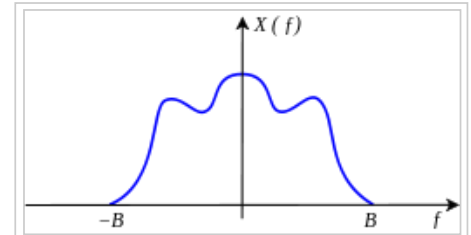


Nyquist–Shannon sampling theorem

From Wikipedia, the free encyclopedia

In the field of digital signal processing, the **sampling theorem** is a fundamental bridge between continuous-time signals (often called "analog signals") and discrete-time signals (often called "digital signals"). It establishes a sufficient condition for a sample rate that permits a discrete sequence of *samples* to capture all the information from a continuous-time signal of finite bandwidth.

Strictly speaking, the theorem only applies to a class of mathematical functions having a Fourier transform that is zero outside of a finite region of frequencies. Intuitively we expect that when one reduces a continuous function to a discrete sequence and interpolates back to a continuous function, the fidelity of the result depends on the density (or sample rate) of the original samples. The sampling theorem introduces the concept of a sample rate that is sufficient for perfect fidelity for the class of functions that are bandlimited to a given bandwidth, such that no actual information is lost in the sampling process. It expresses the sufficient sample rate in terms of the bandwidth for the class of functions. The theorem also leads to a formula for perfectly reconstructing the original continuous-time function from the samples.



Example of magnitude of the Fourier transform of a bandlimited function

Perfect reconstruction may still be possible when the sample-rate criterion is not satisfied, provided other constraints on the signal are known. (See § Sampling of non-baseband signals below, and compressed sensing.) In some cases (when the sample-rate criterion is not satisfied), utilizing additional constraints allows for approximate reconstructions. The fidelity of these reconstructions can be verified and quantified utilizing Bochner's theorem.^[1]

The name *Nyquist–Shannon sampling theorem* honors Harry Nyquist and Claude Shannon. The theorem was also discovered independently by E. T. Whittaker, by Vladimir Kotelnikov, and by others. It is thus also known by the names *Nyquist–Shannon–Kotelnikov*, *Whittaker–Shannon–Kotelnikov*, *Whittaker–Nyquist–Kotelnikov–Shannon*, and *cardinal theorem of interpolation*.

Contents

- 1 Introduction
- 2 Aliasing
- 3 Derivation as a special case of Poisson summation
- 4 Shannon's original proof
 - 4.1 Notes
- 5 Application to multivariable signals and images
- 6 Critical frequency
- 7 Sampling of non-baseband signals
- 8 Nonuniform sampling

- 9 Sampling below the Nyquist rate under additional restrictions
- 10 Historical background
 - 10.1 Other discoverers
 - 10.2 Why Nyquist?
- 11 See also
- 12 Notes
- 13 References
- 14 Further reading
- 15 External links

Introduction

Sampling is a process of converting a signal (for example, a function of continuous time and/or space) into a numeric sequence (a function of discrete time and/or space). Shannon's version of the theorem states:^[2]

If a function $x(t)$ contains no frequencies higher than B hertz, it is completely determined by giving its ordinates at a series of points spaced $1/(2B)$ seconds apart.

A sufficient sample-rate is therefore $2B$ samples/second, or anything larger. Equivalently, for a given sample rate f_s , perfect reconstruction is guaranteed possible for a bandlimit $B < f_s/2$.

When the bandlimit is too high (or there is no bandlimit), the reconstruction exhibits imperfections known as aliasing. Modern statements of the theorem are sometimes careful to explicitly state that $x(t)$ must contain no sinusoidal component at exactly frequency B , or that B must be strictly less than $\frac{1}{2}$ the sample rate. The two thresholds, $2B$ and $f_s/2$ are respectively called the **Nyquist rate** and **Nyquist frequency**. And respectively, they are attributes of $x(t)$ and of the sampling equipment. The condition described by these inequalities is called the **Nyquist criterion**, or sometimes the *Raabe condition*. The theorem is also applicable to functions of other domains, such as *space*, in the case of a digitized image. The only change, in the case of other domains, is the units of measure applied to t , f_s , and B .

The symbol $T = 1/f_s$ is customarily used to represent the interval between samples and is called the **sample period** or **sampling interval**. And the samples of function $x(t)$ are commonly denoted by $x[n] = x(nT)$ (alternatively " x_n " in older signal processing literature), for all integer values of n .

A mathematically ideal way to interpolate the sequence involves the use of sinc functions. Each sample in the sequence is replaced by a sinc function, centered on the time axis at the original location of the sample, nT , with the amplitude of the sinc function scaled to the sample value, $x[n]$. Subsequently, the sinc functions are summed into a continuous function. A mathematically equivalent method is to convolve one sinc function with a series of Dirac delta pulses, weighted

by the sample values. Neither method is numerically practical. Instead, some type of approximation of the sinc functions, finite in length, is used. The imperfections attributable to the approximation are known as *interpolation error*.

Practical digital-to-analog converters produce neither scaled and delayed sinc functions, nor ideal Dirac pulses. Instead they produce a piecewise-constant sequence of scaled and delayed rectangular pulses (the zero-order hold), usually followed by an "anti-imaging filter" to clean up spurious high-frequency content.

