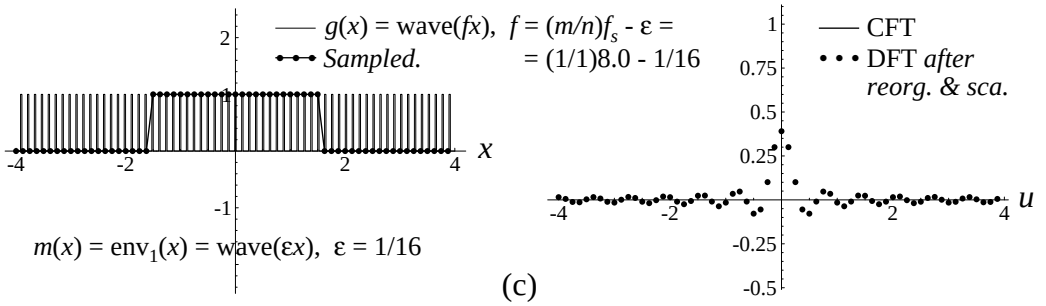
Signal domain $\boxed{\frac{m}{n}=\frac{1}{1}}$ Spectral domain



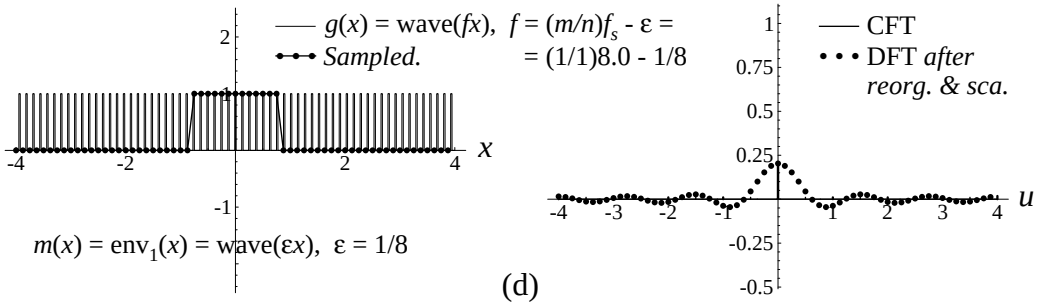

(a)



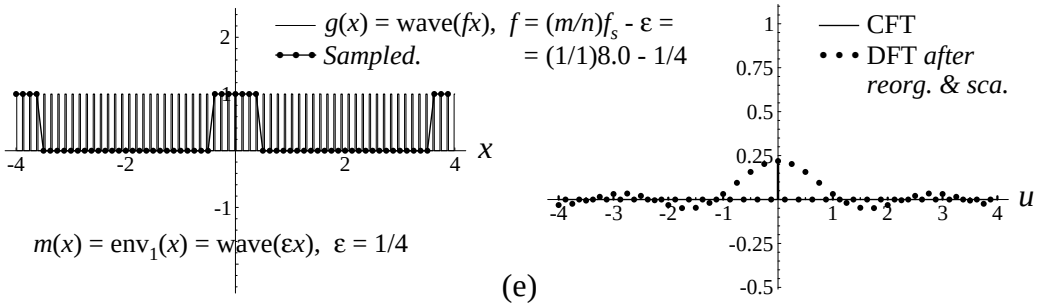

(b)




(c)




(d)




(e)

To illustrate this generalized theorem graphically, consider Figs. 7-9. These figures are similar to Figs. 2-4 above, except that the original continuous function being sampled here is a square wave with frequency $f$ rather than a cosine function with frequency $f$. In Figs. 7 and 8 (just like in their cosine counterparts Figs. 2 and 3), which correspond to the (1/1)- and (2/1)-order cases, respectively, the sampled points $g(x_k)$ fall on exactly one envelope. But in Fig. 9 (just like in Fig. 4, its cosine counterpart), which corresponds to the (1/2)-order case, the sampled points $g(x_k)$ fall on two interlaced envelopes, that are highlighted in our figure by two different colours. In all of Figs. 2-9, each of the envelopes we obtain is simply an expanded (stretched) version of the original continuous function $g(x)$ being sampled, having the frequency $\varepsilon$ (and period $1/\varepsilon$) rather than $f$ (and $p = 1/f$) as in the original function $g(x)$ itself. The magnification rate is, therefore, $(1/\varepsilon)/(1/f) = f/\varepsilon$ along the $x$ axis. Furthermore, whenever $n > 1$ so that interlacing does occur, successive interlaced envelopes are displaced from each other by $\frac{m}{n}$ of their period $1/\varepsilon$, i.e. by a shift of $a = \frac{m}{n\varepsilon}$.[11]

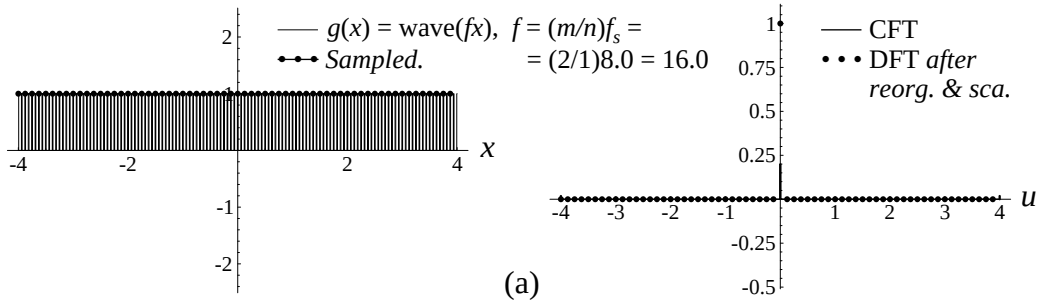
**Remark 2** (*the beating effect due to the modulating envelopes*):

Suppose we are sampling a cosinusoidal signal and that we plot the results, much like on the display of an oscilloscope, by connecting consecutive samples with straight line segments. When plotting the sampled signal densely, i.e. when the number of samples per envelope-period $p_{\text{env}}$ is relatively high, a sub-Nyquist artifact may confer to the sampled signal a typical wavy or fringy appearance. This beating or ripple effect is best observed when we plot the sampled signal $g(x_k)$ alone, without the original continuous-world cosine function $g(x)$ (note that in Figs. 1-9 we always overprinted the original continuous signal, too, for didactic reasons). This beating effect is illustrated for the cases of $(m/n) = (1/1)$, (1,2), (1/3) and (1/4) in Figs. 10-13, respectively. In each of these figures the left-hand column shows the sampled version of the continuous cosine function $g(x) = \cos(2\pi fx)$, and the right-hand column shows the sampled version of the continuous square-wave function $g(x) = \text{wave}(fx)$ having the same frequency $f$. In all cases the sampling frequency remains $f_s = 8$, and the only difference between consecutive rows is in the frequency $f$ of the original continuous function, as indicated in each row. Figs. 10-13 are therefore similar to our previous Figs. 2-9, but they are plotted 8 times more densely and without the original continuous-world signal itself, which was omitted in order not to obscure the sampled signal. Because the sampled

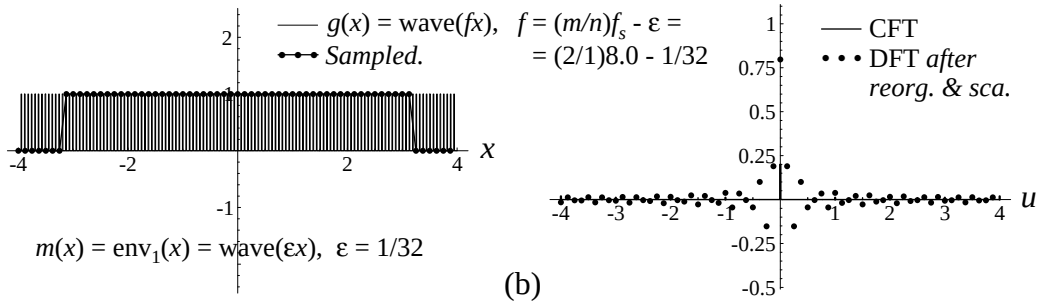
**Figure 8:** This figure is similar to Fig. 3, except that the original continuous function being sampled is the periodic square wave $g(x) = \text{wave}(fx)$ (with frequency $f$ and an opening ratio of $\tau/p = 1/5$). The highly visible (2/1)-order artifact is generated because the sampled points $g(x_k)$ fall along a single low-frequency curve, which is simply a stretched version of $g(x)$.

[11] It is interesting to note that unlike the cosine function, the square wave is not band limited (its spectrum contains infinitely many harmonics of $f$). This means that unlike in the cosinusoidal case, aliasing cannot be avoided here (unless the signal is low-pass filtered before being sampled). See also Remark 9 in Appendix C.

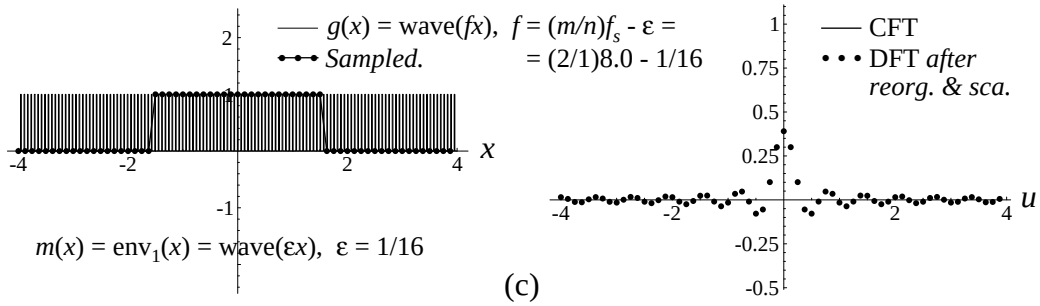
Signal domain $\boxed{\frac{m}{n} = \frac{2}{1}}$ Spectral domain



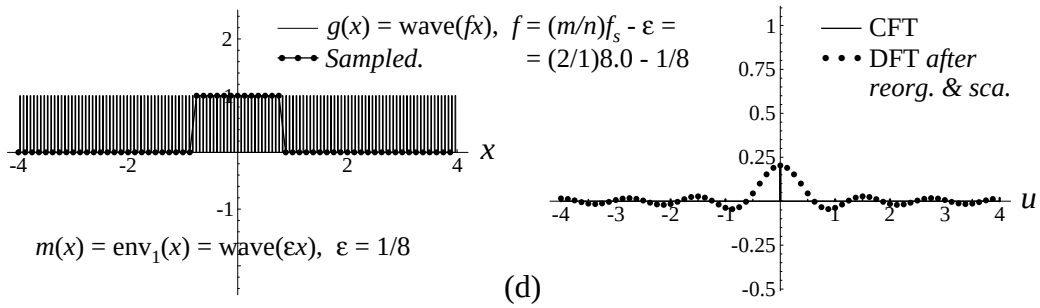

(a)



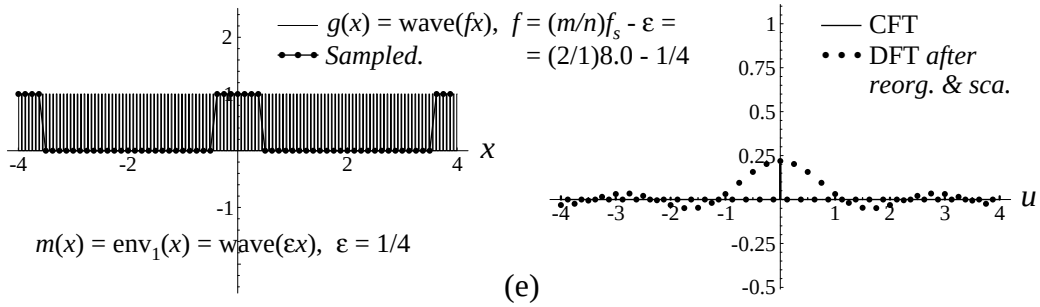

(b)




(c)




(d)




(e)

signal is plotted here more densely than in Figs. 2-9, the beating or ripple effect due to the modulating envelopes is more clearly visible, even without highlighting the envelope curves themselves as we did in Figs. 2-9.[12] In the cosinusoidal case, the depth of this ripple effect (i.e. the amplitude of the resulting corrugations along the top or bottom boundary of the sampled signal) depends on the values of $n$ and $m$: As we can see in the left-hand column of Fig. 11, when $m = 1$ and $n = 2$, i.e. in the case of the simple (1/2)-order sub-Nyquist artifact, this depth equals half of the amplitude of the original continuous cosinusoidal signal (the vertical distance between the node and the maximum of the two modulating envelopes). But as clearly seen in the left-hand columns of Figs. 11-13, the depth of the ripple effect decreases as the number $n$ increases. This depth can be readily found by calculating the height of the intersection points between the interlaced envelopes, and subtracting it from the maximum height of the envelopes. ■

