Tetsuo Sasao and André B. Fletcher
Introduction to VLBI Systems
Chapter 4
Lecture Notes for KVN Students
Partly based on Ajou University Lecture Notes
(to be further edited)
Version 1. (Unfinished.)
Issued on February 19, 2006.

Very Long Baseline Interferometry

Contents

1 Technologies Which Made VLBI Possible

Figure 1: Schematic view of three major technologies and digital data processing which realized VLBI.

There were three major technologies which enabled realization of radio interferometers with very long baselines, exceeding thousands of kilometers. They were:

1. high–stability frequency standard,
2. high–accuracy time synchronization, and
3. high–speed high–density recording, or super–wideband data transmission in nowadays.

In addition, all signal processings essential to VLBI, such as time marking, recording, delay tracking, fringe stopping, and correlation processing are done digitally in modern VLBI systems. In this sense, rapid progress in digital technology in the last decades has formed a fundament of VLBI, as illustrated in Figure 1.

Therefore, we will first examine theoretical bases of digital signal processing, to an extent which is necessary to understand principles and roles of digital circuitries used in VLBI. Then, we will see basic elements of the major technologies mentioned above.

1.1 Basics of Digital Data Processing

1.1.1 Analog Processing Versus Digital Processing

For VLBI, digital processing is much more suited than analog prcessing, as evident from following comparison.

Analog delay circuit	Digital delay circuit
A coaxial cable can be used as a delay cable (Figure 2), if delay is smaller than ~ 1 μsec (~ 300 m). For larger delay, other means such as sonic wave must be used. Almost impossible to use for an intercontinental baseline. Unstable delay value against environmental change. Larger error with larger delay. Change of frequency characteristis with mechanical connection and disconnection of cables.	A ring buffer composed of a large RAM and shift registers can be used for digital delay tracking (Figure 2). Quite stable and highly repeatable in variable environmental conditions. Does not need any special tuning. Accuracy of operation is determined almost solely by accuracy of clock time pulses. Easy to cover large intercontinental delays. Delay is tracked only discretely with some loss of signal power.
Analog correlation processing	**Digital correlation processing**
Multiplication and averaging with analog devices. Operational range is limited by device characteristics. Affected by the environment.	Multiplication and averaging with logic devices and counters. Stable operation and high repeatability. Wide dynamic range.
Analog spectrometer	**Digital spectrometer**
Narrow–band analog BPF's with square–law detectors ("filterbank spectrometer"). Difficult to adjust gains and frequency characteristics of BPF's (frequency channels). Affected by the environment. Frequency resolution is fixed by passbands of BPF's.	Composed of logic devices, shift registers, and counters. Correlation $\Rightarrow$ Fourier transformation (XF–type) and Fourier transformation $\Rightarrow$ correlation (FX–type). Stable operation. Identical characteristics of frequency channels. Frequency resolution could be variable.
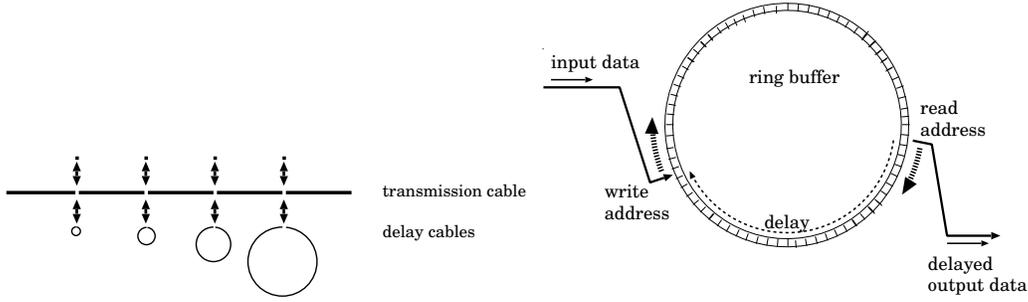
Figure 2: Examples of analog delay circuit using delay cables (left) and digital delay circuit using a ring buffer (right).

1.1.2 Sampling and Clipping

Two important achievements in the theory of digital data processing were vital for VLBI. They are the sampling theorem by Shannon (1949), and the clipping theorem by van Vleck and Middelton (1966, original work was done by van Vleck during World War II).

VLBI data are sampled, and clipped (or digitized) with 1–bit or 2–bit quantization (Figure 3).

In the followings, we will see how the information of the analog data is essentially restored from the sampled and clipped data. We will also consider some loss of information accompanied with the digital data processing.

1.1.3 Discrete–Time Random Process

Discrete sequence of variables $x[1]$, $x[2]$, $x[3]$, $\cdots$, $x[i]$, $\cdots$ is called the "random sequence", or the "discrete–time random process", if $x[i]$ at any i is a random variable, i.e., may vary from trial to trial (Figure 4). This is a "discrete version" of the random process continuously varying in time (henceforth, "continuous–time random process"), which we saw in Chapter 3.

We introduce following statistical concepts for the discrete–time random process.

- Expectation $\eta[i]$ of a discrete–time random process $x[i]$ ($i = 1, 2, 3, \cdots$, n, $\cdots$) is defined by an equation:

$$\eta[i] = \langle x[i] \rangle, \tag{1}$$

where $\langle\ \rangle$ stands for an ensemble average defined by a joint probability distribution of random variables $x[i]$ ($i = 1, 2, 3, \cdots$, n, $\cdots$).

6

Analog signal

Sampling

Clipping (or quantization)

Bit stream

001111111000000000111

To digital circuit
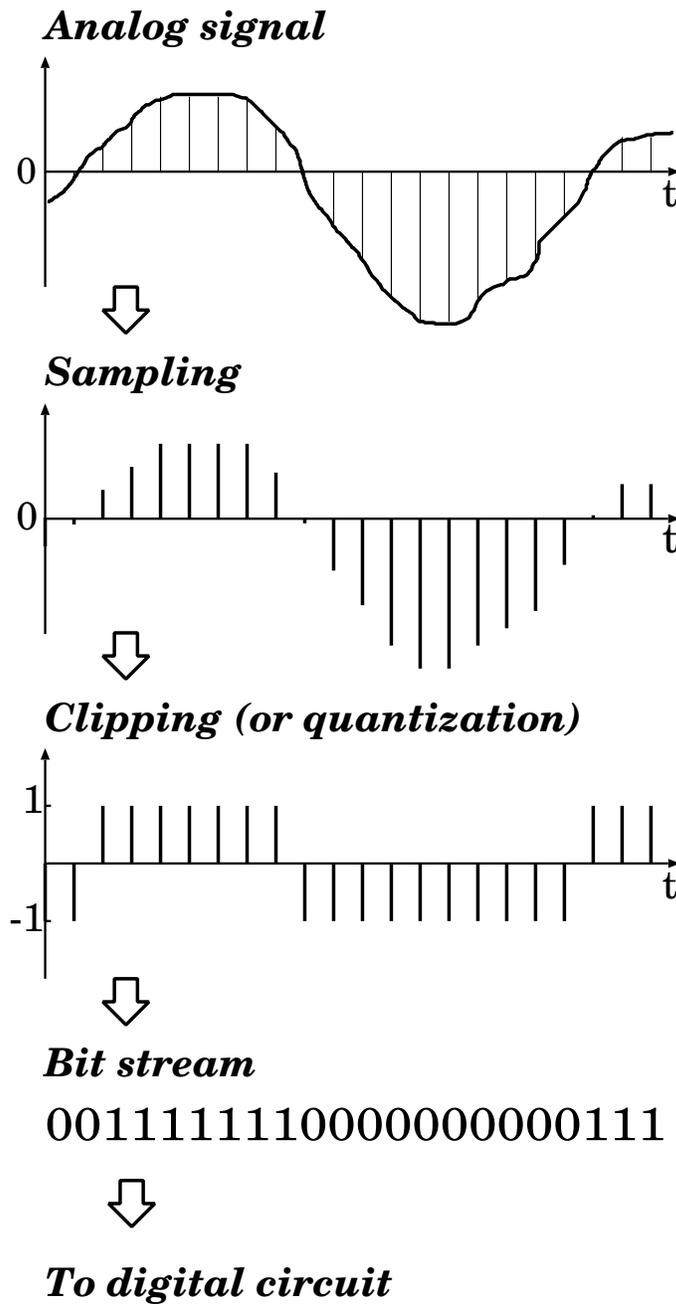
Figure 3: Analog–to–digital (A/D) conversion through sampling, clipping, and bit representation. This figure shows a case of 1–bit quantization.