Aliasing

When $x(t)$ is a function with a Fourier transform, $X(f)$:

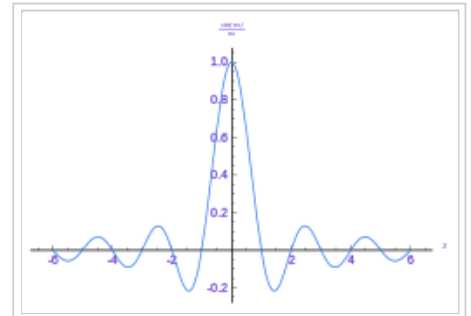
$$X(f) \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} x(t) e^{-i2\pi ft} dt,$$

the Poisson summation formula indicates that the samples, $x(nT)$, of $x(t)$ are sufficient to create a periodic summation of $X(f)$. The result is:

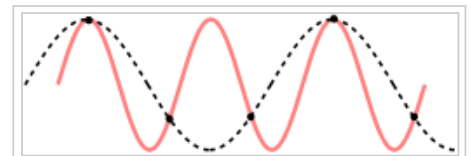
$$X_s(f) \stackrel{\text{def}}{=} \sum_{k=-\infty}^{\infty} X(f - kf_s) = \sum_{n=-\infty}^{\infty} T \cdot x(nT) e^{-i2\pi nTf}, \quad (\text{Eq. 1})$$

which is a periodic function and its equivalent representation as a Fourier series, whose coefficients are $T \cdot x(nT)$. This function is also known as the discrete-time Fourier transform (DTFT) of the sequence $T \cdot x(nT)$, for integers n .

As depicted, copies of $X(f)$ are shifted by multiples of f_s and combined by addition. For a band-limited function ($X(f) = 0$ for all $|f| \geq B$), and sufficiently large f_s it is possible for the copies to remain distinct from each other. But if the Nyquist criterion is not satisfied, adjacent copies overlap, and it is not possible in general to discern an unambiguous $X(f)$. Any frequency component above $f_s/2$ is indistinguishable from a lower-frequency component, called an *alias*, associated with one of the copies. In such cases, the customary interpolation techniques produce the alias, rather than the original component. When the sample-rate is pre-determined by other considerations (such as an industry standard), $x(t)$ is usually filtered to reduce its high frequencies to acceptable levels before it is sampled. The type of filter required is a lowpass filter, and in this application it is called an anti-aliasing filter.



The normalized sinc function: $\sin(\pi x) / (\pi x)$... showing the central peak at $x=0$, and zero-crossings at the other integer values of x .



The samples of two sine waves can be identical when at least one of them is at a frequency above half the sample rate.

Derivation as a special case of Poisson summation

When there is no overlap of the copies (aka "images") of $X(f)$, the $k = 0$ term of $X_s(f)$ can be recovered by the product:

$$X(f) = H(f) \cdot X_s(f), \quad \text{where:}$$

$$H(f) \stackrel{\text{def}}{=} \begin{cases} 1 & |f| < B \\ 0 & |f| > f_s - B. \end{cases}$$

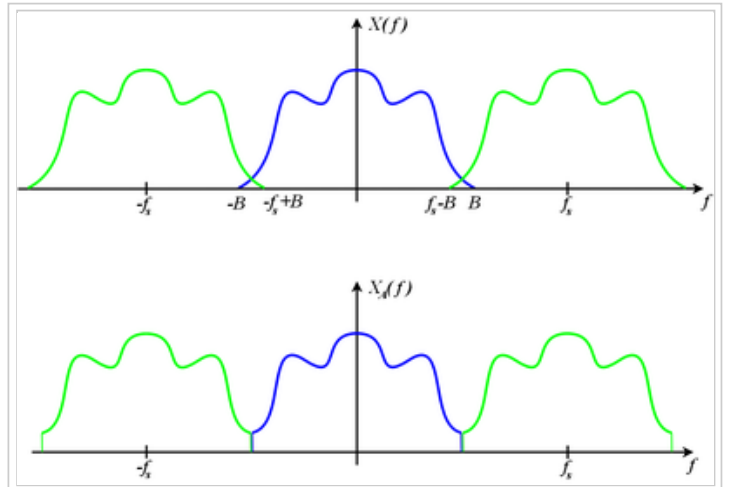
At this point, the sampling theorem is proved, since $X(f)$ uniquely determines $x(t)$.

All that remains is to derive the formula for reconstruction. $H(f)$ need not be precisely defined in the region $[B, f_s - B]$ because $X_s(f)$ is zero in that region. However, the worst case is when $B = f_s/2$, the Nyquist frequency. A function that is sufficient for that and all less severe cases is:

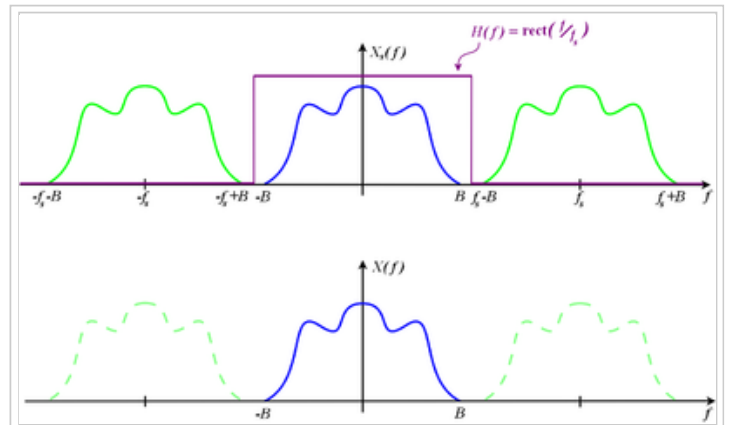
$$H(f) = \text{rect}\left(\frac{f}{f_s}\right) = \begin{cases} 1 & |f| < \frac{f_s}{2} \\ 0 & |f| > \frac{f_s}{2}, \end{cases}$$