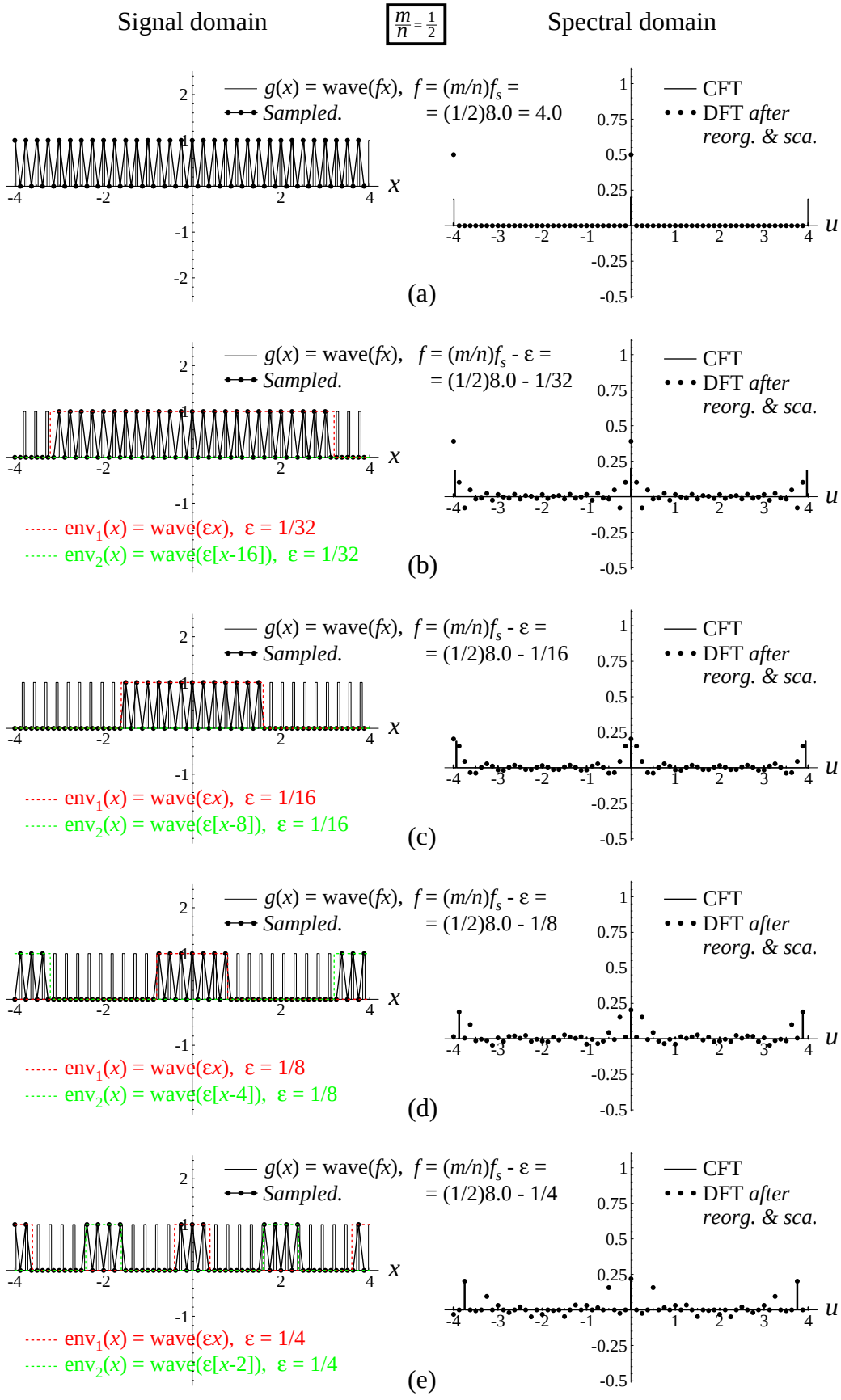
**Remark 3** (*significance of the sign of* $\varepsilon$):

Now that we know how to treat sub-Nyquist artifacts in any periodic function, including functions having an asymmetric period, we are ready to see the effect of the sign of $\varepsilon$. In all our figures so far we always used negative values of $\varepsilon$, meaning that the frequency $f$ of our given periodic function was varying below its singular value, $(m/n)f_s$.[13] Because our given function (cosine or square wave) was symmetric, the results obtained when using positive values of $\varepsilon$ were exactly the same. Let us see now what happens when our given function $g(x)$ is asymmetric. Consider, for example, the continuous periodic sawtooth function $\text{saw}(x) = x \bmod 1$, which consists of a sequence of asymmetric "teeth" having period 1 (or its normalized version, $g(x) = \text{saw}(fx)$, whose frequency is $f$). Fig. 14 shows in its left-hand column what happens to the sampled signal $g(x_k)$ when we take positive values of $\varepsilon$, and in its right-hand column what happens when we take negative values of $\varepsilon$ (see also Fig. 15, which shows the spectral domain, too). Although the behaviour in both cases remains globally the same, the orientation of the sub-Nyquist artifact is flipped (mirror-imaged) when $\varepsilon$ is negative (i.e. when $f$ is slightly

**Figure 9:** This figure is similar to Fig. 4, except that the original continuous function being sampled is the periodic square wave $g(x) = \text{wave}(fx)$ (with frequency $f$ and an opening ratio of $\tau/p = 1/5$). The highly visible (1/2)-order sub-Nyquist artifact is generated because consecutive points $g(x_k)$ of the sampled signal alternately jump from one of the $n = 2$ modulating envelopes to the other (each of the two modulating envelopes being simply a stretched and shifted version of $g(x)$). Note that these two interlaced modulating envelopes are highlighted in the figure in different colours.

[12] Note that the dark geometric patterns that may appear when plotting the sampled signal very densely are only display artifacts. They may differ depending on the display devices being used, and they disappear when plotting the same figure at a larger scale.

[13] The reason we have been using so far negative values of $\varepsilon$ is that in Eq. (6), i.e. in the case of the (1/2)-order sub-Nyquist artifact, we wanted $f$ to remain below the Nyquist limit of $\frac{1}{2}f_s$. We then continued using negative values of $\varepsilon$ in all the other cases, too, for the sake of consistency.


Signal domain
m/n = 1/2
Spectral domain
g(x) = wave(fx), Sampled.
f = (m/n)f_s = = (1/2)8.0 = 4.0
CFT
DFT after reorg. & sca.
x
u
(a)
f = (m/n)f_s - ε = = (1/2)8.0 - 1/32
env1(x) = wave(εx), ε = 1/32
env2(x) = wave(ε[x-16]), ε = 1/32
(b)
f = (m/n)f_s - ε = = (1/2)8.0 - 1/16
env1(x) = wave(εx), ε = 1/16
env2(x) = wave(ε[x-8]), ε = 1/16
(c)
f = (m/n)f_s - ε = = (1/2)8.0 - 1/8
env1(x) = wave(εx), ε = 1/8
env2(x) = wave(ε[x-4]), ε = 1/8
(d)
f = (m/n)f_s - ε = = (1/2)8.0 - 1/4
env1(x) = wave(εx), ε = 1/4
env2(x) = wave(ε[x-2]), ε = 1/4
(e)

below the singular value $(m/n)f_s$). In other words, as $f$ sweeps *backward* along the frequency axis and approaches the singular state from above, the sub-Nyquist artifact gradually becomes bigger, until it becomes infinitely big and disappears when $f$ precisely reaches the singular value $(m/n)f_s$. Then, when $f$ pursues its way backwards below the singular value, the $(m/n)$-order sub-Nyquist artifact "comes back from infinity" and becomes again visible, but this time with an inversed orientation. This behaviour is well known in the moiré theory (see, for example, Fig. C.24 in [9]), and it is interesting to see that it is preserved in the case of sub-Nyquist artifacts, too.

This result also means that the magnification rate $f/\varepsilon$ mentioned above (just before Remark 2) is in fact sign-sensitive, where a negative sign simply means magnification with mirror-inversion. Another effect of negative $\varepsilon$ values consists of an inversion in the direction of the envelope shifts, $a = \frac{m}{n\varepsilon}$. ■

Finally, it should be noted that Theorem 2 only holds when we are sampling a *periodic* function $g(x)$. Nevertheless, this theorem may be of interest even when the function $g(x)$ is aperiodic but still contains a strong periodic component (which manifests itself as a spike in the spectrum of $g(x)$).

## 6. Discussion: sampling and reconstruction considerations

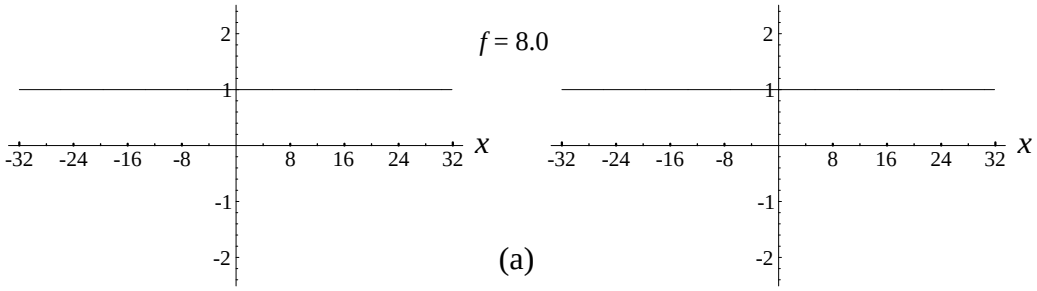
As we have mentioned in Sec. 1, although sub-Nyquist artifacts are generated during the sampling process, they are often considered as reconstruction artifacts (meaning that they appear as a result of *poor reconstruction*). Let us explain this in some more detail.

**Figure 10:** The (1/1)-order artifact (which is, in fact, a true first-order sampling moiré) as it appears in densely plotted periodic signals. Left-hand column: the periodic signal $g(x) = \cos(2\pi fx)$ having frequency $f$, after being sampled with a sampling frequency of $f_s = 8.0$ (i.e. with a sampling interval of $\Delta x = 1/f_s = 1/8$). Right-hand column: the periodic square wave signal $g(x) = \text{wave}(fx)$ (with opening ratio of $\tau/p = 1/5$) having the same frequency $f$, after being sampled with the same sampling frequency of $f_s = 8.0$ (and sampling interval of 1/8). Each row in the figure shows both of the sampled signals for the following values of $f$ (frequency of the signal $g(x)$): (a) $f = f_s$ (the singular state). (b) $f = f_s - 1/128$. (c) $f = f_s - 1/64$. (d) $f = f_s - 1/32$. (e) $f = f_s - 1/16$. All sampled signals are plotted here as *connected line plots*, meaning that consecutive sample points $g(x_k)$ are connected by straight line segments (see [10, Sec. 1.5.1]), just as on the display of an oscilloscope. The signals are plotted much more densely than in Figs. 2-9 (note that we use here the same sampling interval $\Delta x = 1/8$ as in those figures but an 8-fold larger sampling range, and hence 8 times more samples). Unlike in Figs. 2-9 the original continuous signal is not shown here, in order not to obscure the sampled signal itself. The highly visible (1/1)-order artifact is generated because the sampled points of each signal fall along a single low-frequency curve, which is simply a stretched version of $g(x)$, as shown in Figs. 2 and 7.

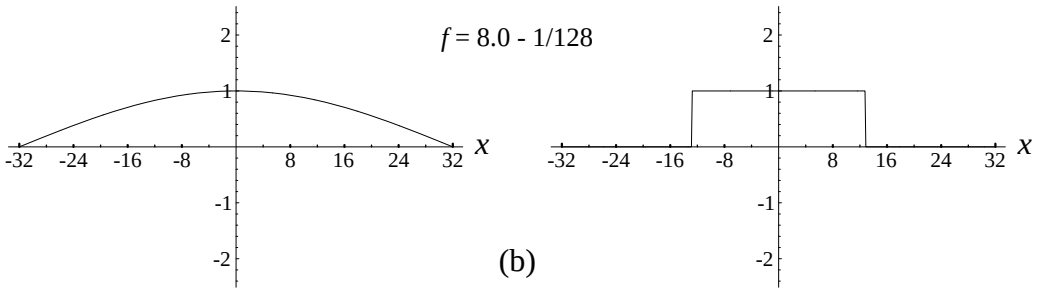
— $g(x) = \cos(2\pi fx)$
*Sampled at* $f_s = 8.0$

$\frac{m}{n} = \frac{1}{1}$

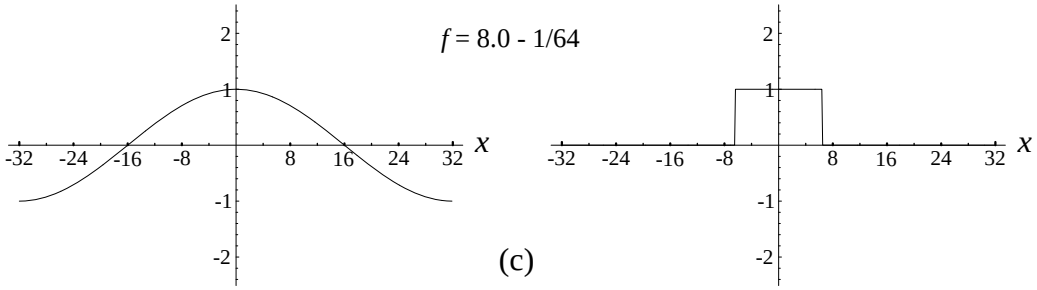
— $g(x) = \text{wave}(fx)$
*Sampled at* $f_s = 8.0$



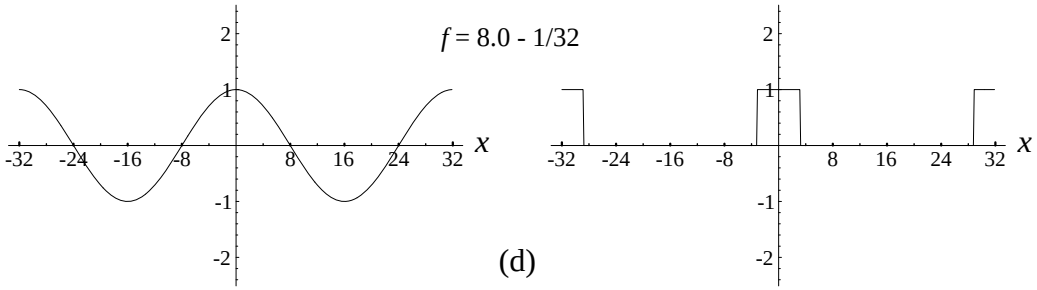

(a)



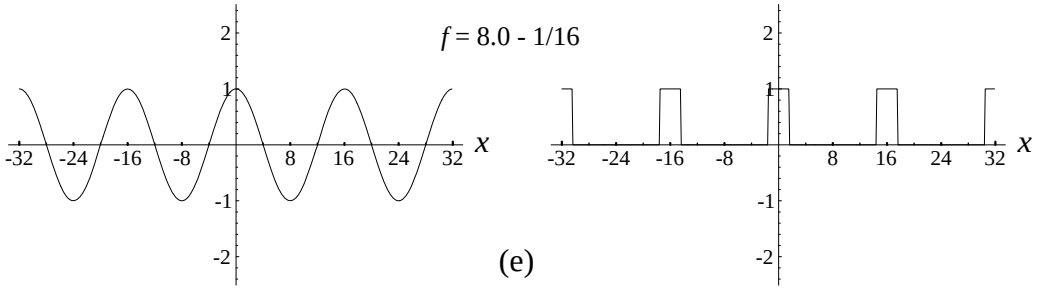

(b)




(c)




(d)




(e)

Sampling is the process that converts a continuous signal to a discrete one, while reconstruction is the process that recreates a continuous signal from its samples [2, Sec. 14.10.5]; [1]. Note that theoretically, a sampled signal consists of zero-width impulses (of varying strengths), and it is precisely the reconstruction process that brings back the "flesh" around each of the sampled "bones". Now, according to the classical sampling theorem (see, for example, [13, Sec. 8-7], [14, p. 115] or [15, p. 1593]), all the information in the original continuous signal $g(x)$ is preserved in its sampled version $g(x_k)$, if the sampling frequency is at least twice the highest frequency contained in $g(x)$. Under this condition, the theorem guarantees that the original continuous signal can be perfectly reconstructed by sinc interpolation, i.e. by convolving the sampled signal $g(x_k)$ with a specified narrow sinc function (that is precisely the inverse Fourier transform of the rectangular window function in the spectral domain that cuts off all the frequencies beyond half of the sampling frequency; see [12, p. 83]). This result remains true even if the sampling process itself introduces into the sampled signal $g(x_k)$ new modulation or beating effects (sub-Nyquist artifacts) that did not exist in the original continuous signal $g(x)$, as it happens, for instance, in Fig. 4. In other words, if reconstruction is done as stipulated by the sampling theorem, the theorem indeed guarantees that these artifacts will disappear in the reconstruction process.