7
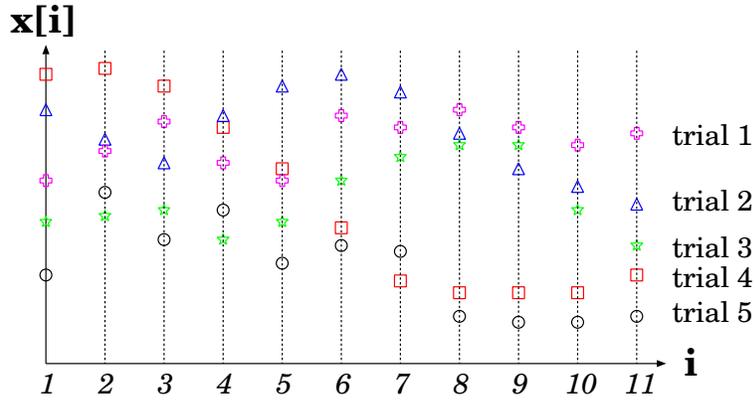
Figure 4: A discrete–time random process is a sequence $x[1]$, $x[2]$, $x[3]$, $\cdots$, $x[i]$, $\cdots$, whose value at any i is a random variable.

- Autocorrelation $R[m, n]$ of the discrete–time random process $x[i]$ is defined by an equation:

$$R[m,\, n] = \langle x[m]\, x^*[n] \rangle, \tag{2}$$

where symbol $\{^*\}$ stands for complex conjugate.

A discrete–time random process $x[i]$ is called the "white noise" if its autocorrelation satisfies

$$R[m,\, n] = \langle |\, x[m]\, |^2 \rangle\, \delta_{mn}, \tag{3}$$

where δ_{mn} is Kronecker's delta symbol:

$$\delta_{mn} = \begin{cases} 1 & (m = n) \\ 0 & (m \neq n) \end{cases}. \tag{4}$$

- Cross–correlation of two discrete–time random processes:

$$x[1],\, x[2],\, x[3],\, \cdots,\, x[n],\, \cdots$$
$$y[1],\, y[2],\, y[3],\, \cdots,\, y[n],\, \cdots$$

is defined by

$$R_{xy}[m,\, n] = \langle x[m]\, y^*[n] \rangle. \tag{5}$$

8

1.1.4 Stationary Discrete–Time Random Process

The concept of stationary random process, which we introduced in Chapter 3 for continuous–time random process, can be transfered to the discrete–time random process in the following way (see, for example, Papoulis, 1984).

- Stationary process.
  A discrete–time random process $x[i]$ is called "stationary" if the expectation

$$\eta[i] = \langle x[i] \rangle = \eta, \tag{6}$$

  is a constant independent of i, and if the autocorrelation

$$R[n+m,\, m] = \langle x[n+m]\, x^*[n] \rangle = R[m], \tag{7}$$

  depends on difference m of arguments only.

  In particular, the stationary random discrete–time process is called the "white noise", if we have

$$R[m] = R[0]\, \delta_{m0}. \tag{8}$$

- Jointly stationary processes.
  Two discrete–time random processes $x[i]$ and $y[j]$ are called "jointly stationary" if they are both stationary, and if their cross–correlation

$$R_{xy}[n+m,\, n] = \langle x[n+m]\, y^*[n] \rangle = R_{xy}[m], \tag{9}$$

  depends on difference m of arguments only.

Similarly to the continuous–time process case, we introduce

- correlation coefficient of a zero–mean stationary discrete–time process $x[i]$:

$$r[m] = \frac{R[m]}{R[0]}, \tag{10}$$

  and

- cross–correlation coefficient of zero–mean jointly stationary discrete–time processes $x[i]$ and $y[j]$:

$$r_{xy}[m] = \frac{R_{xy}[m]}{\sqrt{R_{xx}[0]\, R_{yy}[0]}}, \tag{11}$$

  where, autocovariance is just equal to autocorrelation and cross–covariance is equal to cross–correlation, since we assumed zero–mean processes (i.e. expectations are equal to zero).

1.1.5 Sampling

Let us call the "sampling" an action which makes a discrete–time process by periodically picking up values of a certain continuous–time process with a certain interval of time ("sampling interval"). The discrete–time process thus created is called the "time–sample" of the original continuous–time process.

If a discrete–time random process $x[n]$ is a time–sample of a continuous–time random process $x(t)$ with a sampling interval T, i.e. if

$$x[n] = x(nT), \tag{12}$$

then statistical properties of $x[n]$ are determined by statistical properties (i.e., by probability distribution) of $x(t)$.

- Expectation and autocorrelation of a random time–sample.
  If we denote expectation and autocorrelation of a continuous–time random process $x(t)$ as $\eta(t)$ and $R(t_1, t_2)$, respectively, then expectation and autocorrelation of a random time–sample $x[i] = x(iT)$ are given by

$$\eta[n] = \eta(nT), \tag{13}$$

  and

$$R[m, n] = R(mT, nT), \tag{14}$$

  respectively.

  Proof :

  1. $\eta[n] = \langle x[n] \rangle = \langle x(nT) \rangle = \eta(nT)$.
  2. $R[m, n] = \langle x[m] \, x^*[n] \rangle = \langle x(mT) \, x^*(nT) \rangle = R(mT, nT)$.

- Cross–correlation of random time–samples.
  If we denote cross–correlation of continuous–time random processes $x(t)$ and $y(t)$ as $R_{xy}(t_1, t_2)$, then cross–correlation of random time–samples $x[i] = x(iT)$ and $y[i] = y(iT)$ are given by

$$R_{xy}[m, n] = R_{xy}(mT, nT). \tag{15}$$

  Proof :
  $$R_{xy}[m, n] = \langle x[m] \, y^*[n] \rangle = \langle x(mT) \, y^*(nT) \rangle = R_{xy}(mT, nT).$$

- Stationary random time–sample.

  If a continuous–time random process $x(t)$ is a stationary random process with constant expectation $\langle x(t) \rangle = \eta$ and autocorrelation $\langle x(t+\tau)\,x^*(t) \rangle = R(\tau)$, then a time–sample $x[n] = x(nT)$ is a stationary discrete–time random process with expectation:

$$\eta[n] = \eta, \tag{16}$$

  and autocorrelation:

$$R[n+m,\ n] = R[m] = R(mT). \tag{17}$$

  Proof :

  1. Expectation $\eta[n]$ of the time–sample $x[n]$

$$\eta[n] = \langle x[n] \rangle = \langle x(nT) \rangle = \eta,$$

     is a constant independent of argument n.

  2. Autocorrelation $R[n+m,\ n]$ of the time–sample $x[n]$

$$\begin{aligned}
R[n+m,\ n] &= \langle x[n+m]\,x^*[n] \rangle = \langle x(mT+nT)\,x^*(nT) \rangle \\
&= R(mT) = R[m],
\end{aligned}$$

     depends on difference m of arguments only.

- Jointly–stationary random time–samples.

  If $x(t)$ and $y(t)$ are jointly stationary continuous–time random processes with cross–correlation $\langle x(t+\tau)\,y^*(t) \rangle = R_{xy}(\tau)$, then their time–samples $x[n] = x(nT)$ and $y[n] = y(nT)$ are jointly stationary discrete–time random processes with cross–correlation:

$$R_{xy}[n+m,\ n] = R_{xy}[m] = R_{xy}(mT). \tag{18}$$

  Proof :

  1. Time samples $x[n]$ and $y[n]$ are both stationary discrete–time random processes, as we saw above.

  2. Their cross–correlation

$$\begin{aligned}
R_{xy}[n+m,\ n] &= \langle x[n+m]\,y^*[n] \rangle = \langle x(mT+nT)\,y^*(nT) \rangle \\
&= R_{xy}(mT) = R_{xy}[m],
\end{aligned}$$

     depends on difference m of arguments only.