where $\text{rect}(\cdot)$ is the rectangular function. Therefore:

$$\begin{aligned} X(f) &= \text{rect}\left(\frac{f}{f_s}\right) \cdot X_s(f) \\ &= \text{rect}(Tf) \cdot \sum_{n=-\infty}^{\infty} T \cdot x(nT) e^{-i2\pi nTf} \quad (\text{from Eq.1, above}). \end{aligned}$$



$X(f)$ (top blue) and $X_A(f)$ (bottom blue) are continuous Fourier transforms of two *different* functions, $x(t)$ and $x_A(t)$ (not shown). When the functions are sampled at rate f_s , the images (green) are added to the original transforms (blue) when one examines the discrete-time Fourier transforms (DTFT) of the sequences. In this hypothetical example, the DTFTs are identical, which means *the sampled sequences are identical*, even though the original continuous pre-sampled functions are not. If these were audio signals, $x(t)$ and $x_A(t)$ might not sound the same. But their samples (taken at rate f_s) are identical and would lead to identical reproduced sounds; thus $x_A(t)$ is an alias of $x(t)$ at this sample rate.



Spectrum, $X_s(f)$, of a properly sampled bandlimited signal (blue) and the adjacent DTFT images (green) that do not overlap. A *brick-wall*/low-pass filter, $H(f)$, removes the images, leaves the original spectrum, $X(f)$, and recovers the original signal from its samples.

$$= \sum_{n=-\infty}^{\infty} x(nT) \cdot T \cdot \underbrace{\text{rect}(Tf) \cdot e^{-i2\pi nTf}}_{\mathcal{F}\left\{\text{sinc}\left(\frac{t-nT}{T}\right)\right\}}. \quad [\text{note 1}]$$

The inverse transform of both sides produces the Whittaker–Shannon interpolation formula:

$$x(t) = \sum_{n=-\infty}^{\infty} x(nT) \cdot \text{sinc}\left(\frac{t-nT}{T}\right),$$

which shows how the samples, $x(nT)$, can be combined to reconstruct $x(t)$.

- Larger-than-necessary values of f_s (smaller values of T), called *oversampling*, have no effect on the outcome of the reconstruction and have the benefit of leaving room for a *transition band* in which $H(f)$ is free to take intermediate values. Undersampling, which causes aliasing, is not in general a reversible operation.
- Theoretically, the interpolation formula can be implemented as a low pass filter, whose impulse response is $\text{sinc}(t/T)$ and whose input is $\sum_{n=-\infty}^{\infty} x(nT) \cdot \delta(t-nT)$, which is a Dirac comb function modulated by the signal samples. Practical digital-to-analog converters (DAC) implement an approximation like the zero-order hold. In that case, oversampling can reduce the approximation error.

Shannon's original proof

Poisson shows that the Fourier series in **Eq. 1** produces the periodic summation of $X(f)$, regardless of f_s and B . Shannon, however, only derives the series coefficients for the case $f_s = 2B$. Virtually quoting Shannon's original paper:

Let $X(\omega)$ be the spectrum of $x(t)$. Then

$$\begin{aligned} x(t) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} X(\omega) e^{i\omega t} d\omega \\ &= \frac{1}{2\pi} \int_{-2\pi B}^{2\pi B} X(\omega) e^{i\omega t} d\omega \end{aligned}$$

since $X(\omega)$ is assumed to be zero outside the band $|\frac{\omega}{2\pi}| < B$. If we let

$$t = \frac{n}{2B}$$

where n is any positive or negative integer, we obtain

$$x\left(\frac{n}{2B}\right) = \frac{1}{2\pi} \int_{-2\pi B}^{2\pi B} X(\omega) e^{i\omega \frac{n}{2B}} d\omega.$$

On the left are values of $x(t)$ at the sampling points. The integral on the right will be recognized as essentially^[n 1] the n^{th} coefficient in a Fourier-series expansion of the function $X(\omega)$, taking the interval $-B$ to B as a fundamental period. This means that the values of the samples $x(n/2B)$ determine the Fourier coefficients in the series expansion of $X(\omega)$. Thus they determine $X(\omega)$, since $X(\omega)$ is zero for frequencies greater than B , and for lower frequencies $X(\omega)$ is determined if its Fourier coefficients are determined. But $X(\omega)$ determines the original function $x(t)$ completely, since a function is determined if its spectrum is known. Therefore the original samples determine the function $x(t)$ completely.

Shannon's proof of the theorem is complete at that point, but he goes on to discuss reconstruction via sinc functions, what we now call the Whittaker–Shannon interpolation formula as discussed above. He does not derive or prove the properties of the sinc function, but these would have been familiar to engineers reading his works at the time, since the Fourier pair relationship between rect (the rectangular function) and sinc was well known.

Let x_n be the n^{th} sample. Then the function $x(t)$ is represented by:

$$x(t) = \sum_{n=-\infty}^{\infty} x_n \frac{\sin \pi(2Bt - n)}{\pi(2Bt - n)}.$$

As in the other proof, the existence of the Fourier transform of the original signal is assumed, so the proof does not say whether the sampling theorem extends to bandlimited stationary random processes.