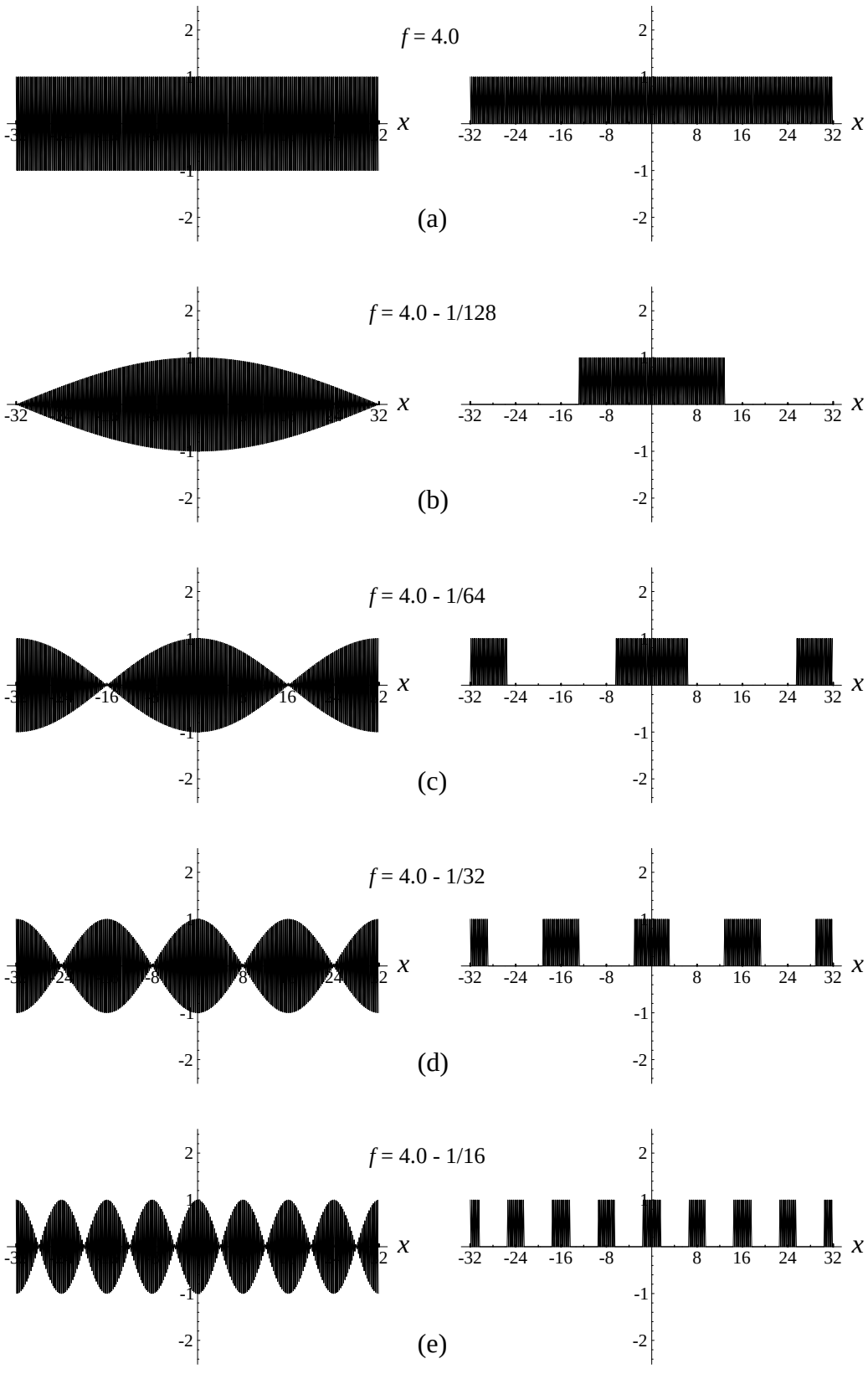
However, in practice, the reconstruction of a sampled signal can never be done by a sinc interpolation as stipulated by the sampling theorem, because the sinc function extends *ad infinitum* to both directions. Instead, reconstruction is very often performed by means of a linear interpolation, meaning that successive sample points are simply connected by straight line segments, just as on the display of an oscilloscope. This is, indeed, what we are also doing in our figures. Now, because the reconstruction process is not performed as stipulated by the sampling theorem, it is clear that the original continuous signal $g(x)$ cannot be perfectly reconstructed from the sampled signal $g(x_k)$. Indeed, if perfect reconstruction by sinc interpolation could be performed, the sub-Nyquist artifacts would have been eliminated in the reconstruction (interpolation) process, and we would obtain a continuous signal that is perfectly identical to $g(x)$. But when the reconstruction is performed by a non-ideal reconstruction method such as linear interpolation, the sub-Nyquist artifacts are not eliminated during the reconstruction process, and they remain visible in our resulting signal. Obviously, the closer we approach sinc interpolation better are the results, but in the real world we can never reach the ideal conditions set by the sampling theorem.

**Figure 11:** The (1/2)-order sub-Nyquist artifact as it appears in densely plotted periodic signals. This figure is similar to Fig. 10, except for the signal-frequency $f$ being used in each row: (a) $f = \frac{1}{2}f_s$ (the singular state). (b) $f = \frac{1}{2}f_s - 1/128$. (c) $f = \frac{1}{2}f_s - 1/64$. (d) $f = \frac{1}{2}f_s - 1/32$. (e) $f = \frac{1}{2}f_s - 1/16$. The highly visible (1/2)-order sub-Nyquist artifact is generated because consecutive points of each sampled signal alternately jump from one of the $n = 2$ modulating envelopes to the other, as shown in greater detail in Figs. 4 and 9, respectively.

— $g(x) = \cos(2\pi fx)$ *Sampled at* $f_s = 8.0$

$\frac{m}{n} = \frac{1}{2}$

— $g(x) = \text{wave}(fx)$ *Sampled at* $f_s = 8.0$



Note, however, that true aliasing moiré effects, those that occur when the Nyquist condition is not satisfied, like in Figs. 2 or 3, would not be eliminated even by an ideal reconstruction: These artifacts, which *are* indeed represented by low frequencies in the spectrum, are embedded in the sampled signal and remain visible in the reconstructed signal, too, whichever reconstruction method we may choose.

Having understood the origin and the nature of sub-Nyquist artifacts, an obvious question may now arise: Since these artifacts cannot be eliminated in the reconstruction stage, is it possible to avoid them by judiciously choosing our sampling frequency? As one could guess, avoiding such "dangerous frequency zones" is difficult, since they may lurk anywhere along the frequency axis, including in zones that fully respect the frequency condition required by the sampling theorem. A possible rule of thumb [7, p. 15] consists of taking $f_S > 10f$ (i.e. $f_S \approx \frac{n}{m}f > 10f$ and hence $\frac{n}{m} > 10$) rather than the Nyquist requirement of $f_S > 2f$ (i.e. $f_S \approx \frac{n}{m}f > 2f$ and hence $\frac{n}{m} > 2$). The reasoning behind this rule of thumb is that for high values of $n$ the artifacts in question become rather negligible (see Remarks 2 and 4). Nevertheless, a good practice when sampling a given signal having a strong periodic component of frequency $f$ would be to avoid sampling-frequencies $f_S$ in the close neighbourhood of $(n/m)f$ for relatively small integers $n$ and $m$. If these "dangerous zones" are carefully avoided, one may choose even lower values of $f_S$, which are often more practical to use. It should be mentioned that in some particular cases signals can be correctly reconstructed even when using a sampling rate much lower than the Nyquist requirement. This subject is being studied in the field of compressive sampling, which continues gaining interest since the work of D. L. Donoho and E. J. Candès first appeared in 2006 [16], [17]. In the present contribution we only address the general case, in which the classical sampling theorem applies, but possible extensions of this subject within other sampling frameworks could also be studied in future works.

## 7. Significance in applications

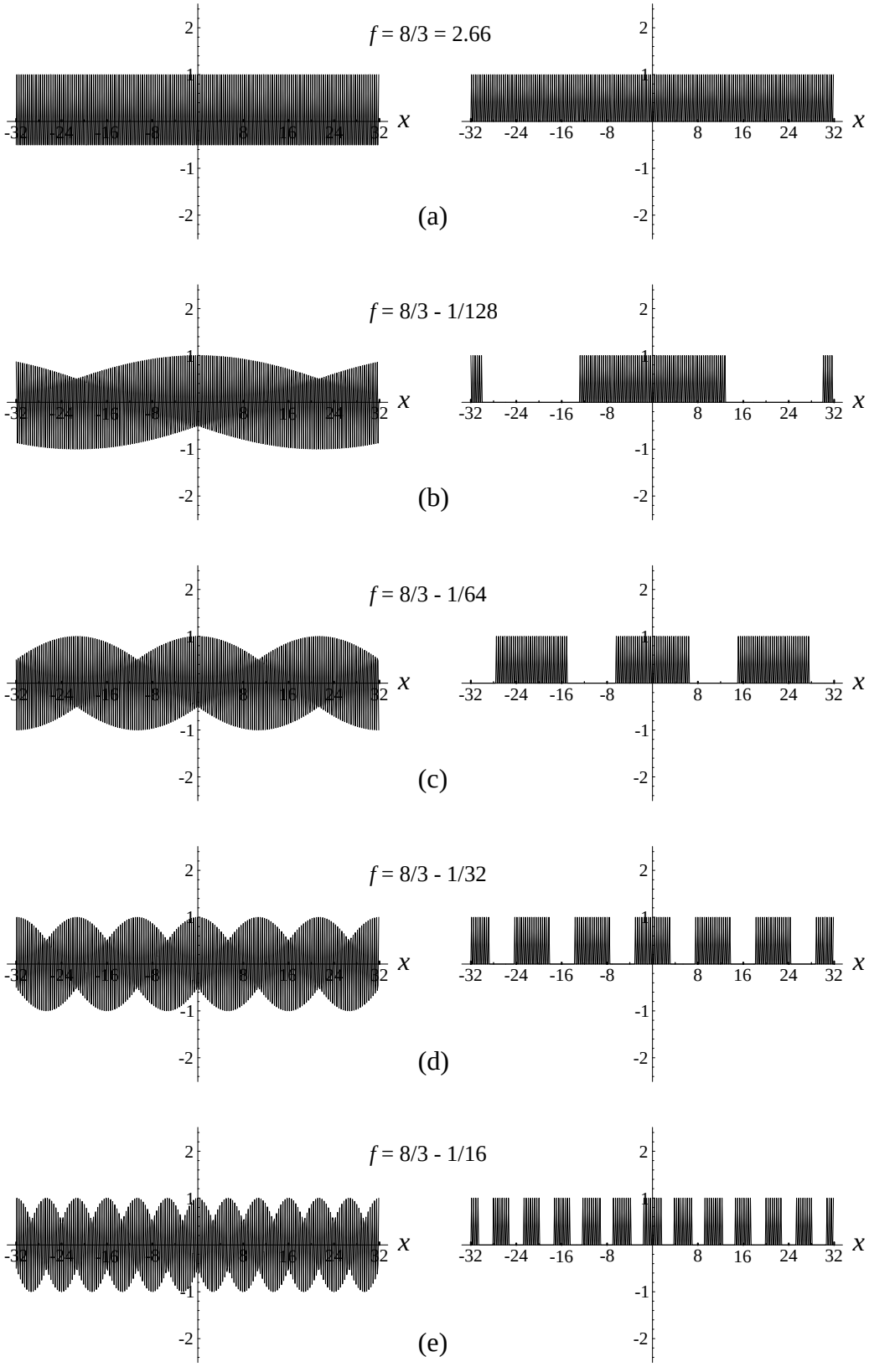
Sub-Nyquist artifacts may occur in various signal processing applications, and one should be aware of their existence and of their correct interpretation. Each application involving analog to digital conversion of continuous periodic signals, including simple sinusoidal waves, may give rise to such artifacts in the resulting digital signal, even if the original signal is band limited and the Nyquist condition is respected.

**Figure 12:** The (1/3)-order sub-Nyquist artifact as it appears in densely plotted periodic signals. This figure is similar to Figs. 10-11, except for the signal-frequency $f$ being used in each row: (a) $f = \frac{1}{3}f_S$ (the singular state). (b) $f = \frac{1}{3}f_S - 1/128$. (c) $f = \frac{1}{3}f_S - 1/64$. (d) $f = \frac{1}{3}f_S - 1/32$. (e) $f = \frac{1}{3}f_S - 1/16$. The highly visible (1/3)-order sub-Nyquist artifact is generated because consecutive points of each sampled signal alternately jump from one of the $n = 3$ modulating envelopes to the others, as shown in greater detail in Fig. 6 for the cosinusoidal case.