1.1.6 Comb Function

Infinite number of delta functions arranged with equal intervals along a horizontal axis, shown in Figure 5, is called the "comb function". Thus a comb function $\sqcup\!\sqcup(t; T)$ with period T is given in terms of delta functions $\delta(t)$ by an equation:

$$\sqcup\!\sqcup(t; T) = \sum_{k=-\infty}^{\infty} \delta(t - kT), \tag{19}$$

where k is an integer.
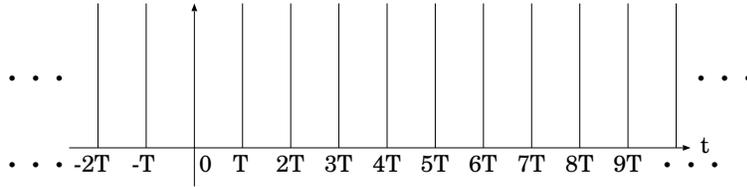
Figure 5: A comb function.

An alternative expression of the comb function is known in a Fourier series form as shown below.

1. Let us expand the comb function to a Fourier series within a range $-\dfrac{T}{2} < t \leq \dfrac{T}{2}$ of interval T:

$$\sum_{k=-\infty}^{\infty} \delta(t - kT) = \sum_{n=-\infty}^{\infty} a_n\, e^{i\frac{2\pi n}{T}t}, \quad \text{for} \quad -\frac{T}{2} < t \leq \frac{T}{2}, \tag{20}$$

where n is an integer.

Folllowing the standard procedure of the Fourier series expansion, we calculate n–th Fourier coefficient a_n by multiplying $e^{-i\frac{2\pi m}{T}t}$, where m is an arbitrary integer, to both sides of equation (20), and integrating them through $\dfrac{T}{2} < t \leq \dfrac{T}{2}$.

2. Then, we have

- LHS $= \displaystyle\sum_{k=-\infty}^{\infty} \int_{-\frac{T}{2}}^{\frac{T}{2}} \delta(t - kT)\, e^{-i\frac{2\pi m}{T}t} dt = \int_{-\frac{T}{2}}^{\frac{T}{2}} \delta(t) e^{-i\frac{2\pi m}{T}t} dt = 1,$

  since t can be equal to kT $(t = kT)$ only when $k = 0$ within the range $\dfrac{T}{2} < t \leq \dfrac{T}{2}$,

12

- RHS $= \displaystyle\sum_{n=-\infty}^{\infty} a_n \int_{-\frac{T}{2}}^{\frac{T}{2}} e^{i \frac{2\pi(n-m)}{T} t} \, dt \; = \; a_m T,$

  since

  $$\int_{-\frac{T}{2}}^{\frac{T}{2}} e^{i \frac{2\pi(n-m)}{T} t} \, dt = \begin{cases} 0 & \text{if } m \neq n \\ T & \text{if } m = n \end{cases}.$$

3. Equating both sides, we have $a_n = \dfrac{1}{T}$, for any n, and hence

$$\sum_{k=-\infty}^{\infty} \delta(t - kT) = \frac{1}{T} \sum_{n=-\infty}^{\infty} e^{i \frac{2\pi n}{T} t}, \quad \text{for} \quad -\frac{T}{2} < t \le \frac{T}{2}. \tag{21}$$

4. Although we derived this equality in a limited range $-\dfrac{T}{2} < t \le \dfrac{T}{2}$, it actually holds for wider range of t. In fact, functions in the both sides of equation (21) do not change if we substitute t with $t + mT$ with an arbitrary integer m. This means that they are both periodic functions with period T. Therefore, equation (21) holds for the whole range of t, i.e. $-\infty < t \le \infty$. Thus, we have a general relation

$$\sum_{k=-\infty}^{\infty} \delta(t - kT) = \frac{1}{T} \sum_{n=-\infty}^{\infty} e^{i \frac{2\pi n}{T} t}, \tag{22}$$

which holds for any t, and, therefore,

$$\sqcup\!\sqcup(t; T) = \frac{1}{T} \sum_{n=-\infty}^{\infty} e^{i \frac{2\pi n}{T} t}, \tag{23}$$

is the alternative expression of the comb function.

1.1.7 Fourier Transform of a Comb Function Is a Comb Function

Fourier transform $\tilde{\sqcup\!\sqcup}(\omega; T)$ of a comb function $\sqcup\!\sqcup(t; T)$ of argument t with period T is a comb function of argument ω with period $\dfrac{2\pi}{T}$ (Figure 6).

Proof :

According to the general formula of Fourier transformation, we have

$$\tilde{\sqcup\!\sqcup}(\omega; T) \;=\; \int_{-\infty}^{\infty} \sqcup\!\sqcup(t; T) e^{-i\omega t} \, dt \;=\; \int_{-\infty}^{\infty} \left(\frac{1}{T} \sum_{n=-\infty}^{\infty} e^{i \frac{2\pi n}{T} t} \right) e^{-i\omega t} \, dt$$

13
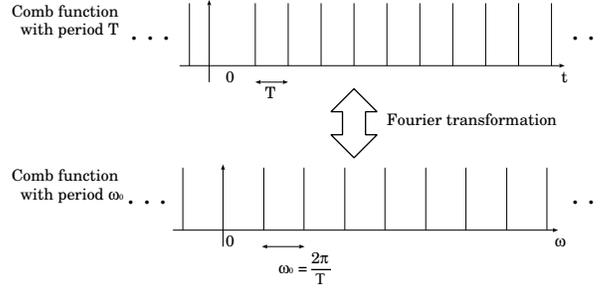
Figure 6: Fourier transform of a comb function of t with period T is a comb function of ω with period $\omega_0 = \dfrac{2\pi}{T}$.

$$
\begin{aligned}
&= \; \frac{1}{T} \sum_{n=-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-i\left(\omega - \frac{2\pi n}{T}\right)t} \, dt \; = \; \frac{2\pi}{T} \sum_{n=-\infty}^{\infty} \delta\left(\omega - \frac{2\pi n}{T}\right) \\
&= \; \omega_0 \sum_{n=-\infty}^{\infty} \delta(\omega - n\,\omega_0),
\end{aligned}
\tag{24}
$$

where we introduced a notation:

$$
\omega_0 = \frac{2\pi}{T},
$$

and used the general formula of the delta function:

$$
\int_{-\infty}^{\infty} e^{-i\omega t} \, dt = 2\pi\delta(\omega).
$$

The RHS of equation (24) is nothing but a comb function of ω with a period $\omega_0 = \dfrac{2\pi}{T}$:

$$
\omega_0 \sum_{n=-\infty}^{\infty} \delta(\omega - n\,\omega_0) = \omega_0 \sqcup\!\sqcup(\omega;\,\omega_0).
\tag{25}
$$

Thus,

$$
\sqcup\!\sqcup(t;\,T) \; \Leftrightarrow \; \omega_0 \sqcup\!\sqcup(\omega;\,\omega_0),
$$

where a symbol $\Leftrightarrow$ implies a Fourier transform pair.

1.1.8 Spectra of Discrete–Time Processes

We introduce following definitions.

14

- Power spectrum.

A power spectrum $S_D(\omega)$ of a stationary random discrete–time process $x[n]$ with autocorrelation $R[m]$ is given by a discrete Fourier transform with an arbitrary parameter $\mathcal{T}$ (Papoulis, 1984):

$$S_D(\omega) = \sum_{m=-\infty}^{\infty} R[m]\, e^{-im\omega\mathcal{T}}. \tag{26}$$

- Cross–power spectrum.

A cross–power spectrum $S_{Dxy}(\omega)$ of jointly stationary discrete–time processes $x[n]$ and $y[n]$ with cross–correlation $R_{xy}[m]$ is given by a discrete Fourier transform with an arbitrary parameter $\mathcal{T}$:

$$S_{Dxy}(\omega) = \sum_{m=-\infty}^{\infty} R_{xy}[m]\, e^{-im\omega\mathcal{T}}. \tag{27}$$

These spectra, as defined by discrete Fourier transforms with an arbitrary parameter $\mathcal{T}$ in equations (26) and (27), are periodic functions of ω with a period $\dfrac{2\pi}{\mathcal{T}}$. They have the same forms as Fourier series, with Fourier coefficients $R[m]$ and $R_{xy}[m]$, respectively. The spectra $S_D(\omega)$ and $S_{Dxy}(\omega)$ are, in general, dependent on the arbitrary parameter $\mathcal{T}$. Later, for particular cases of sampled discrete–time processes (time–samples), we will choose $\mathcal{T}$ to be equal to their sampling intervals. Then, we will be able to establish a relationship between a spectrum of a time–sample and a spectrum of its original continuous–time process.

