Chapter 15

Sampling principles

P. Cochrane

1 PROLOGUE

Applying a sampling technique to the solution of any problem is often an admission of a fundamental inability to encompass all the data involved. In effect sampling is specifically used to reduce information to its essentials - or at least to a more manageable level. At microwave frequencies it is often the only practicable technique available, and in this forum time domain sampling may be classified into three broad functional categories:-

i. Switching between two or more reference/measure paths

ii. Part of an Analogue to Digital conversion process

iii. Overcoming fundamental hardware speed limitations precluding real-time operation.

In turn each of these categories involve processes that may be perceived to be Sub-Nyquist, Nyquist or Super Sampled. It is therefore essential that we examine the action of sampling in some depth before we go on to consider specific microwave measurement applications. This chapter is thus solely concerned with explaining the action and limitations of sampling in the time domain - but with due reference to the frequency domain implications. Some reference is also made to the practical difficulties encountered and the engineering solutions available. In the chapter that follows we concentrate on specific examples of instruments and measurement techniques employing sampling and explain their actions, applications and limitations. An initial consideration is also given to the design and realisation of practical time samplers.

1.1 Mathematical Symbols, Notation and Definitions

i = an integer

n = an integer

T = a period of time

τ = a period of time

f_c = carrier frequency

f_m = modulation frequency

f_s = sampling frequency

B = a finite frequency band

β = modulation index

J_n = Bessel function of the first kind

$g(t)$ = generalised function of time

$G(f)$ = generalised function of frequency

$<=>$ signifies a Fourier transformation

$*$ signifies a convolution integral

$\sum_n$ = a summation over all 'n'

$\sum_n'$ = a summation over all 'n' except n=o

$s(t) = \text{comb}_T[\ ] = \sum_i \delta(t-iT)$ = the sampling function (at T intervals)

$\text{Rep}_{\frac{1}{T}}[\ ]$ = the replication function (at $\frac{1}{T}$ intervals)

$\text{rect}(t) = u(t+\frac{1}{2})-u(t-\frac{1}{2})$ = the rectangular function

$\text{sinc}(t) = \frac{\sin(\pi ft)}{\pi ft}$

$\delta(t) \triangleq \int_{-\infty}^{+\infty} \delta(t)dt = 1;\ \delta(t) = 0,\ t \neq 0$ = the Dirac delta function

2 INTRODUCTION

The derivation of the sampling theorem is often attributed to Nyquist [1], Shannon [2], and less commonly to Gabor [3], but in fact it was known to Cauchy [4] much earlier. In many respects it was rediscovered and refined by Whittaker [5,6] before it emerged in its more popular guise via communication theory [7]. Briefly this theory states:-

> If a signal is band-limited and if the time interval is divided into equal parts forming subintervals such that each subdivision comprises an interval T seconds long where T is less than half the period of the highest significant frequency (B) component of the signal; and if one instantaneous sample is taken from each subinterval in any manner; then a knowledge of the instantaneous magnitude of each sample plus a knowledge of the instants within each subinterval at which the sample is taken contains all the information of the original signal.

There are at least 5 generalised sampling theorems [8] that spring from this definition - and many more [9,10] related to the relaxation of the bandlimiting constraint. Fortunately we are only concerned with two relatively straightforward interpretations pertaining to systematic and random sampling of bandlimited signals.

3 UNIFORM SAMPLING

3.1 The Nyquist Case.

Let us consider a bandlimited signal g(t) with a frequency spectrum G(f), sampled by evenly spaced Dirac elements [11-15] as follows:-

$$g(t)\,.\,s(t) = \mathrm{comb}_T\,[g(t)] \equiv \sum_i \delta(t-iT)\,.\,g(t) \qquad \ldots\ldots (1)$$

$$\Updownarrow \qquad\qquad \Updownarrow$$

$$\frac{1}{T}\,\mathrm{Rep}_{\frac{1}{T}}\,[G(f)] \equiv \frac{1}{T}\sum_n G(f-\tfrac{n}{T}) \qquad \ldots\ldots (2)$$