— $g(x) = \cos(2\pi fx)$ *Sampled at* $f_s = 8.0$

$\frac{m}{n} = \frac{1}{3}$

— $g(x) = \text{wave}(fx)$ *Sampled at* $f_s = 8.0$



In many cases such phenomena (with $n > 1$) are harmless, since they do not truly introduce new false frequencies into the system, as aliasing phenomena do. But in other cases, as witnessed in [3-7], a misunderstanding of these phenomena may lead to undesirable results even when $n > 1$. This may happen, for example, in applications that rely on amplitude peak detection in the sampled signal. In such cases, false peaks may be detected in the sampled signal, even though they do not exist in the original continuous signal (see, for example, Figs. 4-6 or 11-13).

Care should be taken in particular when the resulting signal has to undergo in subsequent processing steps any non-linear operation (such as rectification, non-linear amplification, etc.). A non-linear operation may truly insert the envelope frequencies in question into the signal and its spectral representation [18], and hence flaw the processing results. This may happen, for example, when applying to the beating signal a nonlinear operation such as envelope detection; see also Footnote 5 in Sec. 2.2. Other examples in image processing include operations such as quantization or halftoning.

It should also be noted that when further processing of the sampled signal is required, one should always use the sampled signal itself, and not its reconstructed version (which is necessarily based on non-ideal reconstruction, as we have seen above). The reconstructed version of the sampled signal can be used for plotting or displaying purposes, but not for any further processing steps.

## 8. Conclusions

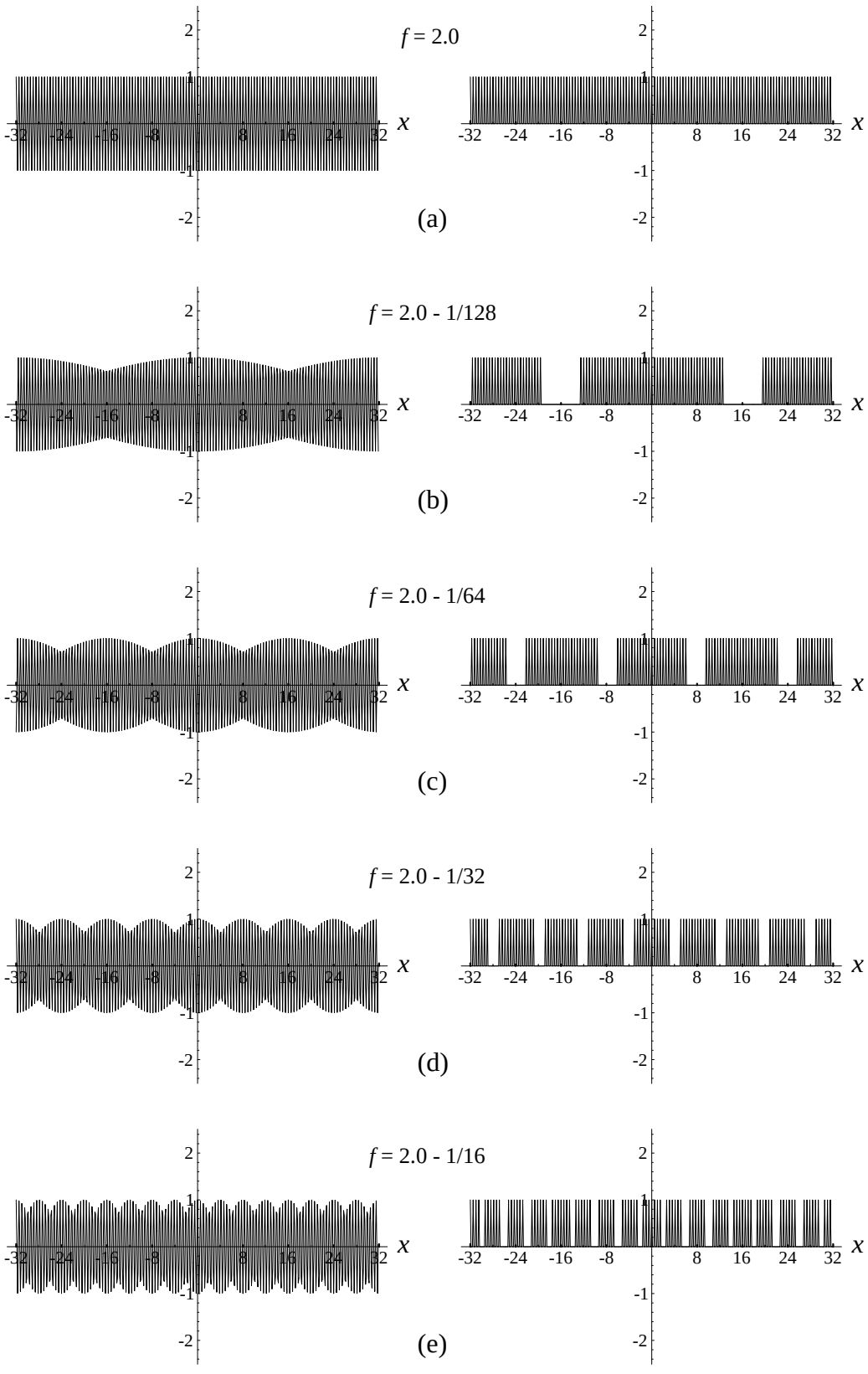
We show in this paper that sub-Nyquist artifacts and sampling moiré effects are, indeed, particular cases of the same phenomenon: Whenever the frequency $f$ of our given continuous periodic function $g(x)$ is close to a critical point $(m/n)f_s$ where $f_s$ is the sampling frequency and $m$ and $n$ are integers, a visible artifact is generated in the resulting sampled signal. This artifact is explained by our main theorems: It occurs since, under these circumstances, the successive sampled points of our original function, $g(x_k)$, $k = 0, 1, 2,...$ fall intermittently on one of $n$ interlaced low-frequency envelopes, which are simply expanded (stretched) versions of $g(x)$ having the frequency $\varepsilon$ and period $1/\varepsilon$, and which only differ from each other in their phase. If $n = 1$ then all the sampled points fall on a single envelope, and the resulting effect is simply a sampling

**Figure 13:** The (1/4)-order sub-Nyquist artifact as it appears in densely plotted periodic signals. This figure is similar to Figs. 10-12, except for the signal-frequency $f$ being used in each row: (a) $f = \frac{1}{4}f_s$ (the singular state). (b) $f = \frac{1}{4}f_s - 1/128$. (c) $f = \frac{1}{4}f_s - 1/64$. (d) $f = \frac{1}{4}f_s - 1/32$. (e) $f = \frac{1}{4}f_s - 1/16$. The highly visible (1/4)-order sub-Nyquist artifact is generated because consecutive points of each sampled signal alternately jump from one of the $n = 4$ modulating envelopes to the others, as shown in greater detail in Fig. 5 for the cosinusoidal case.

— $g(x) = \cos(2\pi fx)$
*Sampled at* $f_s = 8.0$

$\frac{m}{n} = \frac{1}{4}$

— $g(x) = \text{wave}(fx)$
*Sampled at* $f_s = 8.0$



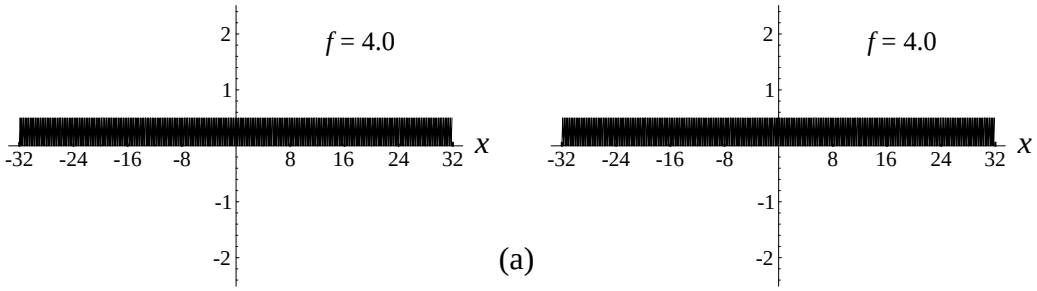
moiré effect, which indeed behaves exactly as predicted in the moiré theory. But when $n > 1$, the successive sampling points $g(x_k)$ alternately jump from one modulating envelope to another, and the resulting effect is a sub-Nyquist artifact. If $g(x)$ does not possess the $n$-th harmonic of its frequency $f$, as in the cosinusoidal case, this is a pure modulation phenomenon and it is not accompanied by a corresponding low-frequency impulse in the spectrum. And yet, as we have seen throughout this paper, in spite of this difference between true moiré effects and sub-Nyquist artifacts, they all behave in a very similar way. Let us mention, for example, their similar behaviour at the singular point or around it, their mirror-inversion depending on the sign of $\varepsilon$ (i.e. depending on which side of the singular point they are located), etc. This striking similarity between true moirés and sub-Nyquist artifacts is indeed explained by the fact that they are both particular cases of the same phenomenon, as shown by our theorems.

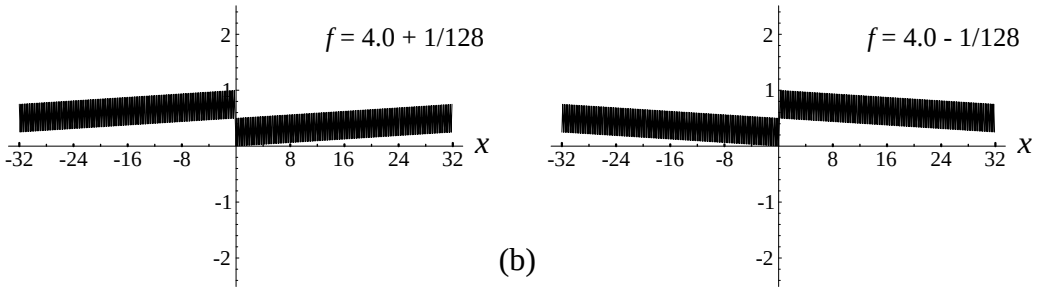
Finally, as we have seen in Sec. 6, sub-Nyquist artifacts are really generated during the sampling process, but they remain visible due to imperfect reconstruction. Our analysis of the sub-Nyquist artifacts would be of interest even if these phenomena only existed in the sampled signal, and then disappeared during the reconstruction process. But because perfect reconstruction does not exist in the real world, so that sub-Nyquist artifacts cannot be completely eliminated from the reconstructed signal, the analysis provided here becomes all the more valuable and pertinent. Our results may be of interest to people working in a wide range of applications, where such artifacts may occur and should be therefore well understood.

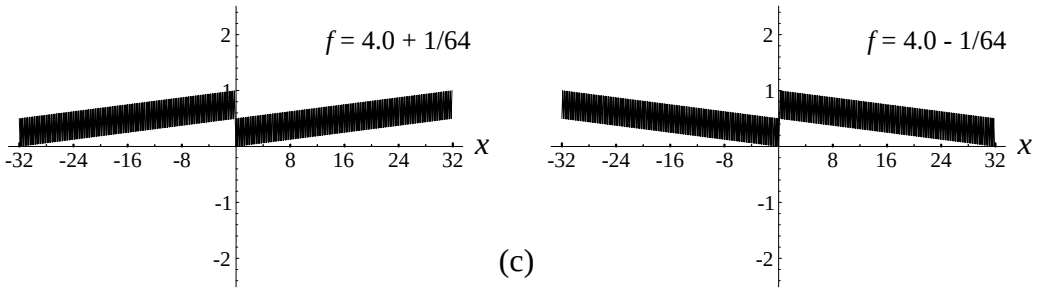
**Figure 14:** The (1/2)-order sub-Nyquist artifact as it appears in densely plotted periodic sawtooth signals. This figure illustrates the role of the sign of $\varepsilon$: It shows in the right-hand column the periodic sawtooth signal $g(x) = \text{saw}(fx) = fx \bmod 1$ (where $f$ is the signal's frequency) after being sampled with a sampling frequency of $f_s = 8.0$ (i.e. with a sampling interval of $\Delta x = 1/8$), with the following values of $f$ (frequency of the original signal $g(x)$): (a) $f = f_s$ (the singular state). (b) $f = f_s - 1/128$. (c) $f = f_s - 1/64$. (d) $f = f_s - 1/32$. (e) $f = f_s - 1/16$. The left-hand side column shows what happens in the very same cases if we only change the sign of $\varepsilon$ so that the values of $f$ become: (a) $f = \frac{1}{2}f_s$ (the singular state). (b) $f = \frac{1}{2}f_s + 1/128$. (c) $f = \frac{1}{2}f_s + 1/64$. (d) $f = \frac{1}{2}f_s + 1/32$. (e) $f = \frac{1}{2}f_s + 1/16$. Note that the modulating envelopes in both columns are simply stretched (and shifted) versions of $g(x)$, but in the right-hand column (where the sign of $\varepsilon$ is negative) they are also mirror-inversed with respect to the original sawtooth $g(x)$. Here, too, the highly visible (1/2)-order sub-Nyquist artifacts are generated because consecutive points of each sampled signal alternately jump from one of the $n = 2$ modulating envelopes to the other. See also Fig. 15, in greater detail.