Notes

1. The actual coefficient formula contains an additional factor of $1/2B=T$. So Shannon's coefficients are $T \cdot x(nT)$, which agrees with **Eq.1**.

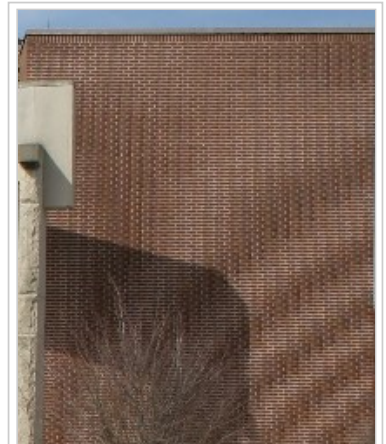
Application to multivariable signals and images

The sampling theorem is usually formulated for functions of a single variable. Consequently, the theorem is directly applicable to time-dependent signals and is normally formulated in that context. However, the sampling theorem can be extended in a straightforward way to functions of arbitrarily many variables. Grayscale images, for example, are often represented as two-dimensional arrays (or matrices) of real numbers representing the relative intensities of pixels (picture elements) located at the intersections of row and column sample locations. As a result, images require two independent variables, or indices, to specify each pixel uniquely—one for the row, and one for the column.

Color images typically consist of a composite of three separate grayscale images, one to represent each of the three primary colors—red, green, and blue, or *RGB* for short. Other colorspace using 3-vectors for colors include HSV, CIELAB, XYZ, etc. Some colorspace such as cyan, magenta, yellow, and black (CMYK) may represent color by four dimensions. All of these are treated as vector-valued functions over a two-dimensional sampled domain.

Similar to one-dimensional discrete-time signals, images can also suffer from aliasing if the sampling resolution, or pixel density, is inadequate. For example, a digital photograph of a striped shirt with high frequencies (in other words, the distance between the stripes is small), can cause aliasing of the shirt when it is sampled by the camera's image sensor. The aliasing appears as a moiré pattern. The "solution" to higher sampling in the spatial domain for this case would be to move closer to the shirt, use a higher resolution sensor, or to optically blur the image before acquiring it with the sensor.

Another example is shown to the right in the brick patterns. The top image shows the effects when the sampling theorem's condition is not satisfied. When software rescales an image (the same process that creates the thumbnail shown in the lower image) it, in effect, runs the image



Subsampled image showing a Moiré pattern



Properly sampled image

through a low-pass filter first and then downsamples the image to result in a smaller image that does not exhibit the moiré pattern. The top image is what happens when the image is downsampled without low-pass filtering: aliasing results.

The application of the sampling theorem to images should be made with care. For example, the sampling process in any standard image sensor (CCD or CMOS camera) is relatively far from the ideal sampling which would measure the image intensity at a single point. Instead these devices have a relatively large sensor area at each sample point in order to obtain sufficient amount of light. In other words, any detector has a finite-width point spread function. The analog optical image intensity function which is sampled by the sensor device is not in general bandlimited, and the non-ideal sampling is itself a useful type of low-pass filter, though not always sufficient to remove enough high frequencies to sufficiently reduce aliasing. When the area of the sampling spot (the size of the pixel sensor) is not large enough to provide sufficient spatial anti-aliasing, a separate anti-aliasing filter (optical low-pass filter) is typically included in a camera system to further blur the optical image. Despite images having these problems in relation to the sampling theorem, the theorem can be used to describe the basics of down and up sampling of images.

Critical frequency

To illustrate the necessity of $f_s > 2B$, consider the family of sinusoids generated by different values of θ in this formula:

$$x(t) = \frac{\cos(2\pi Bt + \theta)}{\cos(\theta)} = \cos(2\pi Bt) - \sin(2\pi Bt) \tan(\theta), \quad -\pi/2 < \theta < \pi/2.$$

With $f_s = 2B$ or equivalently $T = 1/(2B)$, the samples are given by:

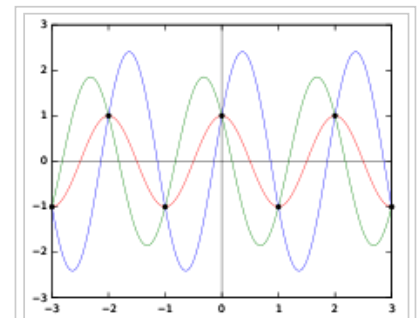
$$x(nT) = \cos(\pi n) - \underbrace{\sin(\pi n)}_0 \tan(\theta) = (-1)^n$$

regardless of the value of θ . That sort of ambiguity is the reason for the *strict* inequality of the sampling theorem's condition.

Sampling of non-baseband signals

As discussed by Shannon:^[2]

A similar result is true if the band does not start at zero frequency but at some higher value, and can be proved by a linear translation (corresponding physically to single-sideband modulation) of the



A family of sinusoids at the critical frequency, all having the same sample sequences of alternating +1 and -1. That is, they all are aliases of each other, even though their frequency is not above half the sample rate.

zero-frequency case. In this case the elementary pulse is obtained from $\sin(x)/x$ by single-side-band modulation.