Alternatively, an amplitude modulation view of this process may be derived:-

$$\frac{1}{T}\sum_n G(f-\tfrac{n}{T}) = \frac{1}{T}\sum_n \delta(f-\tfrac{n}{T}) * G(f) \qquad \ldots\ldots (3)$$

$$= \frac{G(f)}{T} + \frac{1}{T}{\sum_n}' \delta(f-\tfrac{n}{T}) * G(f) \qquad \ldots\ldots (4)$$

Therefore
$$g(t)\,.\,s(t) = \frac{g(t)}{T}\left\{1 + 2\sum_{i=1}^{\infty} \cos 2\pi i\frac{t}{T}\right\} \qquad \ldots\ldots (5)$$

Pictorially the sampling of a bandlimited signal may thus be viewed as an amplitude modulated harmonic series as indicated in Fig 1.

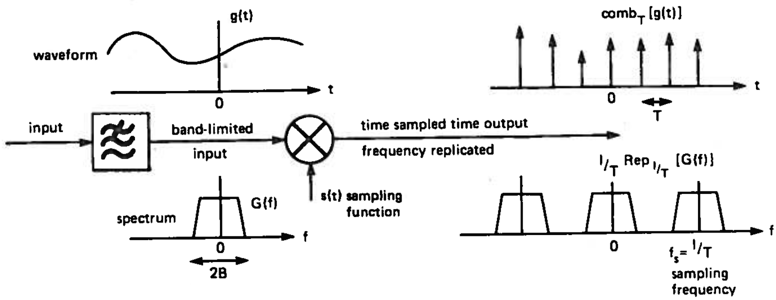

Fig 1 Sampling a band-limited signal

Clearly the limiting condition for separability - or recovery of the original information undistorted, is when $f_s \geqslant 2B$ - this is commonly referred to as the Nyquist condition [1] and 2B called the Nyquist frequency.

3.2 Sub-Sampling.

If the sampling frequency is not at least twice that of the highest component present, or if the signal is not strictly band-limited, aliasing distortion [16] occurs as the frequency bands overlap.

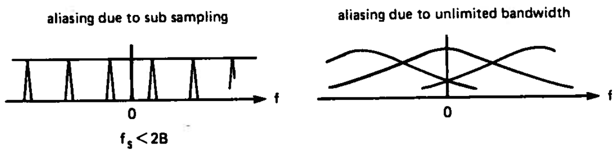

Fig 2 Alias distortion in the frequency domain

This phenomenon of non-ideal-separability may also be conveniently viewed in the time domain as depicted in Fig 3.

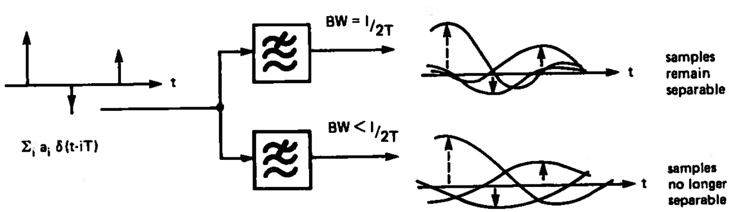

Fig 3 Alias distortion in the time domain

In the case of continuous signals that are not strictly band-limited this effect has to be minimised to prevent serious distortion. The effects of aliasing are often considered as noise-like processes and a power criterion is commonly used to restrict its effect. Typically this might be < 1% of the total recovered sample power depending upon the application [17]. For periodic signals this problem is generally less serious because the frequency spectrum is discontinuous and the spectra can be interleaved by the sampling process. The signal thus remains perfectly recoverable provided the sampling frequency is adjusted to achieve spectral interleaving.