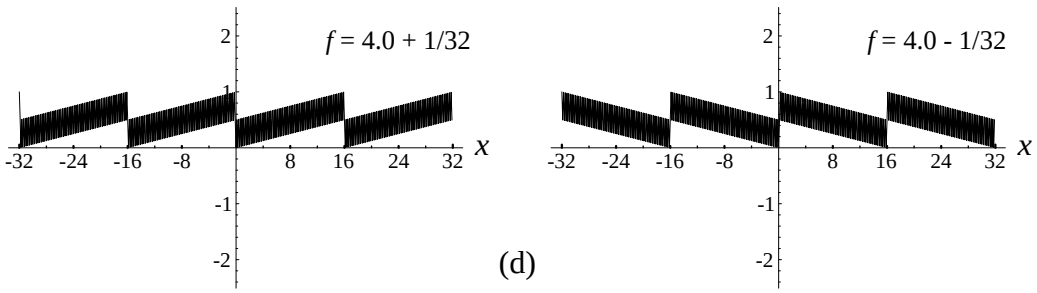
— $g(x)$ = saw($fx$)
*Sampled at* $f_s = 8.0$

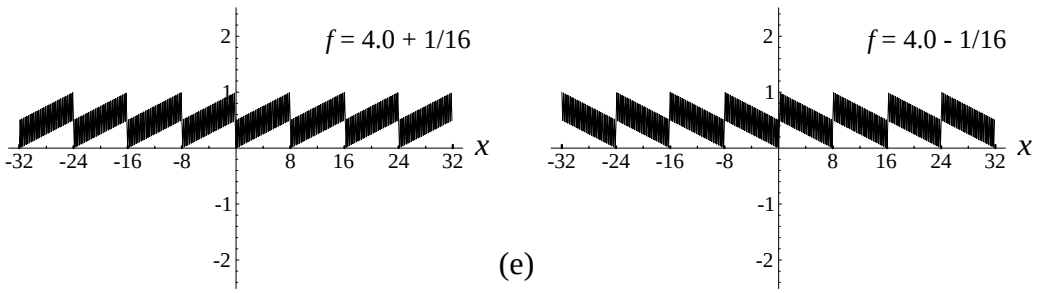
$\frac{m}{n} = \frac{1}{2}$



## Appendix A: Moiré theory notations and terminology

In this appendix we provide a brief review, as far as space limitations allow, of some basic notions and terms from the moiré theory that are being used in the present paper. Interested readers may find further details in publications on the moiré theory such as [9].

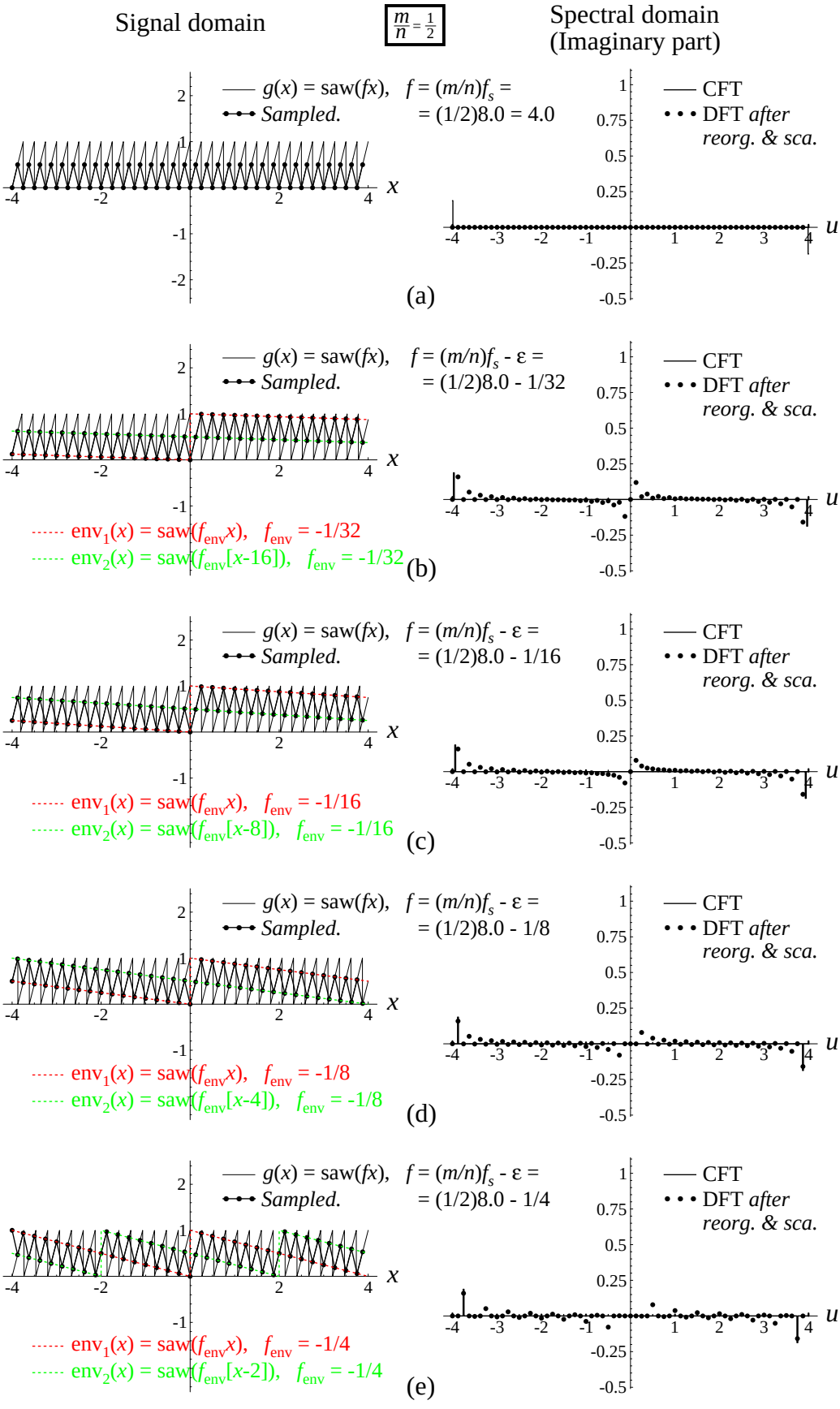
The moiré effect is a well-known phenomenon which occurs when two or more repetitive structures (such as periodic signals, waves, line gratings, grids, etc.) are superposed. It consists of a new pattern of alternating dark and bright areas which is clearly observed at the superposition, although it does not appear in any of the original structures (Fig. A1). Usually moiré effects are studied in two-dimensional settings, but the moiré theory is general and it covers also one-dimensional settings (as in the present paper).

Moiré effects may occur in different circumstances. For example, *superposition moirés* occur in the superposition (e.g. overprinting) of two or more periodic structures. *Sampling moirés* occur in the sampling process when an original periodic function $g(x)$ is sampled by a periodic sampling comb. In both cases the interaction between the individual periodic structures gives rise to the resulting moiré effect.

The moiré phenomenon can be explained both in the signal domain and in the spectral domain. From the point of view of the signal domain, the moiré effect occurs due to an interaction between the superposed structures. It results from the geometric distribution of dark and bright areas in the superposition: areas where dark elements of the original structures fall on top of each other appear brighter than areas in which dark elements fall between each other and fill the spaces better (see Fig. A1).

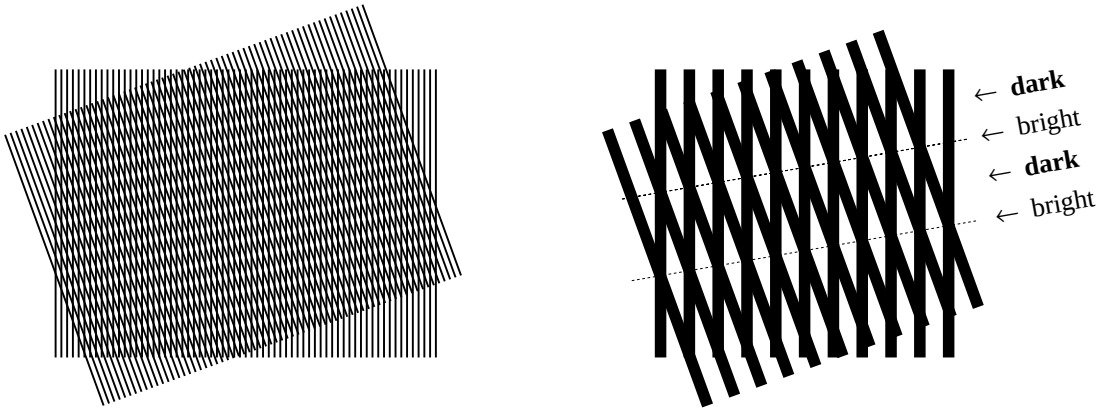
The spectral or Fourier domain explanation, on its part, is based on the frequencies involved in each of the $N$ superposed layers, in the corresponding frequency spectrum (see Fig. A2). The superposition of layers in the signal domain is viewed mathematically as the *product* of the given structures (where 0 represents black, 1 represents white, and intermediate values represent in-between shades). Therefore, in the Fourier frequency domain, the spectrum of the superposition is the *convolution* of

**Figure 15:** A more detailed view of the (1/2)-order sub-Nyquist artifact that is shown in the right-hand column of Fig. 14, along with the corresponding spectral domain representations. Note that saw($fx$) can be seen as an odd function plus the constant 0.5; the spectrum of an odd real-valued function is odd and imaginary-valued [19, p. 15], and the spectrum of the constant 0.5 is a real-valued impulse of strength 0.5 at the origin (the real-valued part of the spectrum is not shown here). Compare with the cosine and square-wave counterparts in Figs. 4 and 9, respectively; note that the envelope frequency $f_{\text{env}}$ has negative $\varepsilon$ values in all of these figures, but it only matters in the present figure, where $g(x)$ is not symmetric.


Signal domain
m/n = 1/2
Spectral domain
(Imaginary part)
g(x) = saw(fx),
Sampled.
f = (m/n)f_s =
= (1/2)8.0 = 4.0
CFT
DFT after
reorg. & sca.
x
u
(a)
f = (m/n)f_s - ε =
= (1/2)8.0 - 1/32
env_1(x) = saw(f_env x), f_env = -1/32
env_2(x) = saw(f_env[x-16]), f_env = -1/32
(b)
= (1/2)8.0 - 1/16
env_1(x) = saw(f_env x), f_env = -1/16
env_2(x) = saw(f_env[x-8]), f_env = -1/16
(c)
= (1/2)8.0 - 1/8
env_1(x) = saw(f_env x), f_env = -1/8
env_2(x) = saw(f_env[x-4]), f_env = -1/8
(d)
= (1/2)8.0 - 1/4
env_1(x) = saw(f_env x), f_env = -1/4
env_2(x) = saw(f_env[x-2]), f_env = -1/4
(e)

the spectra of the $N$ individual layers. This convolution often gives rise to new low frequencies which did not exist in any of the original layers; these frequencies are the spectral-domain representation of the moiré effects that we see in the superposition, back in the signal domain.

Since our original structures are periodic, their spectra consist of impulses (the base frequency of the structure and its higher harmonics, if any). Considering the impulse locations in the spectrum as vectors emanating from the origin, the impulse locations in the convolution of $N$ such spectra can be viewed as a set of vector sums: Each frequency in the convolution is a vector sum of one vector from the first spectrum, one vector from the second spectrum, ... and one vector from the $N$-th spectrum: $k_1f_1 + ... + k_Nf_N$, where $f_i$ is the base frequency of the $i$-th superposed structure and $k_if_i$ is the $k_i$ harmonic (provided that it exists). If any of the new vector sums in the convolution falls close to the spectrum origin, within the visibility circle[14], namely if $k_1f_1 + ... + k_Nf_N \approx 0$, the resulting frequency corresponds to a visible moiré effect in the structure superposition. When the vector sum in question falls exactly on the spectrum origin we get the *singular state* of the moiré, where the moiré period is infinitely long and therefore not visible. But as soon as the vector sum starts moving away from the origin, the corresponding moiré "comes back from infinity" and becomes again visible with a long period. As the moiré vector (the vector sum in question) moves away from the origin to any direction, the frequency of the moiré gradually increases and its period decreases, until finally the moiré effect becomes invisible and disappears.




(a) (b)

**Figure A1:** (a) The moiré effect in the superposition of two identical, mutually rotated line gratings. It consists of a new pattern of alternating dark and bright areas that are only formed in the superposed area. (b) Enlarged view.

[14] The *visibility circle* is a circular region with a small radius about the spectrum origin, which includes all frequencies that are small enough to be visible (under some given viewing conditions). In the moiré theory the visibility circle is usually drawn assuming viewing conditions under which only the moiré effects are visible, but all the original frequencies $f_1,...,f_N$ remain beyond the visibility circle. See, for example, [9, Chapter 2].