That is, a sufficient no-loss condition for sampling signals that do not have baseband components exists that involves the *width* of the non-zero frequency interval as opposed to its highest frequency component. See *Sampling (signal processing)* for more details and examples.

For example, in order to sample the FM radio signals in the frequency range of 100-102 MHz, it is not necessary to sample at 204 MHz (twice the upper frequency), but rather it is sufficient to sample at 4 MHz (twice the width of the frequency interval).

A bandpass condition is that $X(f) = 0$, for all nonnegative f outside the open band of frequencies:

$$\left(\frac{N}{2} f_s, \frac{N+1}{2} f_s \right),$$

for some nonnegative integer N . This formulation includes the normal baseband condition as the case $N=0$.

The corresponding interpolation function is the impulse response of an ideal brick-wall bandpass filter (as opposed to the ideal brick-wall lowpass filter used above) with cutoffs at the upper and lower edges of the specified band, which is the difference between a pair of lowpass impulse responses:

$$(N+1) \operatorname{sinc}\left(\frac{(N+1)t}{T}\right) - N \operatorname{sinc}\left(\frac{Nt}{T}\right).$$

Other generalizations, for example to signals occupying multiple non-contiguous bands, are possible as well. Even the most generalized form of the sampling theorem does not have a provably true converse. That is, one cannot conclude that information is necessarily lost just because the conditions of the sampling theorem are not satisfied; from an engineering perspective, however, it is generally safe to assume that if the sampling theorem is not satisfied then information will most likely be lost.

Nonuniform sampling

The sampling theory of Shannon can be generalized for the case of nonuniform sampling, that is, samples not taken equally spaced in time. The Shannon sampling theory for non-uniform sampling states that a band-limited signal can be perfectly reconstructed from its samples if the average sampling rate satisfies the Nyquist condition.^[3] Therefore, although uniformly spaced samples may result in easier reconstruction algorithms, it is not a necessary condition for perfect reconstruction.

The general theory for non-baseband and nonuniform samples was developed in 1967 by Landau.^[4] He proved that the average sampling rate (uniform or otherwise) must be twice the *occupied* bandwidth of the signal, assuming it is *a priori* known what portion of the spectrum was occupied. In the late 1990s, this work was partially extended to cover signals of when the amount of occupied bandwidth was known, but the actual occupied portion of the spectrum was unknown.^[5] In the 2000s, a complete theory was developed (see the section Beyond Nyquist below) using compressed sensing. In particular, the theory, using signal processing language, is described in this 2009 paper.^[6] They show, among other things, that if the frequency locations are unknown, then it is necessary to sample at least at twice the Nyquist criteria; in other words, you must pay at least a factor of 2 for not knowing the location of the spectrum. Note that minimum sampling requirements do not necessarily guarantee stability.

Sampling below the Nyquist rate under additional restrictions

The Nyquist–Shannon sampling theorem provides a sufficient condition for the sampling and reconstruction of a band-limited signal. When reconstruction is done via the Whittaker–Shannon interpolation formula, the Nyquist criterion is also a necessary condition to avoid aliasing, in the sense that if samples are taken at a slower rate than twice the band limit, then there are some signals that will not be correctly reconstructed. However, if further restrictions are imposed on the signal, then the Nyquist criterion may no longer be a necessary condition.

A non-trivial example of exploiting extra assumptions about the signal is given by the recent field of compressed sensing, which allows for full reconstruction with a sub-Nyquist sampling rate. Specifically, this applies to signals that are sparse (or compressible) in some domain. As an example, compressed sensing deals with signals that may have a low over-all bandwidth (say, the *effective* bandwidth EB), but the frequency locations are unknown, rather than all together in a single band, so that the passband technique doesn't apply. In other words, the frequency spectrum is sparse. Traditionally, the necessary sampling rate is thus $2B$. Using compressed sensing techniques, the signal could be perfectly reconstructed if it is sampled at a rate slightly lower than $2EB$. The downside of this approach is that reconstruction is no longer given by a formula, but instead by the solution to a convex optimization program which requires well-studied but nonlinear methods.

Another example where sub-Nyquist sampling is optimal arises under the additional constraint that the samples are quantized in an optimal manner, as in a combined system of sampling and optimal lossy compression.^[7] This setting is relevant in cases where the joint effect of sampling and quantization is to be considered, and can provide a lower bound for the minimal reconstruction error that can be attained in sampling and quantizing a random signal. For stationary Gaussian random signals, this lower bound is usually attained at a sub-Nyquist sampling rate, indicating that sub-Nyquist sampling is optimal for this signal model under optimal quantization.^[8]

Historical background

The sampling theorem was implied by the work of Harry Nyquist in 1928,^[9] in which he showed that up to $2B$ independent pulse samples could be sent through a system of bandwidth B , but he did not explicitly consider the problem of sampling and reconstruction of continuous signals.

About the same time, Karl Küpfmüller showed a similar result,^[10] and discussed the sinc-function impulse response of a band-limiting filter, via its integral, the step response *Integralsinus*; this bandlimiting and reconstruction filter that is so central to the sampling theorem is sometimes referred to as a *Küpfmüller filter* (but seldom so in English).

The sampling theorem, essentially a dual of Nyquist's result, was proved by Claude E. Shannon.^[2] V. A. Kotelnikov published similar results in 1933,^[11] as did the mathematician E. T. Whittaker in 1915,^[12] J. M. Whittaker in 1935,^[13] and Gabor in 1946 ("Theory of communication"). In 1999, the Eduard Rhein Foundation awarded Kotelnikov their Basic Research Award "for the first theoretically exact formulation of the sampling theorem."