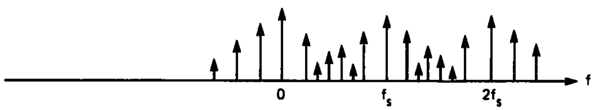

Fig 4 Interleaving of sample spectra

This form of sub-sampling is predominant in instrumentation where sinusoidal or repetitive waves of limited harmonic content are to be processed. For a sinusoidal wave, in particular, the sampling action may be thought of as a "down conversion" modulation:-

$$\mathrm{comb}_T\,[\cos 2\pi f_c t] \Longleftrightarrow \frac{1}{2T}\,\mathrm{Rep}_{\frac{1}{T}}\,[\delta(f-f_c) + \delta(f+f_c)] \qquad \ldots\ldots (6)$$

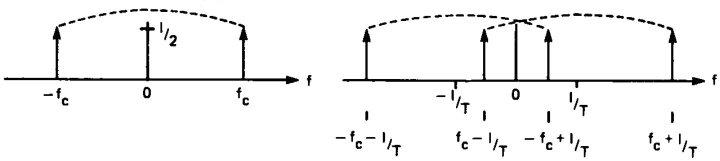

Fig 5 Down conversion via sampling

3.3 Super Sampling.

Strictly speaking any sampling process that uses $f_s>2B$ falls in the super sampling category. In fact the title is more commonly reserved for cases where $f_s>>2B$, which implies a high degree of separability making possible simple filtering. When defining super sampling we have to take care in terms of our definition of the original signal. Consider the two cases of a single carrier and a carrier with bandlimited modulation as shown in Fig 6.

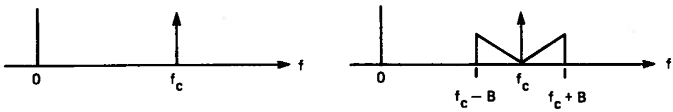

Fig 6 Example spectra

On the one hand to recover the carrier we might expect to sample at $f_s > 2f_c$. On the other, to retrieve the modulation information we have to sample at $f_s > 2B$ - not $f_s > 2(f_c + B)$!! How is this so? Consider the action of the sampler:-

signal $g(t) = [1 + m(t)] \cos\omega_c t$ ..... (7)

sampled signal $h(t) = \text{comb}_T [\cos\omega_c t] + \text{comb}_T [m(t) \cos\omega_c t]$ ..... (8)

$$\Updownarrow$$

$$H(f) = \frac{1}{T} \text{Rep}_{\frac{1}{T}} \left\{ [\delta(f) + M(f)] * \frac{1}{2} [\delta(f-f_c) + \delta(f+f_c)] \right\}$$

..... (9)

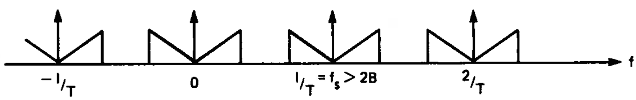

Fig 7 Sampler output spectrum for a carrier plus modulated carrier

Qualitatively we can argue that a carrier alone contains very little information (only amplitude and frequency) and occupies zero bandwidth, therefore requiring a sampling rate << f_c. Similarly the information bearing energy of the sidebands lies in a bandwidth B, and a sampling rate of $f_s \geqslant 2B$ is sufficient. In short, the process of down conversion can be achieved by a sampler of rate $\geqslant 2B$.

It is clearly necessary to take some care to define sampling systems in terms of the information to be recovered; for a repetitive wave containing fixed amplitude and frequency information we may in principle sample as slowly as we wish.

4 NON-UNIFORM SAMPLING

4.1 Statement of Constraints.

The sampling theorem tells us that we may recover all of the original information provided that samples are taken in T spaced slots (satisfying the greater than twice frequency criterion) and that both the amplitude and time information is recorded [8]. This condition is satisfied for random and deterministic distributions of the sample times provided they do not stray beyond the boundary given by ± T/2 on the uniform case [18,19].

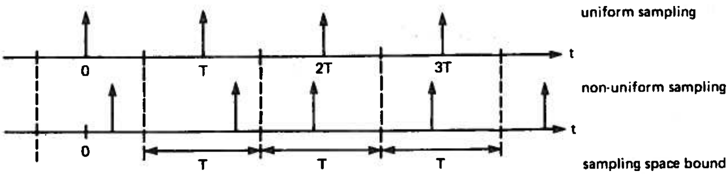

Fig 8 Non-uniform sampling

4.2 Practical Occurrence.

The randomisation of the sampling process can arise via three principal mechanisms:-

i. Uncertainty in the signal arrival time - such as pulse from a non-stationary source.

ii. Jitter in the sampler action.

iii. Deliberate randomisation, or deterministic perturbation, introduced to avoid harmonic or sub-harmonic beats between the signal and sampling frequencies.