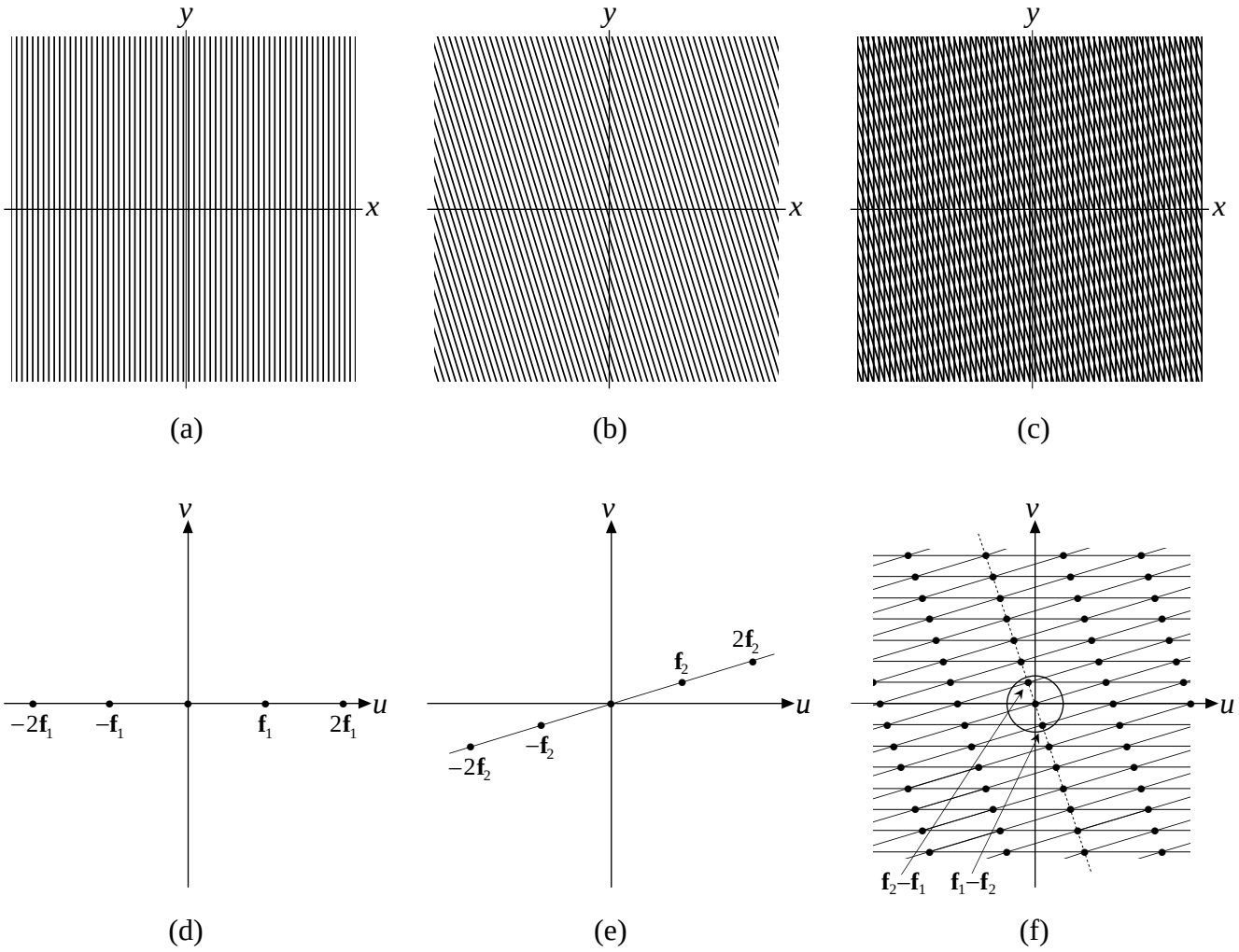



**Figure A2:** Line gratings (a) and (b) and their superposition (c) in the signal domain; the respective spectra are the infinite combs shown in (d) and (e), and their convolution (f). Note in the center of spectrum (f) the new impulse pair near the origin whose frequency vectors are $\mathbf{f}_1 - \mathbf{f}_2$ and $\mathbf{f}_2 - \mathbf{f}_1$. This is the fundamental impulse pair of the (1,-1)-moiré seen in the superposition (c). Note that the spectra (d)-(f) are purely impulsive; the straight lines connecting between the impulse locations have been added for didactic reasons only, to clarify the geometric structure of these spectra. The dotted line in (f) indicates the infinite impulse comb representing the moiré.

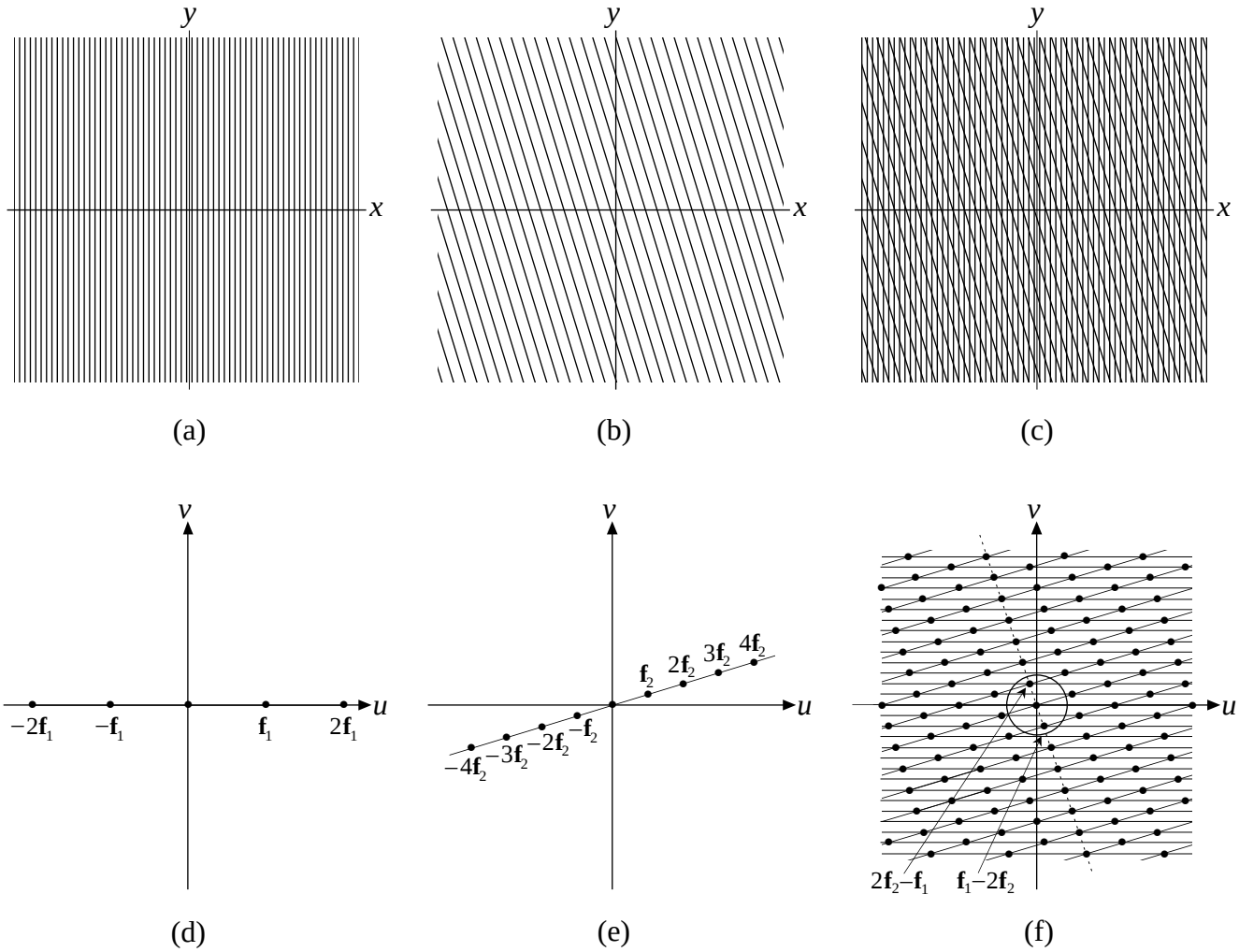



**Figure A3:** Line gratings (a) and (b) as in Fig. A2 but with (b) having half the frequency, and their superposition (c); the respective spectra are the infinite combs shown in (d) and (e), and their convolution (f). Note in the center of spectrum (f) the new impulse pair near the origin whose frequency vectors are $\mathbf{f}_1 - 2\mathbf{f}_2$ and $2\mathbf{f}_2 - \mathbf{f}_1$, which originate from the second harmonic of $\mathbf{f}_2$. This is the fundamental impulse pair of the (1,-2)-moiré seen in the superposition (c). The dotted line in (f) indicates the infinite impulse comb representing the moiré.

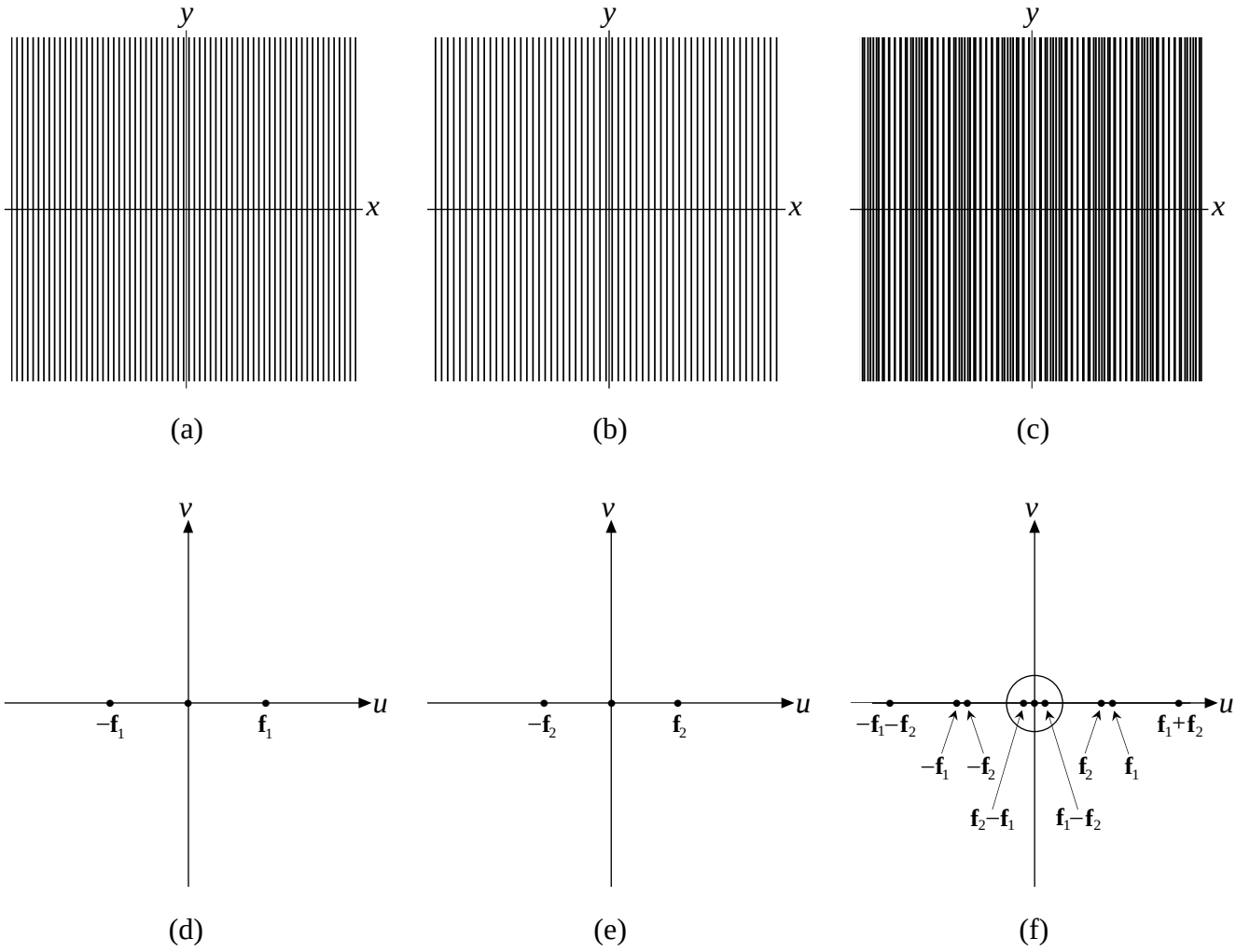



**Figure A4:** Vertical line gratings (a) and (b) and their superposition (c) in the signal domain; the respective spectra are shown in (d) and (e) and their convolution (f). Since in this case all the frequency vectors in the spectral domain fall on the horizontal *u* axis, only the first-harmonic frequencies are shown, to avoid cluttering along this axis. Note in the center of spectrum (f) the new impulse pair near the origin whose frequency vectors are $\mathbf{f}_1 - \mathbf{f}_2$ and $\mathbf{f}_2 - \mathbf{f}_1$. This is the fundamental impulse pair of the (1,-1)-moiré seen in the superposition (c). The present figure illustrates a one-dimensional setting, that has been extended to the two dimensional plane for an improved visibility. The true one-dimensional version of this figure consists of the cross sections passing through the horizontal axes of (a)-(f).

In the moiré theory, the moiré effect generated between $N$ periodic structures having frequencies $f_i$, $i = 1,...,N$ when $k_1f_1 + ... + k_Nf_N \approx 0$ is traditionally called the $(k_1,...,k_N)$-moiré effect, since it is generated by an interaction between the $k_1$-harmonic frequency of the first signal, the $k_2$-harmonic frequency of the second signal, and so forth (see Sec. 2.8 in [9]). For example, the moiré effect generated between two structures when $f_1 \approx f_2$, so that $f_1 - f_2 \approx 0$, is called a (1,-1)-moiré effect (Fig. A2), and the moiré effect generated between two structures when $f_1 \approx 2f_2$, so that $f_1 - 2f_2 \approx 0$, is called a (1,-2)-moiré effect (Fig. A3).

Note that the situation in the one-dimensional setting is analogous, except that all the frequency vectors in the spectral domain and all the periodicities in the signal domain only take place along the horizontal axis. The two-dimensional Fig. A4 illustrates a one-dimensional setting which has been artificially extended to the two dimensional plane. The true one-dimensional setting consists of a cross section through the horizontal axis in both signal and spectral domains. (Note however that this is not true for figures such as A1-A3, which indeed contain true two-dimensional information).

Now, in the case of *sampling moirés* (moiré effects that are generated when sampling a periodic signal $g(x)$ of frequency $f$ using a sampling frequency of $f_s$) only $N$=2 structures are involved, $g(x)$ and the sampling comb. Therefore, using the traditional moiré-theory notation, when $mf_s + nf \approx 0$ we obtain a $(m,n)$-sampling-moiré effect. This moiré is generated in the sampled signal $g(x_k)$ due to an interaction between the $m$-th harmonic of the sampling frequency $f_s$ and the $n$-th harmonic of $f$, the fundamental frequency of $g(x)$.[15] See, for example, the (1,-1)-sampling moiré in Figs. 2 and 7, where $f_s - f \approx 0$, and the (2,-1)-sampling moiré in Figs. 3 and 8, where $2f_s - f \approx 0$. (Note that in these figures the frequencies $f$ and $f_s$ fall outside the spectrum range, and only the folded over moiré frequencies are visible, near the origin).

As we have already seen throughout this paper (see, for example, the discussion before Remark 1), sampling-theory considerations can be often interpreted in a dual way using moiré theory considerations, and vice versa. This duality sheds a new light on the results in question thanks to the cross-fertilisation between these two points of view.

## Appendix B: The phase terminology for periodic functions

Let us introduce some notations and terms in connection with the phase of periodic functions.

Suppose we are given a cosine function $g(x) = \cos(2\pi fx)$ with frequency $f$ and period $p = 1/f$, and that we shift it by $a$ along the $x$ axis:

$$g(x+a) = \cos(2\pi\tfrac{1}{p}(x+a)) = \cos(2\pi\tfrac{1}{p}x + 2\pi\tfrac{1}{p}a) \qquad \text{(B.1)}$$

[15] Of course, this interaction may only occur if the spectrum (Fourier decomposition) of the signal $g(x)$ contains the $n$-th harmonic of the signal's frequency $f$. Otherwise, no sampling moiré may be generated here (although a $(m/n)$-order sub-Nyquist artifact *may* occur, since $mf_s + nf \approx 0$ is equivalent to $f \approx \frac{m}{n} f_s$).