- Inverse relations.

Autocorrelation $R[m]$ and cross–correlation $R_{Dxy}[m]$ of jointly stationary random discrete–time processes in equations (26) and (27) are given through the power spectrum $S_D(\omega)$ and cross–power spectrum $S_{Dxy}(\omega)$ by inverse relations:

$$R[m] \;=\; \frac{\mathcal{T}}{2\pi} \int_{-\frac{\pi}{\mathcal{T}}}^{\frac{\pi}{\mathcal{T}}} S_D(\omega)\, e^{im\omega\mathcal{T}}\, d\omega, \tag{28}$$

$$R_{xy}[m] \;=\; \frac{\mathcal{T}}{2\pi} \int_{-\frac{\pi}{\mathcal{T}}}^{\frac{\pi}{\mathcal{T}}} S_{Dxy}(\omega)\, e^{im\omega\mathcal{T}}\, d\omega, \tag{29}$$

which are nothing but the formulae for Fourier coefficients in the series expansion.

Proof :

We prove the inverse relation for the power spectrum $S_D(\omega)$ given in equation (28) only, since a proof for the cross–power spectrum $S_{Dxy}(\omega)$ (equation (29)) is given just in a similar way.

1. If $S_D(\omega) = \sum\limits_{n=-\infty}^{\infty} R[n]\, e^{-in\omega\mathcal{T}}$, then $R[m] = \dfrac{\mathcal{T}}{2\pi} \int\limits_{-\frac{\pi}{\mathcal{T}}}^{\frac{\pi}{\mathcal{T}}} S_D(\omega)\, e^{im\omega\mathcal{T}}\, d\omega$.

   In fact,

   $$\frac{\mathcal{T}}{2\pi} \int_{-\frac{\pi}{\mathcal{T}}}^{\frac{\pi}{\mathcal{T}}} S_D(\omega)\, e^{im\omega\mathcal{T}}\, d\omega = \frac{\mathcal{T}}{2\pi} \sum_{n=-\infty}^{\infty} R[n] \int_{-\frac{\pi}{\mathcal{T}}}^{\frac{\pi}{\mathcal{T}}} e^{i(m-n)\omega\mathcal{T}}\, d\omega = R[m],$$

   since

   $$\int_{-\frac{\pi}{\mathcal{T}}}^{\frac{\pi}{\mathcal{T}}} e^{i(m-n)\omega\mathcal{T}}\, d\omega = \begin{cases} \dfrac{2\pi}{\mathcal{T}} & \text{if } n = m, \\[2mm] 0 & \text{otherwise.} \end{cases}$$

2. If $R[m] = \dfrac{\mathcal{T}}{2\pi} \int\limits_{-\frac{\pi}{\mathcal{T}}}^{\frac{\pi}{\mathcal{T}}} S_D(\omega)\, e^{im\omega\mathcal{T}}\, d\omega$, then $S_D(\omega) = \sum\limits_{n=-\infty}^{\infty} R[n]\, e^{-in\omega\mathcal{T}}$.

   We first prove this statement for a limited range of ω confined within an interval $-\dfrac{\pi}{\mathcal{T}} < \omega \le \dfrac{\pi}{\mathcal{T}}$. Inserting first equation to the RHS of second equation, we have

   $$\sum_{n=-\infty}^{\infty} R[n]\, e^{-in\omega\mathcal{T}} = \frac{\mathcal{T}}{2\pi} \int_{-\frac{\pi}{\mathcal{T}}}^{\frac{\pi}{\mathcal{T}}} S_D(\omega') \sum_{n=-\infty}^{\infty} e^{in(\omega'-\omega)\mathcal{T}}\, d\omega'.$$

   Note here that $\sum\limits_{n=-\infty}^{\infty} e^{in(\omega'-\omega)\mathcal{T}}$ is a comb function given in equation (23), since, introducing a notation $\omega_0 = \dfrac{2\pi}{\mathcal{T}}$, we have

   $$\begin{aligned} \sum_{n=-\infty}^{\infty} e^{in(\omega'-\omega)\mathcal{T}} &= \sum_{n=-\infty}^{\infty} e^{i\frac{2\pi n}{\omega_0}(\omega'-\omega)} = \omega_0 \sqcup\!\sqcup(\omega'-\omega\,;\,\omega_0) \\[2mm] &= \omega_0 \sum_{k=-\infty}^{\infty} \delta(\omega'-\omega-k\,\omega_0) \\[2mm] &= \frac{2\pi}{\mathcal{T}} \sum_{k=-\infty}^{\infty} \delta\left(\omega'-\omega-k\,\frac{2\pi}{\mathcal{T}}\right). \end{aligned} \qquad (30)$$

16

Therefore, we obtain

$$
\begin{aligned}
\sum_{n=-\infty}^{\infty} R[n]\, e^{-in\omega\mathcal{T}} &= \sum_{k=-\infty}^{\infty} \int_{-\frac{\pi}{\mathcal{T}}}^{\frac{\pi}{\mathcal{T}}} S_D(\omega')\delta\left(\omega' - \omega - k\frac{2\pi}{\mathcal{T}}\right) d\omega' \\
&= S_D(\omega),
\end{aligned}
$$

since the delta function in the integrand takes non–zero value when

$$
\omega' = \omega + k\frac{2\pi}{\mathcal{T}},
$$

and this condition holds only when $k = 0$, provided that ω is confined within the interval $-\dfrac{\pi}{\mathcal{T}} < \omega \le \dfrac{\pi}{\mathcal{T}}$.

Now, if we extend the function $S_D(\omega)$ to a periodic function with a period of $\dfrac{2\pi}{\mathcal{T}}$, beyond the initially imposed interval $-\dfrac{\pi}{\mathcal{T}} < \omega \le \dfrac{\pi}{\mathcal{T}}$, we have

$$
\sum_{n=-\infty}^{\infty} R[n]\, e^{-in\omega\mathcal{T}} = S_D(\omega),
$$

for any range of ω.

1.1.9 Spectra of Sampled Data

Let us consider discrete–time processes $x[n]$ and $y[n]$, which are time–samples obtained by sampling jointly stationary continuous–time random processes $x(t)$ and $y(t)$ with a sampling interval T:

$$
x[n] = x(nT), \quad \text{and} \quad y[n] = y(nT).
$$

Let autocorrelation of $x[n]$, and cross–correlation of $x[n]$ and $y[n]$, be $R[m]$, and $R_{xy}[m]$, respectively. They satisfy

$$
R[m] = R(mT), \quad \text{and} \quad R_{xy}[m] = R_{xy}(mT),
$$

in view of equations (17) and (18). If we choose the arbitrary parameter $\mathcal{T}$ in the power spectrum $S_D(\omega)$ and the cross–power spectrum $S_{Dxy}(\omega)$ of the discrete–time processes $x[n]$ and $y[n]$, as defined in equations (26) and (27), to be equal to the sampling interval T, i.e.,

$$
\mathcal{T} = T, \tag{31}
$$

17

then $S_D(\omega)$ and $S_{Dxy}(\omega)$ are related to power spectrum $S(\omega)$ and cross–power spectrum $S_{xy}(\omega)$ of the original continuous–time processes $x(t)$ and $y(t)$ by equations:

$$S_D(\omega) = \frac{1}{T} \sum_{k=-\infty}^{\infty} S(\omega + k\,\omega_0), \tag{32}$$

$$S_{Dxy}(\omega) = \frac{1}{T} \sum_{k=-\infty}^{\infty} S_{xy}(\omega + k\,\omega_0), \tag{33}$$

where $\omega_0 = \dfrac{2\pi}{T}$.

Proof:

We prove equation (32) for the power spectrum $S_D(\omega)$ only, since a proof of equation (33) for the cross–power spectrum $S_{Dxy}(\omega)$ is given just in a similar way.