In 1948 and 1949, Claude E. Shannon published the two revolutionary papers in which he founded the information theory.^{[14][15][2]} In Shannon 1948 the sampling theorem is formulated as "Theorem 13": Let $f(t)$ contain no frequencies over W . Then

$$f(t) = \sum_{n=-\infty}^{\infty} X_n \frac{\sin \pi(2Wt - n)}{\pi(2Wt - n)},$$

where $X_n = f(n/2W)$. It was not until these papers were published that the theorem known as "Shannon's sampling theorem" became common property among communication engineers, although Shannon himself writes that this is a fact which is common knowledge in the communication art.^[note 2] A few lines further on, however, he adds: ... "but in spite of its evident importance [it] seems not to have appeared explicitly in the literature of communication theory".

Other discoverers

Others who have independently discovered or played roles in the development of the sampling theorem have been discussed in several historical articles, for example by Jerri^[16] and by Lüke.^[17] For example, Lüke points out that H. Raabe, an assistant to Küpfmüller, proved the theorem in his 1939 Ph.D. dissertation; the term *Raabe condition* came to be associated with the criterion for unambiguous representation (sampling rate greater than twice the bandwidth). Meijering^[18] mentions several other discoverers and names in a paragraph and pair of footnotes:

As pointed out by Higgins [135], the sampling theorem should really be considered in two parts, as done above: the first stating the fact that a bandlimited function is completely determined by its samples, the second describing how to reconstruct the function using its samples. Both parts of the sampling theorem were given in a somewhat different form by J. M. Whittaker [350, 351, 353] and before him also by Ogura [241, 242]. They were probably not aware of the fact that the first part of the

theorem had been stated as early as 1897 by Borel [25].²⁷ As we have seen, Borel also used around that time what became known as the cardinal series. However, he appears not to have made the link [135]. In later years it became known that the sampling theorem had been presented before Shannon to the Russian communication community by Kotel'nikov [173]. In more implicit, verbal form, it had also been described in the German literature by Raabe [257]. Several authors [33, 205] have mentioned that Someya [296] introduced the theorem in the Japanese literature parallel to Shannon. In the English literature, Weston [347] introduced it independently of Shannon around the same time.²⁸

²⁷ Several authors, following Black [16], have claimed that this first part of the sampling theorem was stated even earlier by Cauchy, in a paper [41] published in 1841. However, the paper of Cauchy does not contain such a statement, as has been pointed out by Higgins [135].

²⁸ As a consequence of the discovery of the several independent introductions of the sampling theorem, people started to refer to the theorem by including the names of the aforementioned authors, resulting in such catchphrases as "the Whittaker–Kotel'nikov–Shannon (WKS) sampling theorem" [155] or even "the Whittaker–Kotel'nikov–Raabe–Shannon–Someya sampling theorem" [33]. To avoid confusion, perhaps the best thing to do is to refer to it as the sampling theorem, "rather than trying to find a title that does justice to all claimants" [136].

Why Nyquist?

Exactly how, when, or why Harry Nyquist had his name attached to the sampling theorem remains obscure. The term *Nyquist Sampling Theorem* (capitalized thus) appeared as early as 1959 in a book from his former employer, Bell Labs,^[19] and appeared again in 1963,^[20] and not capitalized in 1965.^[21] It had been called the *Shannon Sampling Theorem* as early as 1954,^[22] but also just *the sampling theorem* by several other books in the early 1950s.

In 1958, Blackman and Tukey cited Nyquist's 1928 paper as a reference for *the sampling theorem of information theory*,^[23] even though that paper does not treat sampling and reconstruction of continuous signals as others did. Their glossary of terms includes these entries:

Sampling theorem (of information theory)

Nyquist's result that equi-spaced data, with two or more points per cycle of highest frequency, allows reconstruction of band-limited functions. (See *Cardinal theorem*.)

Cardinal theorem (of interpolation theory)

A precise statement of the conditions under which values given at a doubly infinite set of equally spaced points can be interpolated to yield a continuous band-limited function with the aid of the function

$$\frac{\sin(x - x_i)}{x - x_i}.$$

Exactly what "Nyquist's result" they are referring to remains mysterious.

When Shannon stated and proved the sampling theorem in his 1949 paper, according to Meijering^[18] "he referred to the critical sampling interval $T = 1/(2W)$ as the *Nyquist interval* corresponding to the band W , in recognition of Nyquist's discovery of the fundamental importance of this interval in connection with telegraphy." This explains Nyquist's name on the critical interval, but not on the theorem.

Similarly, Nyquist's name was attached to *Nyquist rate* in 1953 by Harold S. Black:^[24]

"If the essential frequency range is limited to B cycles per second, $2B$ was given by Nyquist as the maximum number of code elements per second that could be unambiguously resolved, assuming the peak interference is less half a quantum step. This rate is generally referred to as **signaling at the Nyquist rate** and $1/(2B)$ has been termed a *Nyquist interval*." (bold added for emphasis; italics as in the original)

According to the OED, this may be the origin of the term *Nyquist rate*. In Black's usage, it is not a sampling rate, but a signaling rate.