We define $\varphi = 2\pi\frac{1}{p}a = 2\pi fa$ as the *phase* of our shifted cosine. We therefore have:

$$= \cos(2\pi\tfrac{1}{p}x + \varphi) \tag{B.2}$$

As we vary the shift $a$ from 0 to $p$, the phase $\varphi$ of our cosine varies from 0 to $2\pi$, and completes one full cycle of $2\pi$. Note that with respect to $\varphi$ our cosine is cyclical modulo $2\pi$: for any integer $k$, $\cos(2\pi\frac{1}{p}x + [\varphi + 2\pi k]) = \cos(2\pi\frac{1}{p}x + \varphi)$. Hence, a phase of $2\pi k$ is equivalent to the phase 0.

We now define the *relative shift* (or the *period shift*) $\phi = \frac{a}{p} = fa$ as the shift undergone by $g(x+a)$, expressed as a number (integer or not) of periods $p$ of $g(x)$. We therefore have from Eq. (B.1):

$$= \cos(2\pi\tfrac{1}{p}x + 2\pi\phi) \tag{B.3}$$

Note that with respect to the relative shift $\phi$ our shifted cosine is cyclical modulo 1: for any integer $k$, $\cos(2\pi\frac{1}{p}x + 2\pi[\phi + k]) = \cos(2\pi\frac{1}{p}x + 2\pi\phi)$. As we vary the shift $a$ from 0 to $p$, the relative shift $\phi$ varies from 0 to 1, completing one full cycle of the cosine. Hence, a relative shift of $\phi = k$ is equivalent to $\phi = 0$.

Because we have $\varphi = 2\pi\phi$ it is clear that $\varphi$ and $\phi$ are equivalent, and both of them may be used to denote the phase of any shifted periodic function $g(x+a)$. However, the phase $\varphi$ is mainly useful in functions such as sine or cosine, where the constant $2\pi$ has indeed an intrinsic significance. In other periodic functions such as a square wave, a triangular wave, etc., where $2\pi$ does not explicitly appear in the function definition itself, the use of $\phi$ to designate the phase may be more natural.

In the case of a general periodic function $g(x)$ having the Fourier series expansion

$$g(x) = \sum_{l=-\infty}^{\infty} c_l\, e^{i2\pi lfx} \tag{B.4}$$

the influence of a shift of $a$ on the phase of $g(x)$ is expressed in terms of the relative shift $\phi$ harmonic by harmonic, as follows:

$$g(x+a) = \sum_{l=-\infty}^{\infty} c_l\, e^{i2\pi lf(x+a)} = \sum_{l=-\infty}^{\infty} c_l\, e^{i2\pi lfx + i2\pi lfa}$$

$$= \sum_{l=-\infty}^{\infty} c_l\, e^{i2\pi lfx + i2\pi l\phi} \tag{B.5}$$

Note that here, too, with respect to the relative shift $\phi$ our shifted function (B.5) is cyclical modulo 1: for any integer $k$,

$$\sum_{l=-\infty}^{\infty} c_l\, e^{i2\pi lfx + i2\pi l[\phi+k]} = \sum_{l=-\infty}^{\infty} c_l\, e^{i2\pi lfx + i2\pi l\phi + i2\pi lk}$$

$$= \sum_{l=-\infty}^{\infty} c_l\, e^{i2\pi lfx + i2\pi l\phi}\; e^{i2\pi lk} = \sum_{l=-\infty}^{\infty} c_l\, e^{i2\pi lfx + i2\pi l\phi}$$

since $e^{i2\pi lk} = \cos(2\pi lk) + i\,\sin(2\pi lk) = 1$ for any integers $k$ and $l$ [12, p. 96]. As we vary the shift $a$ from 0 to $p$, the relative shift $\phi$ varies from 0 to 1, completing one full cycle. This result will be used in Appendix D.

More detailed information on the phase terminology for periodic functions can be found in [9, Secs. 7.3 and C.4]. Interested readers may also have their own hands-on

experience using the provided interactive applications, which allow one to gradually vary $\phi$ between 0 and 1 and observe the resulting effects both in the signal and in the spectral domains.

## Appendix C: Miscellaneous remarks

In this appendix we provide some further remarks that consolidate the results obtained so far, and shed new light on various aspects of the sub-Nyquist artifacts.

**Remark 4** (*omnipresence of modulating envelopes*):

It is important to note that modulating envelopes can be traced through the samples of our continuous-world signal in all circumstances, whether the signal's frequency $f$ is close to $(m/n)f_s$ or not (note that Theorems 1 and 2 hold for *any* value of $\varepsilon$, be it small or large). Moreover, for any given $f$ and $f_s$, different sets of modulating envelopes belonging to different $m$, $n$ values can be always traced through the very same samples $g(x_k)$, $k = 0, 1, 2,...$[16] Nevertheless, these modulating envelopes only become relevant when $f$ is sufficiently close to $(m/n)f_s$, i.e. when $\varepsilon = f - (m/n)f_s$ is small, thanks to the visible beating or ripple effect that is generated by the $(m/n)$-order sub-Nyquist artifact in these cases. Furthermore, since rational numbers $m/n$ can be always found that closely approximate any real number $r$, be it rational or irrational, it turns out that $(m/n)$-order modulation effects may occur at any frequency $f$ along the frequency axis. But of course, the larger the integer number $n$ (and hence the number of interlaced envelopes), the less visible and prominent the envelopes become. The reason is that when several envelopes are intermingled together it becomes more difficult for the eye to detect and follow each of the envelopes separately, i.e. to detect a visible order within the sampled points. Furthermore, in the case of a sampled cosinusoidal signal, in larger values of $n$ the depth of the ripple effect becomes smaller (see Remark 2 above) and hence the beats are less conspicuous. Thus, although modulating envelopes are always present in the sampled signal $g(x_k)$, in practice the resulting beating effect (sub-Nyquist artifact) is only visible for relatively low values of $n$. In other words, the sampled points of a periodic function $g(x)$ are always located on (i.e. they can always be decomposed into) various sets of interlaced envelopes; but these envelopes truly become visible only when the value of $n$ is relatively low.

We therefore obtain the following result: If we slowly vary the frequency $f$ of the original periodic function $g(x)$ and let it sweep along the frequency axis, we will only find a limited number of zones in which $f$ is close to a $(m/n)f_s$ value (singular point) with a small $n$, where the envelopes give a clearly visible beating effect. Although infinitely many singular points $(m/n)f_s$ exist throughout the frequency axis (in fact, they are everywhere dense along this axis), only few of them will give a "dangerous zone" that $f$ should avoid when sweeping along the frequency axis in order not to generate visible sub-Nyquist artifacts.

---

[16] Note that for any given $f$ and $f_S$ there exist infinitely many values of $m$,$n$ and $\varepsilon$ that satisfy Eq. (15), $f = (m/n) f_S + \varepsilon$. For any chosen $m$ and $n$, there exists a corresponding $\varepsilon$ value that satisfies the equation.

The dual result obtained by considering our question the other way around may be even more interesting: If we fix the frequency $f$ of the original periodic function $g(x)$ and let the sampling frequency $f_s$ slowly vary and sweep along the frequency axis, we will only find a limited number of "dangerous zones" of sampling frequencies $f_s$ that should be avoided when sampling our given periodic function $g(x)$. These "dangerous zones" occur when $f_s$ is located around one of the frequencies $(n/m)f$ with small integers $m$, $n$ (note that $f \approx (m/n)f_s$ is equivalent to $f_s \approx (n/m)f$). ■

**Remark 5** (*singular state of the (m/n)-order sub-Nyquist artifact*):

When the frequency $f$ of the original continuous-world function $g(x)$ equals exactly $(m/n)f_s$, i.e. when $\varepsilon = 0$, the period of the $(m/n)$-order sub-Nyquist artifact becomes infinitely large, and hence the corresponding beating or ripple is no longer visible and the sampled signal becomes uniform and periodic, as shown in rows (a) of Figs. 2 and on. We call this critical frequency the *singular state* of the $(m/n)$-order sub-Nyquist artifact, in analogy to the singular state of the $(m,n)$-moiré effect (see Appendix A, or Sec. 2.9 in [9]). And indeed, just like in the case of the moiré, as soon as the frequency $f$ starts moving away from the singular state (to either direction), the $(m/n)$-order sub-Nyquist artifact "comes back from infinity" and becomes again visible, with a long-period ripple. As $f$ moves further away from the singular state, i.e. as $|\varepsilon|$ gradually increases (see rows (b)-(e) in Figs. 2 and on), the period of the beating or ripple effect becomes smaller, until finally it becomes very small and completely disappears.

As the frequency $f$ sweeps along the frequency axis, it passes through the singular states of many different $(m/n)$-order sub-Nyquist artifacts. Each time, a similar scenario occurs as the frequency $f$ approaches the singular state, coincides with it, and then passes beyond it. Interested readers may get a vivid, dynamic demonstration using the provided interactive applications, which allow one to slowly vary the frequency $f$ and observe the resulting effects on the sampled signal. It should be noted, however, that the sub-Nyquist artifacts we thus meet on our way while varying $f$ are all different, depending on their $m,n$ values: They differ in their prominence, their number of modulating envelopes, their even or odd nature (see Remark 1), etc.

Because the singular state of an $(m/n)$-order sub-Nyquist artifact occurs at the frequency $f = (m/n)f_s$, it follows that for any frequency $f$ of the original function $g(x)$, close-by $(m/n)$-order singular states will always exist in the neighbourhood of $f$ for some integer pairs $m$, $n$ that satisfy $f \approx (m/n)f_s$, i.e. $m/n \approx f/f_s$. However, as explained in Remarks 2 and 4, most of the resulting $(m/n)$-order sub-Nyquist artifacts are negligible and hardly visible, since their $n$ value is relatively high. ■

**Remark 6** (*how to identify to which (m/n)-order belongs a given case*):

Suppose we are given a beating or ripple effect that occurs when sampling a periodic continuous-world function $g(x)$ of frequency $f$ using the sampling frequency $f_s$. How can we find out to which $(m/n)$-order sub-Nyquist artifact belongs our case? An easy, empirical way consists of the following steps: We first plot the given case (possibly after scaling the values of $f$ and $f_s$ to adapt them to our plotting conventions, but obviously using the same scaling factor for both frequencies so as to keep the ratio $f/f_s$

unchanged; this frequency scaling corresponds, in fact, to a horizontal stretching of the plot along the $x$ axis). Then, we slightly increment or decrement $f$ by a very small value $\delta$, and plot the resulting sampled signals having the frequencies $f + \delta$ and $f - \delta$. The aim of this step is to determine to which direction we must vary $f$ along the frequency axis in order to increase the beat's period (i.e. in order to get closer to the singular frequency of our given beating effect). We then continue varying $f$ very slowly in that direction, each time re-plotting the sampled signal, and we stop when the beats become infinitely large and hence invisible (note that beyond this critical value of $f$ the beats start to reappear once again). Having thus identified the singular frequency $f$ belonging to our case, we simply have to find which mutually prime integers $m$ and $n$ satisfy $m/n = f/f_s$. Note that different $(m/n)$-order cases may possess very close singular frequencies, so the step $\delta$ we are using must be very small in order not to miss the correct singular state; often several digits beyond the decimal point may be required.

As an illustrative example, let us apply this method to analyze the nature of the beating effect reported in [7], where $f_s = 12170$ and $f = 1367$. We first rescale these values to adapt them to our plotting conventions, in which we are using $f_s = 8$; this is done by taking $f = 1367 \times 8/12170 = 0.8986$. Now, by slowly varying $f$ and each time plotting the resulting sampled signal, as explained above, we find that the corresponding singular state occurs at $f = 0.888...$ Therefore we obtain $m/n = f/f_s = 0.111...$, and we conclude that $(m/n) = (1/9)$. This is, indeed, an odd-type ripple (see Remark 1), having 9 interlaced modulating envelopes. Note, however, that the case shown in Fig. 7 of [7] is different, and corresponds to an odd-type (1/5)-order sub-Nyquist artifact having 5 interlaced envelopes, as we can clearly see in that figure. ■

**Remark 7** (*coexistence of sub-Nyquist artifacts with leakage*):

Although the sub-Nyquist artifact is a discrete-world phenomenon, it is not affected by leakage (the leakage phenomenon is explained, for example, in [10, Ch. 6] or in [12, pp. 98-107]). In other words, the modulating envelopes in the signal domain remain perfectly accurate, without any flaws, even in cases where the frequency $f$ and/or the frequency $(m/n)f_s$ do not fall exactly on a discrete point of the DFT spectrum, but rather between two such points. See, for example, rows (b) and (c) in Fig. 5, in which the CFT impulses corresponding to the cosine frequency $f$ fall between DFT elements and suffer from leakage; or Fig. 6, in which the frequency $(1/3)f_s$, too, falls between DFT elements. In all of these cases the modulating envelopes in the image domain behave exactly as predicted, and they are not flawed by the existence of leakage. ■

**Remark 8** (*coexistence of sub-Nyquist artifacts with aliasing*):

Similarly, sub-Nyquist artifacts may also coexist with aliasing (folding-over). In other words, sub-Nyquist artifacts occur even when the frequency $f$ of the original function $g(x)$ (or its higher harmonics, if any) fall beyond half of the sampling frequency, i.e. beyond the boundaries of the DFT spectrum (note that the DFT spectrum always extends between minus and plus half of the sampling frequency [10, p. 73]). And indeed, the modulating envelopes we obtain thanks to the sub-Nyquist artifact remain identical for folded-over or non folded-over frequencies. This is illustrated, for example, in rows (f) and (h) of Fig. 1, which look identical although row (h) is a folded-over

counterpart of row (f).[17] Note that when $m > n/2$ so that the frequency $f \approx (m/n)f_s$ is higher than $0.5f_s$, which also includes all cases with $n = 1$ (true moiré effects), aliasing does occur, and the term "sub-Nyquist artifacts" is no longer adequate; but we will continue using it here, too, for consistency reasons and for want of a better term. It should be noted that when a folded-over frequency due to aliasing falls close to the spectrum origin, a true moiré effect becomes visible in the sampled signal (Fig. 1 (i),(j)). This occurs, for example, in the (1/1)-order or (2/1)-order cases. ■

**Remark 9** (*higher harmonics*):

We have seen that when $n = 1$, the sampled points $g(x_k)$, $k = 0, 1, 2,...$ fall on a single low-frequency curve, meaning that the $(m/1)$-order sub-Nyquist artifact is, indeed, a true $(m,-1)$-moiré effect. We know from the moiré theory (see Appendix A) that true moiré effects may also occur for higher values of $n$, whenever $mf_s + nf \approx 0$. However, this is only possible if the periodic function $g(x)$ contains the $n$-th harmonic of its frequency $f$. For example, a (1/2)-order sub-Nyquist artifact cannot be a true moiré effect when the function being sampled is $g(x) = \cos(2\pi fx)$, since the cosine function does not possess a second harmonic of $f$. But when the function being sampled is the square wave $g(x) = \text{wave}(fx)$ having frequency $f$ (see Fig. 9), which does have the second harmonic of $f$, the (1/2)-order sub-Nyquist artifact can be considered as a true (1,-2)-moiré. And indeed, looking at the spectral domain in Fig. 9, we see that the spectra (DFT) of the sampled signals do contain new low frequencies near the origin, that did not exist in the spectra (CFT) of the original continuous signals. These new low frequencies correspond to the true (1,-2)-moiré effect. But as expected, this does not occur in the cosinusoidal counterpart of Fig. 9, shown in Fig. 4, which remains a pure (1/2)-order sub-Nyquist artifact.