According to equations (17), (26), and (31), the power spectrum of the time–sample $x[n] = x(nT)$ is given by

$$S_D(\omega) = \sum_{n=-\infty}^{\infty} R[n]\, e^{-in\omega T} = \sum_{n=-\infty}^{\infty} R(nT)\, e^{-in\omega T},$$

where T is the sampling interval. Describing the autocorrelation $R(\tau)$ of the continuous–time stationary random process $x(t)$ in terms of the power spectrum $S(\omega)$ through inverse Fourier transformation:

$$R(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S(\omega')\, e^{i\omega'\tau}\, d\omega',$$

we have

$$
\begin{aligned}
S_D(\omega) &= \frac{1}{2\pi} \sum_{n=-\infty}^{\infty} \int_{-\infty}^{\infty} S(\omega')\, e^{in(\omega'-\omega)T}\, d\omega' \\
&= \frac{1}{2\pi} \int_{-\infty}^{\infty} S(\omega') \sum_{n=-\infty}^{\infty} e^{in(\omega'-\omega)T}\, d\omega'.
\end{aligned}
$$

According to equation (30), $\displaystyle\sum_{n=-\infty}^{\infty} e^{in(\omega'-\omega)T}$ is a comb function:

$$\sum_{n=-\infty}^{\infty} e^{in(\omega'-\omega)T} = \frac{2\pi}{T} \sum_{k=-\infty}^{\infty} \delta\left(\omega' - \omega - k\,\frac{2\pi}{T}\right).$$

18

Therefore,

$$
\begin{aligned}
S_D(\omega) &= \frac{1}{T} \sum_{k=-\infty}^{\infty} \int_{-\infty}^{\infty} S(\omega')\, \delta(\omega' - \omega - k\frac{2\pi}{T})\, d\omega' \\
&= \frac{1}{T} \sum_{k=-\infty}^{\infty} S\left(\omega + k\frac{2\pi}{T}\right) = \frac{1}{T} \sum_{k=-\infty}^{\infty} S(\omega + k\,\omega_0),
\end{aligned}
$$

where $\omega_0 = \dfrac{2\pi}{T}$.

1.1.10   Inverse Relations for Spectra of Sampled Data

The inverse relation for the power spectrum $S_D(\omega)$ of a discrete–time stationary random process $x[n]$:

$$
R[m] = \frac{T}{2\pi} \int_{-\frac{\pi}{T}}^{\frac{\pi}{T}} S_D(\omega)\, e^{im\omega T}\, d\omega,
$$

as given in equation (28), must yield an autocorrelation which satisfies $R[m] = R(mT)$, if the process $x[n]$ is a time–sample $x[n] = x(nT)$ of a continuous–time stationary random process $x(t)$.

Proof :

Substituting equation (32) to the inverse relation, we obtain

$$
\begin{aligned}
R[m] &= \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \int_{-\frac{\pi}{T}}^{\frac{\pi}{T}} S(\omega + k\frac{2\pi}{T})\, e^{im\omega T}\, d\omega \\
&= \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \int_{-\frac{\pi}{T}+k\frac{2\pi}{T}}^{\frac{\pi}{T}+k\frac{2\pi}{T}} S(\omega')\, e^{im(\omega' - k\frac{2\pi}{T})T}\, d\omega' \\
&= \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \int_{-\frac{\pi}{T}+k\frac{2\pi}{T}}^{\frac{\pi}{T}+k\frac{2\pi}{T}} S(\omega')\, e^{i(m\omega'T - 2\pi km)}\, d\omega' \\
&= \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \int_{-\frac{\pi}{T}+k\frac{2\pi}{T}}^{\frac{\pi}{T}+k\frac{2\pi}{T}} S(\omega')\, e^{im\omega'T}\, d\omega' \\
&= \frac{1}{2\pi} \int_{-\infty}^{\infty} S(\omega')\, e^{im\omega'T}\, d\omega' = R(mT).
\end{aligned}
$$

19

Similarly, we confirm that the inverse relation:

$$R_{xy}[m] = \frac{T}{2\pi} \int\limits_{-\frac{\pi}{T}}^{\frac{\pi}{T}} S_{Dxy}(\omega)\, e^{im\omega T}\, d\omega,$$

in equation (29), gives a cross–correlation of time–samples $x[n]$ and $y[n]$: $R_{xy}[m] = R_{xy}(mT)$.

1.1.11 Sampling Theorem

Shannon (1949) gave a beautiful proof of the sampling theorem, which he formulated as follows:
"If a function $f(t)$ contains no frequencies higher than B cps, it is completely determined by giving its ordinates at a series of points spaced $1/2B$ seconds apart."
A mathematical proof showing that "this is not only approximately, but exactly, true" was given as follows.

"Let $F(\omega)$ be the spectrum of $f(t)$. Then

$$
\begin{aligned}
f(t) &= \frac{1}{2\pi} \int\limits_{-\infty}^{\infty} F(\omega)\, e^{i\omega t}\, d\omega \\
&= \frac{1}{2\pi} \int\limits_{-2\pi B}^{2\pi B} F(\omega)\, e^{i\omega t}\, d\omega,
\end{aligned}
$$

since $F(\omega)$ is assumed zero outside the band B. If we let

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

where n is any positive or negative integer, we obtain

$$f\left(\frac{n}{2B}\right) = \frac{1}{2\pi} \int\limits_{-2\pi B}^{2\pi B} F(\omega)\, e^{i\omega \frac{n}{2B}}\, d\omega.$$

On the left are the values of $f(t)$ at the sampling points. The integral on the right will be recognized as essentially the n–th coefficient in a Fourier–series expansion of the function $F(\omega)$, taking the interval $-B$ to $+B$ as a fundamental period. This means that the values of the samples $f(n/2B)$ determine the Fourier coefficients in the series expansion of $F(\omega)$. Thus they determine

20

$F(\omega)$, since $F(\omega)$ is zero for frequencies greater than B, and for lower frequencies $F(\omega)$ is determined if its Fourier coefficients are determined. But $F(\omega)$ determines the original function $f(t)$ completely, since a function is determined if its spectrum is known. Therefore the original samples determine the function $f(t)$ completely."

Shannon (1949) mentioned that Nyquist had pointed out the fundamental importance of the time interval $1/2B$ seconds in connection with telegraphy, and proposed to call this the "**Nyquist interval**" corresponding to the band B.

Nowadays, we formulate the sampling theorem in a slightly wider form (Figure 7).

Analog continuous-time signal

Spectrum of analog continuous-time data

Sampled signal
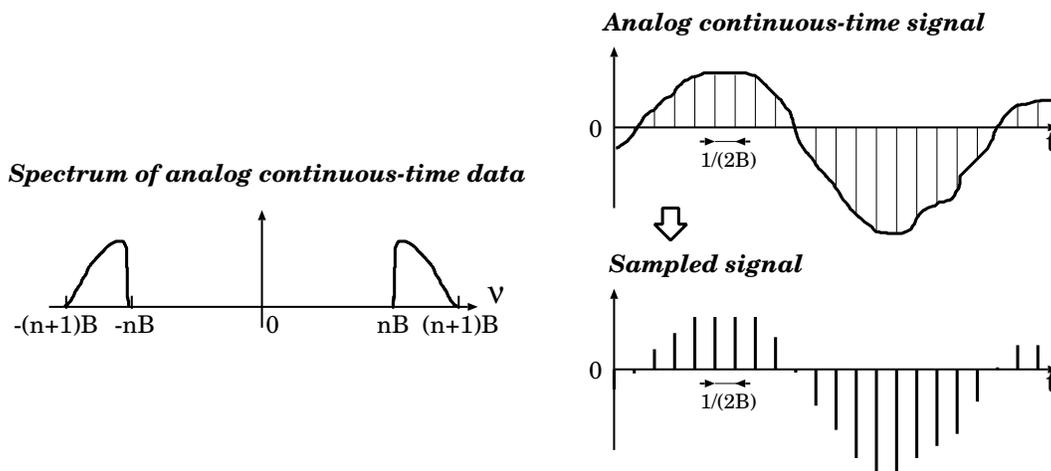
Figure 7: Sampling theorem.

Sampling Theorem:
All the information in an analog continuous–time signal with a passband spectrum limited within a frequency range $nB \leq \nu < (n+1)B$, where B is a bandwidth and $n \geq 0$ is an integer, can be preserved, provided the signal is sampled with the Nyquist interval $1/(2B)$.

Here we assume a real process with even or Hermitian symmetric spectrum with respect to frequency. Thus, "spectrum" here implies positive frequency part of the spectrum. Sampling frequecncy $2B$, with Nyquist interval $1/(2B)$, is called the "Nyquist rate".