See also

- Balian–Low theorem, a similar theoretical lower-bound on sampling rates, but which applies to time–frequency transforms.
- The Cheung–Marks theorem specifies conditions where restoration of a signal by the sampling theorem can become ill-posed.
- Hartley's law
- Nyquist ISI criterion
- Reconstruction from zero crossings
- Zero-order hold

Notes

1. The sinc function follows from rows 202 and 102 of the transform tables
2. Shannon 1949, p 448.

References

1. Nemirovsky, Jonathan; Shimron, Efrat (2015). "Utilizing Bochners Theorem for Constrained Evaluation of Missing Fourier Data". arXiv:1506.03300 (<https://arxiv.org/abs/1506.03300>) [physics.med-ph (<https://arxiv.org/archive/physics/physics.med-ph>)].
2. Shannon, Claude E. (January 1949). "Communication in the presence of noise". *Proc. Institute of Radio Engineers*. **37** (1): 10–21. doi:10.1109/jrproc.1949.232969 (<https://doi.org/10.1109%2Fjrproc.1949.232969>). Reprint as classic paper in: *Proc. IEEE*, Vol. 86, No. 2, (Feb 1998) (<http://www.stanford.edu/class/ee104/shannonpaper.pdf>)
3. Marvasti (ed), F. (2000). *Nonuniform Sampling, Theory and Practice*. New York: Kluwer Academic/Plenum Publishers.
4. Landau, H. J. (1967). "Necessary density conditions for sampling and interpolation of certain entire functions". *Acta Math*. **117** (1): 37–52. doi:10.1007/BF02395039 (<https://doi.org/10.1007%2FBF02395039>).
5. see, e.g., Feng, P. (1997). *Universal minimum-rate sampling and spectrum-blind reconstruction for multiband signals*. Ph.D. dissertation, University of Illinois at Urbana-Champaign.
6. Mishali, Moshe; Eldar, Yonina C. (March 2009). "Blind Multiband Signal Reconstruction: Compressed Sensing for Analog Signals". *IEEE Trans. Signal Processing*. **57** (3). CiteSeerX 10.1.1.154.4255 (<http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.154.4255>)
7. Kipnis, Alon; Goldsmith, Andrea J.; Eldar, Yonina C.; Weissman, Tsachy (January 2016). "Distortion rate function of sub-Nyquist sampled Gaussian sources". *IEEE Transactions on Information Theory*. **62**: 401–429. doi:10.1109/tit.2015.2485271 (<https://doi.org/10.1109%2Ftit.2015.2485271>).
8. Kipnis, Alon; Goldsmith, Andrea J.; Eldar, Yonina C. (May 2015). "Sub-Nyquist sampling achieves optimal rate-distortion". *IEEE Information Theory Workshop (ITW)*.
9. Nyquist, Harry (April 1928). "Certain topics in telegraph transmission theory". *Trans. AIEE*. **47**: 617–644. doi:10.1109/t-aiee.1928.5055024 (<https://doi.org/10.1109%2Ft-aiee.1928.5055024>). Reprint as classic paper in: *Proc. IEEE, Vol. 90, No. 2, Feb 2002* (http://www.ieee.org/publications_standards/publications/proceedings/nyquist.pdf)
10. Küpfmüller, Karl (1928). "Über die Dynamik der selbsttätigen Verstärkungsregler". *Elektrische Nachrichtentechnik* (in German). **5** (11): 459–467. (English translation 2005) (<http://ict.open.ac.uk/classics/2.pdf>).
11. Kotelnikov, V.A. (1933). "On the carrying capacity of the ether and wire in telecommunications". *Material for the First All-Union Conference on Questions of Communication, Izd. Red. Upr. Svyazi RSKA* (in Russian). Moscow. (English translation, PDF) (<http://ict.open.ac.uk/classics/1.pdf>)
12. Whittaker, E.T. (1915). "On the Functions Which are Represented by the Expansions of the Interpolation Theory". *Proc. Royal Soc. Edinburgh*. **35**: 181–194. doi:10.1017/s0370164600017806 (<https://doi.org/10.1017%2Fs0370164600017806>). ("Theorie der Kardinalfunktionen").
13. Whittaker, J.M. (1935). *Interpolatory Function Theory*. Cambridge, England: Cambridge Univ. Press..
14. Shannon, Claude E. (July 1948). "A Mathematical Theory of Communication" (<http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=6773024>). *Bell System Technical Journal*. **27** (3): 379–423. doi:10.1002/j.1538-7305.1948.tb01338.x (<https://doi.org/10.1002%2Fj.1538-7305.1948.tb01338.x>)..
15. Shannon, Claude E. (October 1948). "A Mathematical Theory of Communication" (<http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=6773067>). *Bell System Technical Journal*. **27** (4): 623–666. doi:10.1002/j.1538-7305.1948.tb00917.x (<https://doi.org/10.1002%2Fj.1538-7305.1948.tb00917.x>)..
16. Jerri, Abdul (November 1977). "The Shannon Sampling Theorem—Its Various Extensions and Applications: A Tutorial Review" (<http://web.archive.org/web/20080605171238/http://ieeexplore.ieee.org/search/wrapper.jsp?arnumber=1455040>). *Proceedings of the IEEE*. **65** (11): 1565–1596. doi:10.1109/proc.1977.10771 (<https://doi.org/10.1109%2Fproc.1977.10771>). See also Jerri, Abdul (April 1979). "Correction to "The Shannon sampling theorem—Its various extensions and applications: A tutorial review" " (<http://web.archive.org/web/20090120203809/http://ieeexplore.ieee.org/search/wrapper.jsp?arnumber=1455576>). *Proceedings of the IEEE*. **67** (4): 695. doi:10.1109/proc.1979.11307 (<https://doi.org/10.1109%2Fproc.1979.11307>).