Providing the sampling time shifts are constrained and both amplitude and position information is retained each of these categories may be catered for in practice [21]. However, in the field of nuclear physics the above conditions are further compounded by amplitude and shape variations from sample to sample. Fortunately this lies outside our sphere of interest!

4.3 Deterministic Perturbation Analysis.

As any well behaved deterministic distribution of the sample instants may be described by a suitable Fourier series, we consider the special case of a sinusoidal variation, about the uniform points. Extension to an arbitrary variation is then a straightforward matter of including similarly disposed components [22,23]. For manipulative convenience we therefore apply a sinusoidal modulation to the sampling instants via a cosine series description:-

$$s(t) = \frac{1}{T}\left\{1 + 2\sum_i \cos 2\pi f_s it\right\} \qquad \ldots\ldots (10)$$

With the introduction of a sinusoidal sampling deviation this takes on the deterministic form:-

$$s_d(t) = \frac{1}{T}\left\{1 + 2\sum_i \cos(\omega_s it + \beta \sin\omega_m t)\right\} \qquad \ldots\ldots (11)$$

$$= \frac{1}{T}\left\{1 + 2\sum_i \sum_n J_n(\beta)\cos(i\omega_s + n\omega_m)t\right\} \qquad \ldots\ldots (12)$$

Where $J_n(\beta)$ are Bessel functions of the first kind

β is the modulation index

ω_m is the modulation frequency

Sampling a bandlimited signal with this wave gives:-

$$s_d(t).g(t) <=> \frac{1}{T}\left\{\delta(f) + \sum_{i,n} J_n(\beta)[\,\delta(f-if_s-nf_m)+\delta(f+if_s+nf_m)]\right\} * G(f)$$

$$\ldots\ldots (13)$$

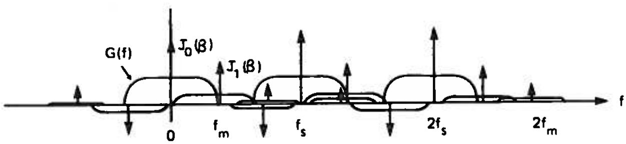

Fig 9 Sampling spectra with deterministic perturbation

Hence provided f_s is made sufficiently large, the deviation βf_m is constrained, and we have a knowledge of f_s (ie when the samples were taken) we can recover g(t) intact. When $f_s >> \hat{f}_G$ - the highest frequency in the bandlimited signal - we need only consider the i = 0 component in isolation:-

$$s_d(f) * G(f)\big|_{i=0} = \frac{G(f)}{T} + \sum_n \frac{J_n(\beta)}{T}\,[G(f) - nf_m) + G(f + nf_m)]$$

$$\ldots\ldots (14)$$

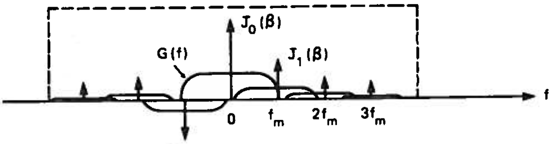

Fig 10 The fundamental sampling spectrum with a deterministic perturbation

A further separability condition is now apparent. When $2\hat{f}_G < f_m$; G(f) may be recovered by filtering out the baseband component about f = 0.

When the above conditions for separability do not apply our only recourse is to deconvolve [24] or log both amplitude and time information of each sample [25,27] - simple recovery by filtering is not generally possible.