The proof of the above theorem is given by equations (32) and (33):

$$S_D(\omega) = \frac{1}{T} \sum_{k=-\infty}^{\infty} S\left(\omega + k\frac{2\pi}{T}\right),$$

$$S_{Dxy}(\omega) = \frac{1}{T} \sum_{k=-\infty}^{\infty} S_{xy}\left(\omega + k\frac{2\pi}{T}\right),$$

as illustrated in Figure 8.

In fact, equations (32) and (33), and Figure 8, show

- if positive frequency part of the analog continuous–time spectrum $S(\omega)$ is confined within a passband $nB \leq \nu < (n+1)B$, where $n \geq 0$ is an integer, B is bandwidth, and ν is frequency, and

- if sampling interval T is equal to the Nyquist interval: $T = 1/(2B)$,

then the analog continuous–time spectrum $S(\omega)$ is completely preserved in the spectrum $S_D(\omega)$ of the sampled data (First and second panels from the top of Figure 8). Therefore, all the information of the original analog continuous–time signal is preserved in the sampled data. This proves the sampling theorem.

Note, however, that the spectrum $S_D(\omega)$ of the sampled data in a range $0 \leq \nu < B$ is inverted in frequency compared with the original analog continuous–time spectrum $S(\omega)$, if the integer n is odd (second panel from the top of Figure 8).

On the other hand,

- if the sampling interval T is larger than the Nyquist interval, i.e., $T > 1/(2B)$, as shown in the third panel from the top of Figure 8, or, even though $T = 1/(2B)$, if the analog continuous–time spectrum is confined within $aB \leq \nu < (a+1)B$, where a is not an integer, as shown in the bottom panel of Figure 8,

then foots of spectral components with different n's in equations (32) and (33) are overlapped with each other (this is called the "aliasing"). Therefore, information of the analog continuous–time spectrum $S(\omega)$ is no longer preserved in the spectrum $S_D(\omega)$ of the sampled data.

1.1.12 Optimum Sampling Interval

In order to see that the Nyquist interval is the optimum interval for sampling, let us consider an analog continuous–time spectrum which is confined within

22

Spectrum of continuous-time data   **Spectrum of sampled data**
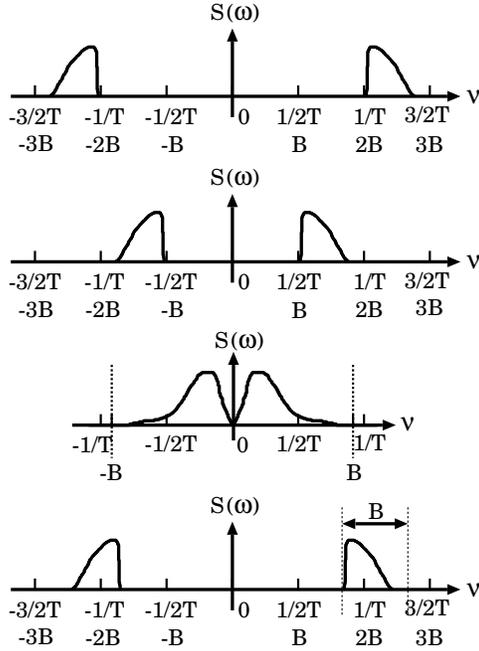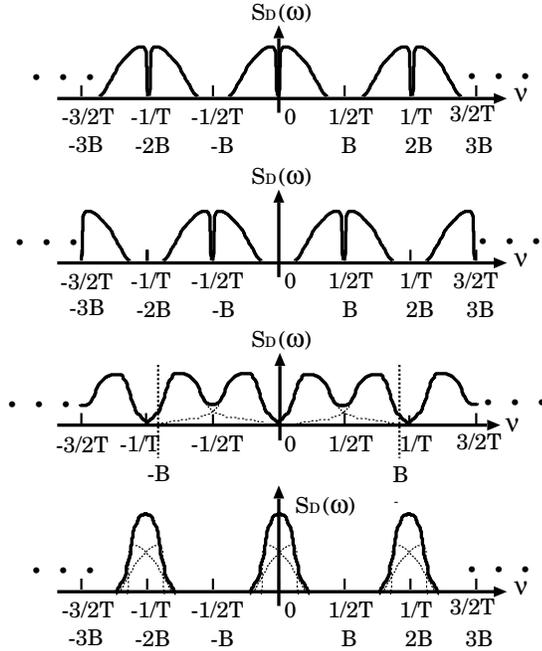
Figure 8: Four cases of relation between spectrum of analog continuous–time data and spectrum of sampled data given by equations (32) and (33). Top: analog continuous–time spectrum is confined within a passband $2mB \leq |\nu| < (2m+1)B$ and sampling interval T is equal to Nyquist interval $T = 1/(2B)$. Second from the top: analog continuous–time spectrum is confined within a passband $(2m+1)B \leq |\nu| < 2(m+1)B$ and $T = 1/(2B)$. Third from the top: $T > 1/(2B)$. Bottom: analog continuous–time spectrun is confined within a passband with boundaries of non–integer multiples of B, i.e., $aB \leq |\nu| < (a+1)B$, and $T = 1/(2B)$. Here we adopted notations, ν: frequency, B: bandwidth of the analog continuous–time spectrum, $m \geq 0$: an integer, and a: a non–integer number. We assume a real process, and, therefore, an even or Hermitian symmetric spectrum with respect to frequency. All information of the original analog continuous–time data is completely preserved in the sampled data in the first two cases, but a part of the information is lost after sampling in the last two cases.
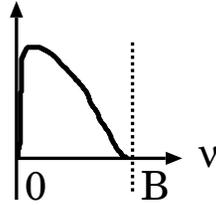
Figure 9: A band–limited baseband spectrum confined within a frequency range $0 \le \nu < B$. Here, positive frequency part only is shown.

a baseband $0 \le \nu < B$. Here, the "baseband", or otherwise called the "video–band", implies a frequency band containing 0 Hz (or "DC", which means "direct current") as the lowest frequency, such as shown in Figure 9. This is a particular case of the passband spectrum within $nB \le \nu < (n+1)B$ when $n = 0$. Also, we assume that the bandwidth B here corresponds to an actual extent of the spectrum, that means the spectrum is non–zero in the inside of the interval $0 \le \nu < B$, but zero in the outside.

Then, we can conceive three cases which are shown in Figure 10.
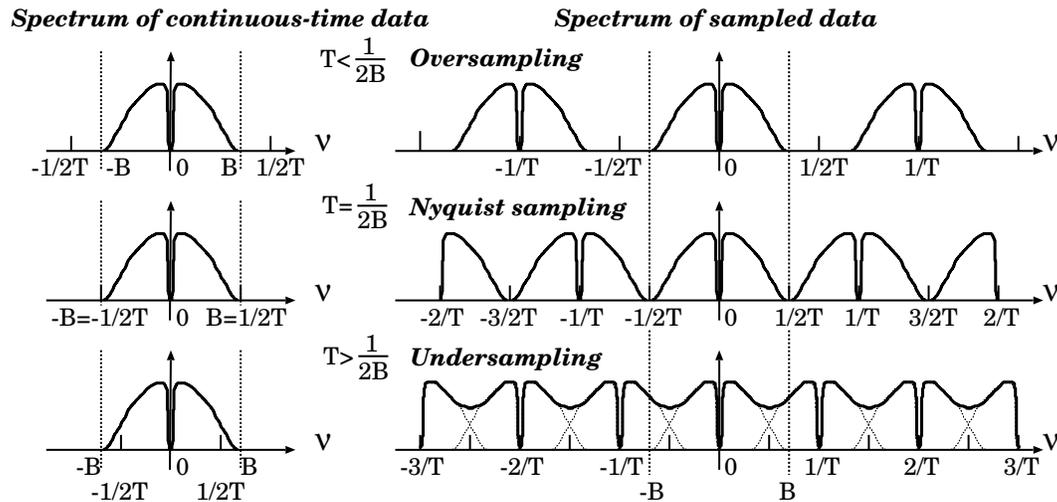
Figure 10: Spectra of sampled data in oversampling $T < 1/(2B)$ (top), Nyquist sampling $T = 1/(2B)$ (middle), and undersampling $T > 1/(2B)$ (bottom).