17. Lücke, Hans Dieter (April 1999). "The Origins of the Sampling Theorem". *IEEE Communications Magazine*. **37** (4): 106–108. doi:10.1109/35.755459 (<https://doi.org/10.1109%2F35.755459>).
18. Meijering, Erik (March 2002). "A Chronology of Interpolation From Ancient Astronomy to Modern Signal and Image Processing". *Proc. IEEE*. **90** (3): 319–342. doi:10.1109/5.993400 (<https://doi.org/10.1109%2F5.993400>).
19. Members of the Technical Staff of Bell Telephone Laboratories (1959). *Transmission Systems for Communications*. AT&T. pp. 26–4 (Vol.2).
20. Guillemin, Ernst Adolph (1963). *Theory of Linear Physical Systems* (<https://books.google.com/books?id=jtI-AAAAIAAJ>). Wiley.
21. Roberts, Richard A.; Barton, Ben F. (1965). *Theory of Signal Detectability: Composite Deferred Decision Theory*.
22. Gray, Truman S. (1954). *Applied Electronics: A First Course in Electronics, Electron Tubes, and Associated Circuits*.
23. Blackman, R. B.; Tukey, J. W. (1958). *The Measurement of Power Spectra : From the Point of View of Communications Engineering* (<http://alcatel-lucent.com/bstj/vol37-1958/articles/bstj37-1-185.pdf>) (PDF). New York: Dover.
24. Black, Harold S. (1953). *Modulation Theory*.

Further reading

- Higgins, J.R.: *Five short stories about the cardinal series*, Bulletin of the AMS 12(1985)
- Küpfmüller, Karl, "Utjämningsförlopp inom Telegraf- och Telefontekniken", ("Transients in telegraph and telephone engineering"), Teknisk Tidskrift, no. 9 pp. 153–160 and 10 pp. 178–182, 1931. [1] (<http://runeberg.org/tektid/1931e/0157.html>) [2] (<http://runeberg.org/tektid/1931e/0182.html>)
- Marks, R.J.(II): *Introduction to Shannon Sampling and Interpolation Theory* (http://marksmannet.com/RobertMarks/REPRINTS/1999_IntroductionToShannonSamplingAndInterpolationTheory.pdf), Springer-Verlag, 1991.
- Marks, R.J.(II), Editor: *Advanced Topics in Shannon Sampling and Interpolation Theory* (http://marksmannet.com/RobertMarks/REPRINTS/1993_AdvancedTopicsOnShannon.pdf), Springer-Verlag, 1993.
- Marks, R.J.(II), *Handbook of Fourier Analysis and Its Applications*, Oxford University Press, (2009), Chapters 5-8. Google books (<https://books.google.com/books?id=Sp704bocjPAC>)
- Press, WH; Teukolsky, SA; Vetterling, WT; Flannery, BP (2007), "Section 13.11. Numerical Use of the Sampling Theorem" (<http://apps.nrbook.com/empanel/index.html#pg=717>), *Numerical Recipes: The Art of Scientific Computing* (3rd ed.), New York: Cambridge University Press, ISBN 978-0-521-88068-8
- Unser, Michael: *Sampling-50 Years after Shannon* (<http://bigwww.epfl.ch/publications/unser0001.htm>), Proc. IEEE, vol. 88, no. 4, pp. 569–587, April 2000

External links

- Learning by Simulations (http://www.vias.org/simulations/simusoft_nykvist.html) Interactive simulation of the effects of inadequate sampling
- Interactive presentation of the sampling and reconstruction in a web-demo (http://webdemo.inue.uni-stuttgart.de/webdemos/02_lectures/uebertragungstechnik_1/sampling_theorem/) Institute of Telecommunications, University of Stuttgart
- Undersampling and an application of it (<http://spazioscuola.altervista.org/UndersamplingAR/UndersamplingARnv.htm>)
- Sampling Theory For Digital Audio (https://web.archive.org/web/20060614125302/http://www.lavryengineering.com/documents/Sampling_Theory.pdf)
- Journal devoted to Sampling Theory (<http://www.stsip.org/>)

- Sampling Theorem with Constant Amplitude Variable Width Pulse (http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=5699377&url=http%3A%2F%2Fieeexplore.ieee.org%2Fexpls%2Fabs_all.jsp%3Farnumber%3D5699377)
- Hans Dieter Lüke (April 1999). "The Origins of the Sampling Theorem". *IEEE Communications Magazine*. CiteSeerX 10.1.1.163.2887 (<https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.163.2887>) .

Retrieved from "https://en.wikipedia.org/w/index.php?title=Nyquist-Shannon_sampling_theorem&oldid=791056734"

Categories: Digital signal processing | Information theory | Theorems in Fourier analysis
 | Mathematical theorems in theoretical computer science | Claude Shannon
 | Telecommunication theory

-
- This page was last edited on 17 July 2017, at 21:19.
 - Text is available under the Creative Commons Attribution-ShareAlike License; additional terms may apply. By using this site, you agree to the Terms of Use and Privacy Policy. Wikipedia® is a registered trademark of the Wikimedia Foundation, Inc., a non-profit organization.