4.4 Random Perturbation Analysis.

Unfortunately there is no concise closed form analysis available for this case [28] and we thus adopt a series solution to illustrate the process and its similarities with deterministic perturbation. Stipulating that the random modulation introduced by $\phi(t)$ obeys the condition $|\phi(t)| \ll 1$, we can assume the following approximations:-

$$s_R(t) = \frac{1}{T}\left\{1 + 2\sum_i [\cos\omega_s it - \phi(t)\sin\omega_s it]\right\} \qquad \text{..... (15)}$$

$$S_R(f) = \frac{1}{T}\ \delta(f) + \sum_i [\delta(f-if_s) + \delta(f+if_s)] + j\sum [\Phi(f-if_s) - \Phi(f+if_s)] \qquad \text{..... (16)}$$

$$S_R(f)*G(f) = \frac{G(f)}{T} + \frac{1}{T}\sum_i [G(f-if_s)+G(f+if_s)] + \frac{j}{T}\sum [\Phi(f-if_s)-\Phi(f+if_s)]*G(f) \qquad \text{..... (17)}$$

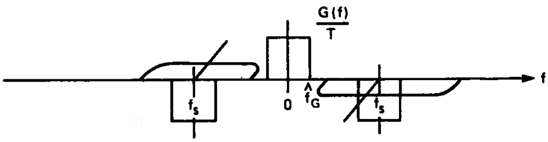

Fig 11 Recoverable spectra with constrained random perturbation

Again separability is possible provided; $f_s >> \hat{f}_G$ and $f_s >> (\hat{f}_G + \hat{\Phi})$, otherwise a record of all sample amplitude and times has to be made as per the deterministic case.

5 PRACTICAL LIMITATIONS

5.1 Aperture Distortion.

The Dirac δ function is physically unrealizable and in practice we have to content ourselves with more modest functions [29,30]. These are not only limited in their amplitude and width, but also in their shape. Although a rectangular pulse is unrealiseable - as is a rectangular filter in the frequency domain, we use this ideal as a convenient approximation to demonstrate the practical limitations imposed by finite amplitude and width sampling pulses.

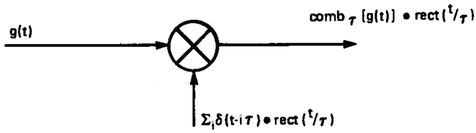

Fig 12 Aperture distortion model

With this imperfect sampling pulse the output is both amplitude scaled and time domain smeared by the convolution process, which leads to a reduction of effective bandwidth in the frequency domain:-

$$\underbrace{\mathrm{comb}_T\,[g(t)]}_{\text{perfect sampling}} * \underbrace{\mathrm{rect}\left(\frac{t}{\tau}\right)}_{\text{smearing function}} <=> \underbrace{\frac{1}{T}\,\mathrm{Rep}_{\frac{1}{T}}\,[G(f)]}_{\text{perfect spectrum}} \cdot \underbrace{\tau\,\mathrm{sinc}\,(f\tau)}_{\text{bandwidth reduction}} \quad ..... (18)$$

The sampler output thus suffers a frequency domain droop dependent upon the particular shape of the sampling pulse. For all practical pulses of interest the resulting distortion is bounded - best to worst - by sinc and sinc^2, ie the pulse shape lies between the rectangular and triangular. The resulting amplitude distortion introduced by these functions, as well as that for a Gaussian pulse are given in Fig 13.

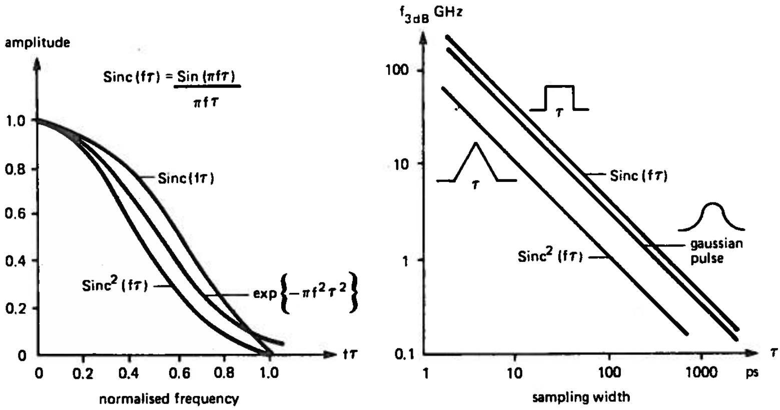

Fig 13 Aperture distortion with pulse width and shape

For some applications this limitation can be partly off-set by introducing an amount of high frequency compensation prior to the sampler as indicated in Fig 14. However, this is not a popular technique at microwave frequencies as it often introduces other complications related to phase distortion which is also compounded by the variability of practical samplers - generally speaking engineering solutions tend towards producing the best sampler possible.