1. **Oversampling**

   If we sample an analog continuous–time signal with sampling interval

24

smaller than the Nyquist interval $T < 1/(2B)$, as shown in top panel of Figure 10, we will have larger number of data points per unit duration of time, but information contained is not improved at all, compared with the Nyquist sampling (sampling with Nyquist interval) case shown in the middle panel of Figure 10. Such a samping with an interval $T < 1/(2B)$ is called the "oversampling".

2. **Undersampling**
   On the contrary, if the sampling interval is larger than the Nyquist interval $T > 1/(2B)$, a part of information in the original analog continuous–time signal is lost in the sampled data due to the aliasing, as shown in bottom panel of Figure 10. This case is called the "undersampling".

3. **Nyquist sampling**
   Therefore, the Nyquist sampling, i.e. sampling with the Nyquist interval $T = 1/(2B)$, is the optimum sampling for an analog continuous–time signal with a band–limited baseband spectrum (middle panel of Figure 10).

1.1.13   Sampling Function

We saw above that autocorrelation $R(\tau)$ and cross–correlation $R_{xy}(\tau)$ of jointly stationary continuous–time random processes $x(t)$ and $y(t)$ with band–limited baseband spectra are completely restored from autocorrelation $R[n] = R(nT)$ and cross–correlation $R_{xy}[n] = R_{xy}(nT)$ of time samples of the processes $x[n] = x(nT)$ and $y[n] = y(nT)$, provided that the sampling interval T is shorter than or equal to the Nyquist interval, i.e. $T \leq 1/(2B)$, where B is bandwidth of the spectra. But how can we functionally express $R(\tau)$ and $R_{xy}(\tau)$ through $R[n]$ and $R_{xy}[n]$?

The answer is given by the so–called "second part of the sampling theorem", which states that they satisfy equations:

$$R(\tau) = \sum_{n=-\infty}^{\infty} R[n] \frac{\sin\left[\frac{\pi}{T}(\tau - nT)\right]}{\frac{\pi}{T}(\tau - nT)}, \tag{34}$$

and

$$R_{xy}(\tau) = \sum_{n=-\infty}^{\infty} R_{xy}[n] \frac{\sin\left[\frac{\pi}{T}(\tau - nT)\right]}{\frac{\pi}{T}(\tau - nT)}. \tag{35}$$

The sinc function here:

$$S_{An}(\tau) \equiv \frac{\sin\left[\frac{\pi}{T}(\tau - nT)\right]}{\frac{\pi}{T}(\tau - nT)}, \tag{36}$$

is called the "sampling function".

Proof:

We prove equation (34) for the autocorrelation only, since equation (35) for the cross–correlation can be derived exactly in the same way.

According to equation (32), power spectrum $S(\omega)$ of a stationary random continuous–time process $x(t)$ and power spectrum $S_D(\omega)$ of its time–sample $x[n] = x(nT)$ with a sampling interval T are related to each other by

$$S_D(\omega) = \frac{1}{T} \sum_{k=-\infty}^{\infty} S\left(\omega + k\frac{2\pi}{T}\right).$$

Therefore, if the sampling interval T is shorter than or equal to the Nyquist interval, $T \leq 1/(2B)$, we obtain the full analog continuous–time baseband spectrum $S(\omega)$ by multiplying to the spectrum of the time sample $S_D(\omega)$ a rectangular window function $P(\omega)$ satisfying

$$P(\omega) = \begin{cases} 1 & -\frac{\pi}{T} < \omega \leq \frac{\pi}{T}, \\ 0 & \text{otherwise}, \end{cases} \tag{37}$$

as illustrated in Figure 11. In fact, we have

$$S(\omega) = T\,P(\omega)\,S_D(\omega), \tag{38}$$

from equation (32).
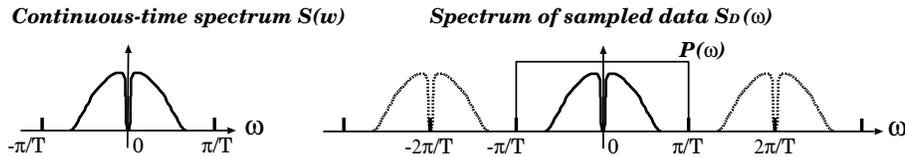
Figure 11: Analog Continuous–Time spectrum $S(\omega)$ (left panel) is fully restored from spectrum of time samples $S_D(\omega)$ (right panel) by multiplying a rectangular window function $P(\omega)$ given in equation (37).

If we introduce Fourier transform pairs $S(\omega) \Leftrightarrow R(\tau)$, $P(\omega) \Leftrightarrow p(\tau)$ and $S_D(\omega) \Leftrightarrow R_D(\tau)$, equation (38) implies

$$R(\tau) = T\, p(\tau) \,*\, R_D(\tau) = T \int\limits_{-\infty}^{\infty} p(\tau - \alpha)\, R_D(\alpha)\, d\alpha, \qquad (39)$$

in view of the convolution theorem which we saw in Chapter 3, where symbol "$*$" stands for the operation of convolution.

We saw in Chapter 3 that inverse Fourier transform of the rectangular window function is a sinc function:

$$p(\tau) = \frac{1}{2\pi} \int\limits_{-\infty}^{\infty} P(\omega) e^{i\omega\tau}\, d\omega = \frac{1}{2\pi} \int\limits_{-\frac{\pi}{T}}^{\frac{\pi}{T}} e^{i\omega\tau}\, d\omega = \frac{\sin\left(\frac{\pi\tau}{T}\right)}{\pi\tau}. \qquad (40)$$

On the other hand, inverse Fourier transform of equation (32) yields

$$
\begin{aligned}
R_D(\tau) &= \frac{1}{2\pi} \int\limits_{-\infty}^{\infty} S_D(\omega) e^{i\omega\tau}\, d\omega = \frac{1}{2\pi T} \int\limits_{-\infty}^{\infty} \sum_{k=-\infty}^{\infty} S(\omega + k\frac{2\pi}{T}) e^{i\omega\tau}\, d\omega \\
&= \frac{1}{T} R(\tau) \sum_{k=-\infty}^{\infty} e^{-i\frac{2\pi k}{T}\tau} = R(\tau) \sum_{n=-\infty}^{\infty} \delta(\tau - nT), \qquad (41)
\end{aligned}
$$

where we used the shift theorem and the property of the comb function given in equation (22):

$$\frac{1}{T} \sum_{k=-\infty}^{\infty} e^{-i\frac{2\pi k}{T}\tau} = \sum_{n=-\infty}^{\infty} \delta(\tau - nT).$$

Therefore, equation (39) is reduced to

$$
\begin{aligned}
R(\tau) &= T \sum_{n=-\infty}^{\infty} \int\limits_{-\infty}^{\infty} \frac{\sin\left(\frac{\pi}{T}(\tau - \alpha)\right)}{\pi(\tau - \alpha)} R(\alpha)\, \delta(\alpha - nT)\, d\alpha \\
&= \sum_{n=-\infty}^{\infty} R(nT) \frac{\sin\left(\frac{\pi}{T}(\tau - nT)\right)}{\frac{\pi}{T}(\tau - nT)},
\end{aligned}
$$

which proves equation (34), since $R(nT) = R[n]$.

Now we are in position to answer to an interesting question: why accuracy of delay determination in VLBI can be much superior (i.e. smaller)

than a sampling interval of digitized voltage signals, from which the delay is determined? For example, typical delay accuracy of Mark III VLBI system has been 0.1 nanosecond (1×10^{-10} sec), while a typical sampling interval in the Mark III observation has been 125 nanosecond ($= 1/(2 \times 2 \text{ MHz})$). Details apart, an essential point of the answer is in the sampling theorem: Nyquist sampled data are capable of determining the delay as accurately as continuous–time data, from which the sampled data are formed, since they are equivalent to each other in view of the sampling theorem.

The $2B$ optimal rate and the sampling function have been independently discovered by a number of researchers in different countries, besides Shannon (1949). The history even goes back to the 19th Century. Interested readers could consult with a review paper by Meijering (2002).