Interestingly, although the (1/2)-order sub-Nyquist artifact of the square wave (see Fig. 9) can be considered as a true moiré effect, the sampled points $g(x_k)$, $k = 0, 1, 2,...$ in its signal domain do jump alternately between two modulating envelopes. And indeed, the mere fact that $g(x_k)$ alternately jumps between several modulating envelopes does not yet mean that the artifact in question is not a moiré effect. In fact, we can say that our example is a *hybrid* case, sharing properties both of a moiré effect (existence of a true low frequency) and of a sub-Nyquist artifact (jumping alternately between several modulation envelopes). Comparing the sampled signals in Figs. 4 and 9 we can find the clue: In Fig. 4, the moving average of the jumpy sampled signal is identically zero, meaning that no low-frequency content is present in the sampled signal. But in Fig. 9, the moving average of the jumpy sampled signal *does* have a true low-frequency content, which precisely corresponds to a true moiré effect. ■

**Remark 10** (*on the notations used for moirés and sub-Nyquist artifacts*):

As we have seen throughout this work, sampling moirés and sub-Nyquist artifacts are particular cases of the same phenomenon, and they obey the same mathematical rules. So why don't we adopt the same notations for both cases?

---

[17] The (1/2)-sub-Nyquist artifact is indeed unique in that its singular point $(1/2)f_s$ exactly coincides with the Nyquist frequency $0.5f_s$, so that for $\varepsilon < 0$ no aliasing exists (Fig. 1(e),(f)), while for $\varepsilon > 0$ aliasing does occur (Fig. 1(h)).

In the moiré theory, the moiré effect generated between $N$ periodic structures having frequencies $f_i$, $i = 1,...,N$ when $k_1f_1 + ... + k_Nf_N \approx 0$ is traditionally called the $(k_1,...,k_N)$-moiré effect (see Appendix A, or Sec. 2.8 in [9]). For example, the moiré effect generated between two structures when $f_1 \approx f_2$, so that $f_1 - f_2 \approx 0$, is called the (1,-1)-moiré effect. Now, in the case of sampling moirés (moiré effects that are generated when sampling a periodic function $g(x)$ of frequency $f$ using a sampling frequency of $f_s$) only two structures are involved, $g(x)$ and the sampling comb. Using the traditional moiré-theory notation, when $mf_s + nf \approx 0$ we obtain a $(m,n)$-moiré effect. But unlike in cases with $N > 2$, when $N = 2$ it is the ratio $m/n$ that plays the main role, and the integer vector notation can be reduced into the $(m/n)$-notation. And indeed, this notation is particularly convenient and advantageous in the case of sub-Nyquist artifacts (note, for example, the explicit use of the ratio $m/n$ in Theorems 1 and 2). Nevertheless, when specifically referring to a moiré effect, we prefer to continue using the classical moiré notations, for reasons of consistency with the existing moiré literature. ■

## Appendix D: Derivation of Theorem 2

Let $g(x)$ be a periodic function with frequency $f$ and period $p = 1/f$. The Fourier series expansion of $g(x)$ is given, using exponential notation, by:

$$g(x) = \sum_{l=-\infty}^{\infty} c_l \, e^{i2\pi lfx} \tag{D.1}$$

where the $l$-th Fourier series coefficient $c_l$ is

$$c_l = (1/p) \int_p g(x) \, e^{-i2\pi lfx} \, dx$$

(see, for example, [9, Sec. A.1] or [20, pp. 4-6, 173-190]).

When we sample $g(x)$ at the sampling frequency $f_s$, i.e. using a sampling step of $\Delta x = 1/f_s$, we obtain the sampled signal:

$$g(x_k) = \sum_{l=-\infty}^{\infty} c_l \, e^{i2\pi lfx_k} \tag{D.2}$$

with $x_k = k\Delta x = k/f_s$.

Suppose, first, that the frequency of the original continuous function $g(x)$ is exactly

$$f = \tfrac{m}{n} f_s \tag{D.3}$$

In this case $p = \frac{n}{m}\Delta x$, meaning that we have exactly $\frac{n}{m}$ samples in each period $p$ of $g(x)$, and the sampling step $\Delta x$ is $\frac{m}{n}$ of the period $p$. The sampled signal we obtain in this case is:

$$g(x_k) = \sum_{l=-\infty}^{\infty} c_l \, e^{i2\pi l[(m/n)f_s]k/f_s} = \sum_{l=-\infty}^{\infty} c_l \, e^{i2\pi lk(m/n)} \tag{D.4}$$

This corresponds, indeed, to the singular state of the $(m/n)$-order sub-Nyquist artifact, which is systematically plotted in Figs. 2 and on in row (a).

Now, suppose that the frequency of the given function $g(x)$ is not exactly $f = \frac{m}{n} f_s$, but rather:

$$f = \frac{m}{n} f_s + \varepsilon \qquad \text{(D.5)}$$

(where $\varepsilon$ may be positive or negative). The sampling frequency $f_s$ remains unchanged, so that we still have $x_k = k\Delta x = k/f_s$. The sampled signal is, in this case:

$$\begin{aligned} g(x_k) &= \sum_{l=-\infty}^{\infty} c_l \, e^{i2\pi l[(m/n)f_s + \varepsilon]k/f_s} \\ &= \sum_{l=-\infty}^{\infty} c_l \, e^{i2\pi l[k(m/n) + \varepsilon k/f_s]} \\ &= \sum_{l=-\infty}^{\infty} c_l \, e^{i2\pi l[x_k \varepsilon + \phi]} \qquad \text{with } \phi = k\frac{m}{n} \\ &= \sum_{l=-\infty}^{\infty} c_l \, e^{i2\pi l\varepsilon[x_k + (\phi/\varepsilon)]} \end{aligned} \qquad \text{(D.6)}$$

As we can see by comparing with Eqs. (D.1) and (D.2), this is a sampled version of the counterpart of $g(x)$ having frequency $\varepsilon$ (i.e. period $1/\varepsilon$) and a nominal shift of $\frac{\phi}{\varepsilon}$, i.e. a *relative shift* of $\phi$ periods (Appendix B gives the definition of this term, and shows that in terms of $\phi$ Eq. (D.6) is cyclic modulo 1).

This means that:

(1) For integer $\phi = k\frac{m}{n}$, namely for $k = 0, n, 2n, 3n, \ldots$ we have:

$$g(x_k) = \sum_{l=-\infty}^{\infty} c_l \, e^{i2\pi l\varepsilon[x_k + 0]}$$

This signal has a shift of 0 (or of an integer multiple of its period $\frac{1}{\varepsilon}$).

(2) For $k = 1, n+1, 2n+1, 3n+1, \ldots$ we have:

$$g(x_k) = \sum_{l=-\infty}^{\infty} c_l \, e^{i2\pi l\varepsilon[x_k + (m/n)(1/\varepsilon)]}$$

This signal has a shift of $\frac{m}{n}$ times its period $\frac{1}{\varepsilon}$.

(3) For $k = 2, n+2, 2n+2, 3n+2, \ldots$ we have:

$$g(x_k) = \sum_{l=-\infty}^{\infty} c_l \, e^{i2\pi l\varepsilon[x_k + 2(m/n)(1/\varepsilon)]}$$

This signal has a shift of $2\frac{m}{n}$ times its period $\frac{1}{\varepsilon}$.

. . .

($n$) For $k = n-1, 2n-1, 3n-1, 4n-1, \ldots$ we have:

$$g(x_k) = \sum_{l=-\infty}^{\infty} c_l \, e^{i2\pi l\varepsilon[x_k + (n\text{-}1)(m/n)(1/\varepsilon)]}$$

This signal has a shift of $(n-1)\frac{m}{n}$ times its period $\frac{1}{\varepsilon}$.

This means that the successive sampled points of our original function, $g(x_k)$, $k = 0, 1, 2,\ldots$ fall intermittently on one of $n$ interlaced $g(x)$-shaped curves (that we call *envelopes*), which have all the same frequency $\varepsilon$ and period $1/\varepsilon$, and which only differ from each other in their phase. More precisely, these $n$ envelopes only differ from each other by successive relative shifts of $\frac{m}{n}$ periods $1/\varepsilon$, i.e. by successive shifts of $a = \frac{m}{n\varepsilon}$. This corresponds, indeed, to the situation we see in our figures in rows (b)-(e).

We have thus obtained Theorem 2.

# References


[1] D. P. Mitchell and A. N. Netravali, "Reconstruction filters in computer graphics," Proceedings of SIGGRAPH 1988, *Computer Graphics*, Vol. 22, 1988, pp. 221–228.

[2] J. D. Foley, A. van Dam, S. K. Feiner and J. F. Hughes, *Computer Graphics: Principles and Practice.* Addison-Wesely, Reading, Massachusetts, 1990 (second edition).

[3] G. Fielding, R. Hsu, P. Jones and C. DuMont, "Aliasing and reconstruction distortion in digital intermediates," *SMPTE Motion Imaging Journal*, Vol. 115, 2006, pp. 128–136.

[4] http://devmaster.net/posts/12295/nyquist-theorem-oddity - "Nyquist theorem oddity" (accessed March 20, 2014).

[5] http://www.hydrogenaudio.org/forums/index.php?showtopic=8177&mode=threaded&pid=86676 - "Nyquist was wrong?!" (accessed March 20, 2014).

[6] http://dsp.stackexchange.com/questions/9986/where-can-i-find-an-authoritative-peer-reviewed-or-textbook-reference-to-sampl - "Where can I find an authoritative (peer-reviewed or textbook) reference to sampling-induced beating?" (accessed March 20, 2014).

[7] G. L. Williams, "Sub-Nyquist distortions in sampled data, waveform recording, and video imaging," NASA Technical Memorandum TM-2000-210381, Ohio, 2000.

[8] M. Schwartz, *Information, Transmission, Modulation and Noise*. McGraw-Hill, NY, 1990 (fourth edition).

[9] I. Amidror, *The Theory of the Moiré Phenomenon, Volume I: Periodic Layers*. Springer, NY, 2009 (second edition).

[10] I. Amidror, *Mastering the Discrete Fourier Transform in One, Two or Several Dimensions: Pitfalls and Artifacts*. Springer, NY, 2013.

[11] I. Amidror and R. D. Hersch, "The role of Fourier theory and of modulation in the prediction of visible moiré effects," *Journal of Modern Optics*, Vol. 56, 2009, pp. 1103–1118.

[12] E. O. Brigham, *The Fast Fourier Transform and Its Applications.* Prentice-Hall, NJ, 1988.

[13] J. D. Gaskill, *Linear Systems, Fourier Transforms, and Optics*. John Wiley & Sons, NY, 1978.

[14] R. W. Ramirez, *The FFT: Fundamentals and Concepts*. Prentice Hall, NJ, 1985.

[15] E. W. Weisstein, *CRC Concise Encyclopedia of Mathematics*. CRC, Boca Raton, 1999.

[16] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, Vol. 52, 2006, pp. 1289–1306.

[17] E. J. Candès, "Compressive sampling," *Proceedings of the International Congress of Mathematicians,* Madrid, 2006, pp. 1433–1452. Available online at: http://statweb.stanford.edu/~candes/papers/CompressiveSampling.pdf (accessed August 22, 2014).

[18] R. Eschbach, "Generation of moiré by nonlinear transfer characteristics," *Jour. of the Optical Society of America A*, Vol. 5, 1988, pp. 1828–1835.

[19] R. N. Bracewell, *The Fourier Transform and its Applications*. McGraw-Hill, Boston, 2000 (third edition).

[20] D. M. Kammler, *A First Course in Fourier Analysis*. Cambridge University Press, Cambridge, 2007 (revised edition).

[21] I. Amidror, "Sub-Nyquist artefacts and sampling moiré effects," *Royal Society Open Science*, 2: 140550, 2015, pp. 1–15. Available online at: http://dx.doi.org/10.1098/rsos.140550 (accessed March 18, 2015).

[22] I. Amidror, "Sub-Nyquist artifacts and sampling moiré effects, Part 2: Spectral-domain analysis," Technical Report No. 206457, EPFL, Lausanne, 2015. Available online at: http://infoscience.epfl.ch/record/206457 (accessed March 18, 2015).

**Acknowledgements:**

A significantly shortened version of this tutorial has been published in the *Royal Society Open Science*, 2015 [21].

A complementary discussion based on spectral-domain considerations can be found in [22].