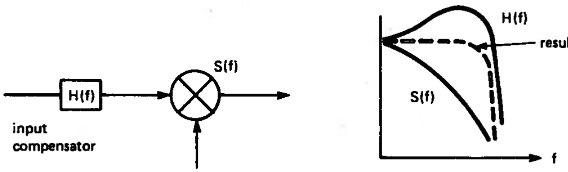

Fig 14 Sampler frequency compensation

For many applications the concept of a bandwidth reduction penalty is sufficient - in others it is somewhat imprecise. Consider the examples given in Fig 15; clearly for the smooth sinusoid the lack of bandwidth is not critically important - it only results in an amplitude scaling. For step or pulse type signals however the penalty can be more severe. In these cases an engineering rule of thumb says; the sampling pulse width should not exceed 1/10 the transient period.

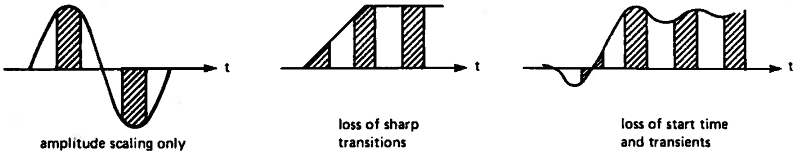

Fig 15 The importance of sampling width

5.2 Synchronised Sampling.

When dealing with periodic signals the sampling process works provided the sampling frequency and signal frequency are not directly (or closely) related. To illustrate this feature let us consider the case when:-

$$g(t) = \cos\left[\frac{2\pi t}{nT}\right] \qquad \ldots\ldots (19)$$

where n is an integer and $f_s = \frac{1}{T}$

$$\text{now } \operatorname{comb}_T\left[\cos\frac{2\pi t}{nT}\right] <=> \frac{1}{2T}\operatorname{Rep}_{\frac{1}{T}}\left[\delta\left(f - \frac{1}{nT}\right) + \delta\left(f + \frac{1}{nT}\right)\right] \qquad \ldots\ldots (20)$$

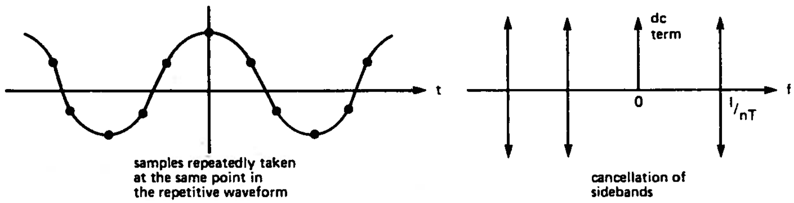

Fig 16 Synchronised sampling of a sinusoid

This clearly results in a repetitive sampling of the same value in the wave which manifests itself as sideband cancellation in the frequency domain. Recovery of the original wave is thus impossible and hence the sampling theorem requirement $f_s > 2B$. For the case of closely related sampling and period time - ie 'n' is 'not quite' an integer - a slow roll or progression becomes evident.

$$\operatorname{comb}_T\left[\cos\frac{2\pi t}{(n+\Delta)T}\right] <=> \frac{1}{2T}\operatorname{Rep}_{\frac{1}{T}}\left[\delta\left(f - \frac{1}{(n+\Delta)T}\right) + \delta\left(f + \frac{1}{(n+\Delta)T}\right)\right] \qquad \ldots\ldots (21)$$

The resulting sampling process thus suffers from a "beat frequency", which in terms of any measurement or observation is undesirable and should be avoided. In contrast, when the sampling frequency and repetition rate of a periodic wave are related in a sub-harmonic manner ($f_s << f_g$), then synchronised sub-sampling is achieved giving an expanded time display of the original wave as shown in Fig 17. At present this is the most commonly used sampling CRO display technique for microwave frequencies, and as per the previous case, precise synchronisation must be maintained or a beat frequency (display roll) is produced.

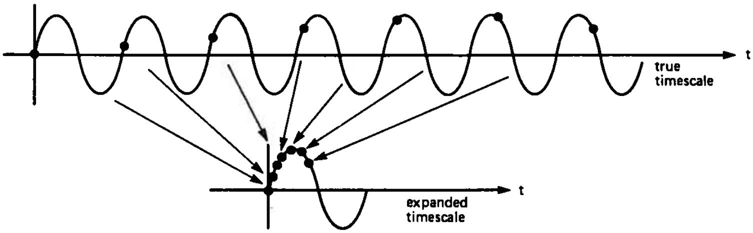

Fig 17 Synchronised sub-sampling of a sinusoid

5.3 The Hold Process.

Because we inherently use narrow sample pulses it is often necessary to stretch the output from the sample gate before it can be further processed [26,29]. Attempting an A/D conversion or trying to display these direct would clearly be impossible. It is thus necessary to hold the sample values so that an A/D converter can converge [31-34] or a CRO tube can be sufficiently illuminated.

The physical realisation of the hold function is dependent upon the particular application, but in general it takes the form of an integrator, which leads naturally to the integrated product description for sampling depicted in Fig 18. In many microwave applications it is necessary to distribute this process using dispersive transmission line, amplification and the sophisticated integration to counteract temperature, voltage and other drift parameters.

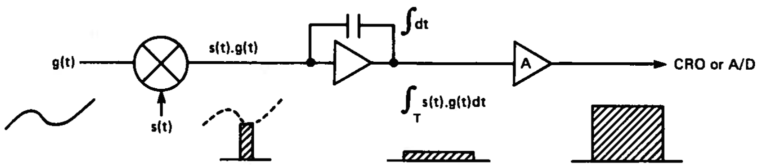

Fig 18 Sample and hold model

5.4 Sampling Jitter and Noise.

All practical sampling schemes suffer from fundamental circuit limitations and signal uncertainties that give rise to both amplitude noise and phase jitter. Given that every care is exercised to minimise these at the design stage, further reduction is possible by signal averaging along both the amplitude and time axes, as well as more sophisticated image processing [31,35]. Broadly speaking averaging techniques achieve performance improvements by the voltage addition of the wanted signal and power addition of the noise/jitter as depicted in Fig 19.

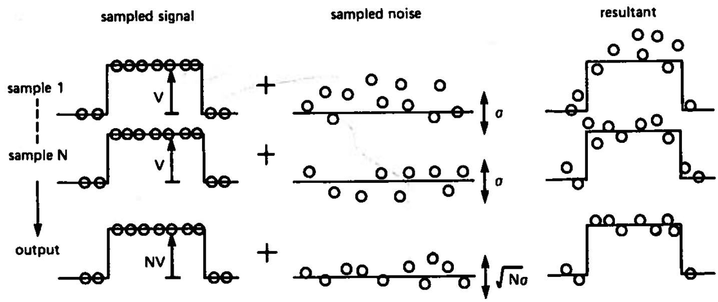

Fig 19 Signal averaging example

$$\text{Original S/N} = \left(\frac{v}{\sigma}\right)^2 \quad \text{..... (22)}$$

$$\text{New S/N} = \frac{(nv)^2}{n\sigma^2} \quad \text{..... (23)}$$

The improvement attained is thus α 10 log n dB.

For small deviations sampling jitter may be considered to be a noise-like process and, by virtue of phase to amplitude conversion, the above averaging description is applicable. However, amplitude and time axis averaging may be applied independently by operating on the full sample array [31,36-38].

6 A FINAL NOTE ON THE PRINCIPLES

Although we have concentrated on time domain sampling alone, it should be recognised that many of the developments described have a dual role in the frequency domain. This duality is principally introduced by the nature of the Fourier Transform Pair. Sampling in time or frequency produce very similar phenomenon in the other transform domain. Further analogies may also be drawn with travelling wave devices which effectively use spatial sampling to achieve wide bandwidth operation. A good introductory text explaining some of these aspects, has been written by Bracewell [13].

7 REFERENCES

1. Nyquist, H.: 1928, Trans AIEE, 47, 617-644.

2. Shannon, C.E.: 1949, Proc IRE, 37, 10-21.

3. Gabor, D.: 1946, Journal IEE, 93, pt 3, 429-457.

4. Cauchy, A.L.: 1841, Compt Rend Paris, 12, 283-298.

5. Whittaker, E.T.: 1915, Proc Royal Society Edinburgh, 35, 181-194.

6. Whittaker, E.T.: 1935, Cambridge Tracts in Mathematical Physics, 33.

7. Luke, H.D.: 1978, Nachrichtentechnik, 31/4, 271-273.

8. Yen, J.L.: 1956, IRE Trans on Cct Theory, 251-257.

9. Kishi, G. & Sakanina, K.: 1981, Electron and Comms in Japan, 64-A/3, pp 34-42.

10. Jerri, A.J.: 1977, Proc IEEE, 65/11, 1565-1596.

11. Muth, J.M.: 1977, Transform Methods with Applications to Engineering and Operations Research. Prentice Hall.

12. Brigham, E.O.: 1974, The Fast Fourier Transform. Prentice Hall.

13. Bracewell, R.: 1965, The Fourier Transform and its Applications. McGraw-Hill.

14. Van der Pol, B. & Bremmer, H.: 1955, Operational Calculus Based on the Two-Sided Laplace Integral. Cambridge University Press.

15. Lathi, B.P.: 1968, An Introduction to Random Signals and Communication Theory. Intertext Books London.

16. Wozencraft, J.M. & Jacobs, I.M.: 1965, Principles of Communication Engineering. Wiley, New York.

17. Schwartz, M.: 1970, Information Transmission Modulation and Noise. McGraw-Hill, New York.

18. Papoulis: 1965, Probability Random Variables and Stochastic Processes. McGraw-Hill, New York.

19. Middleton, D.: 1960, Introduction to Statistical Communication Theory. McGraw-Hill, New York.

20. Nahman, N.S.: 1983, IEE Trans IM-32/1, 117-124.

21. Nahman, N.S.: 1978, Proc IEEE, 66/4, 441-454.

22. Sarhadi, M. & Aitchison, C.S.: 1980, IEE Elec Lett, 16/10, 350-352.

23. Sarhadi, M. & Aitchison, C.S.: 1979, IEE Proc EMC, Brighton, England, 345-349.

24. Nahman, N.S. & Guillaume, M.E.: 1981, Deconvolution of Time Domain Waveforms in the Presence of Noise. US NBS, Technical Note 1047.

25. Andrews, J.R.: 1973, Trans IEEE, IM-22/4, 375-381.

26. Frye, G.J. & Nahman, N.S.: 1964, IEE Trans, Vol IM-13/1, 8-13.

27. Sugarman, R.: 1957, Rev Sci Inst, 28/11, 933-938.

28. Bendat, J.S. & Piersol, A.G.: 1971, Random Data: Analysis and Measurement Procedures. Wiley, New York.

29. Grove, W.M.: 1966, IEEE Trans, MTT-14/12, 629-635.

30. Tielert, R.: 1976, IEE Elec Lett, 12/3, 84-85.

31. Jones, N.B.: 1982, Digital Signal Processing. IEE Control Engineering Series 22. Peter Perigrinus.

32. Isaacson, R. et al: 1974, Electron, 19-33.

33. Bucciarelli, T. and Picardi, G.: 1975, Alta Frequenza, XLIV/8, 454-461.

34. Kinneman, G. & Wilde, R.: 1978, Nachrichtentechnik Elecktronik, 28/8, 337-339.

35. Reed, R.C.: 1977, IEE Euromeas Conf Pub 152, 101-103.

36. Haralick, R.M., Shanmucam, K. & Dinstein, I.: 1973, IEEE Trans, SME-3/6, 610-621.

37. Stuller, J.A. & Kurz, R.: 1976, IEEE Trans Comm, 1148-1195.

38. Nahman, N.S.: Proc IEEE, 55/6, 855-864.