1.1.14 Correlations of Nyquist Sampled Data with Rectangular Passband Spectra

Let us consider continuous–time stationary random processes $x(t)$ and $y(t)$ with rectangular power spectra $S(\omega)$ with a passband of bandwidth B:

$$S(\omega) = \begin{cases} a & 2\pi nB \leq \mid \omega \mid < 2\pi(n+1)B, \\ 0 & \text{otherwise,} \end{cases} \tag{42}$$

where n is an integer, and $n = 0$ corresponds to the particular case of the baseband spectrum (Figure 12).
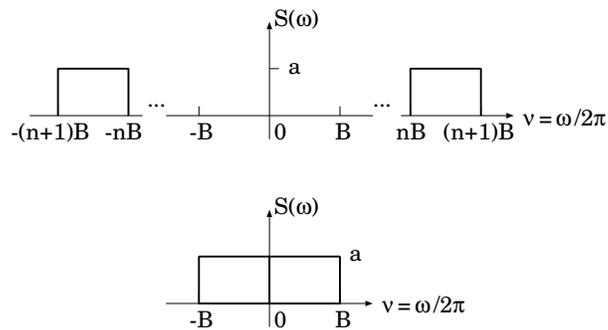
Figure 12: Rectangular passband (top) and baseband (bottom) power spectra.

If we sample the data with Nyquist interval $T = 1/(2B)$, then we obtain following properties for correlations of the time samples.

Autocorrelation:

Autocorrelation $R(\tau)$ of an original continuous–time process $x(t)$ is obtained by an inverse Fourier transformation of the passband power spectrum $S(\omega)$ (top panel of Figure 12):

$$
\begin{aligned}
R(\tau) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} S(\omega)\, e^{i\omega\tau}\, d\omega = \frac{1}{\pi} \Re \int_{0}^{\infty} S(\omega)\, e^{i\omega\tau}\, d\omega \\[2mm]
&= \frac{a}{\pi} \Re \int_{2\pi n B}^{2\pi(n+1)B} e^{i\omega\tau}\, d\omega = \frac{a}{\pi} \Re \left[e^{i2\pi\left(n+\frac{1}{2}\right)B\tau} \int_{-\pi B}^{\pi B} e^{i\omega'\tau}\, d\omega' \right] \\[2mm]
&= \frac{a}{\pi} \Re \left[e^{i2\pi\left(n+\frac{1}{2}\right)B\tau} \frac{e^{i\pi B\tau} - e^{-i\pi B\tau}}{i\,\tau} \right] \\[2mm]
&= 2aB \frac{\sin(\pi B\tau)}{\pi B\tau} \cos\left[2\pi \left(n + \frac{1}{2} \right) B\tau \right],
\end{aligned}
\tag{43}
$$

which has the familiar "white–fringe" form with the fringe pattern enclosed by the bandwidth pattern, as we saw in Chapter 3.

In the baseband spectrum (bottom panel of Figure 12), we have $n = 0$, and the autocorrelation of the continuous–time process has a sinc function form:

$$
R(\tau) = 2aB \frac{\sin(2\pi B\tau)}{2\pi B\tau}.
\tag{44}
$$

For the correlation coefficient of the continuous–time process:

$$
r(\tau) = \frac{R(\tau)}{R(0)},
$$

we have,

$$
r(\tau) = \frac{\sin(\pi B\tau)}{\pi B\tau} \cos\left[2\pi \left(n + \frac{1}{2} \right) B\tau \right],
\tag{45}
$$

in a case of the general passband spectrum, and

$$
r(\tau) = \frac{\sin(2\pi B\tau)}{2\pi B\tau},
\tag{46}
$$

in the particular case of the baseband spectrum.

Now, if we sample the continuous–time process $x(t)$ with the Nyquist interval $T = 1/(2B)$, correlation coefficient of the time sample is given by

$$
r[m] = r(mT) = \frac{\sin\left(\frac{m\pi}{2}\right)}{\frac{m\pi}{2}} \cos\left[m\pi \left(n + \frac{1}{2} \right) \right],
\tag{47}
$$

29
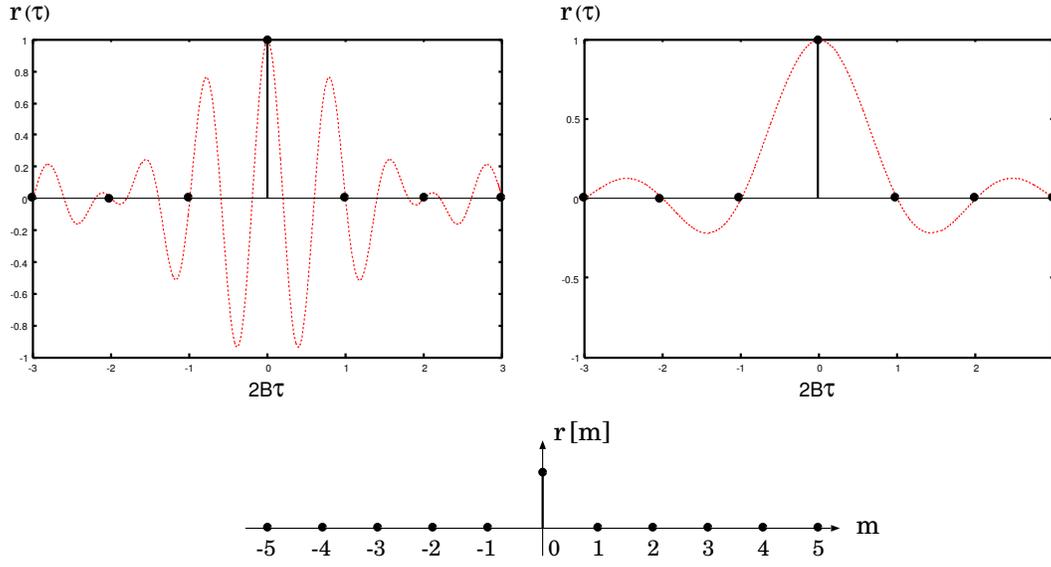
Figure 13: Correlation coefficient $r(\tau)$ of a continuous–time process with rectangular passband spectrum when $n = 2$ (top left) and baseband spectrum (top right), as given in equations (45) and (46). When this process is sampled with the Nyquist interval $T = 1/(2B)$, correlation coefficient $r[m]$ of the time sample has the "white–noise" form which is equal to 1 if $m = 0$, and equal to 0 if $m \neq 0$ (bottom), since $r[m] = r(mT) = 0$ for all m except for $m = 0$, as shown in the top two panels.

for the passband spectrum, and

$$r[m] = r(mT) = \frac{\sin(m\pi)}{m\pi}, \tag{48}$$

for the baseband spectrum, in particular. Both equations (47) and (48) show the "white–noise" form of the time sample:

$$r[m] = \delta_{m0} = \begin{cases} 1 & \text{if } m = 0, \\ 0 & \text{if } m \neq 0, \end{cases} \tag{49}$$

as given in equation (8), where δ_{ij} is Kronecker's delta symbol. This shows that different sample points are not correlated, and therefore independent of each other, in time samples of Nyquist sampled data with rectangular passband spectra.

Relationship between the correlation coefficient of the original continuous–time data and that of the sampled data is illustrated in Figure 13.

30

Cross–correlation:

If a cross–power spectrum $S_{xy}(\omega)$ of jointly stationary continuous–time random processes $x(t)$ and $y(t)$ is real (i.e., has zero phase), and rectangular with bandwidth B, such as shown in Figure 12, the situation is much the same with the autocorrelation case discussed above, and their cross–correlation has the same functional form as equation (43) or (44). Therefore, cross–correlation coefficient of their time samples has the "white–noise" form, proportional to the one given in equation (49).

Let us now consider a little more general case, when amplitude $A(\omega)$ of the cross–power spectrum $S_{xy}(\omega)$ is rectangular, as given in equation (42), but phase is non–zero due to some delay τ_d between correlated signals in processes $x(t)$ and $y(t)$, which may in general contain both the signals and uncorrelated noises, just like in an interferometer problem. In such a case, the cross–power spectrum, which contains the signal contribution only, has a form:

$$S_{xy}(\omega) = A(\omega)\, e^{-i\omega\tau_d}, \tag{50}$$

as we saw in Chapter 3.

Strictly speaking, actual passband spectra to be sampled in realistic interferometers are IF spectra after the frequency–conversion, and hence their phase spectra usually do not cross the origin, i.e., phases are non–zero at $\omega = 0$, unlike in equation (50), as we discussed in Chapter 3. Nevertheless, we adopt equation (50) for simplicity, assuming an idealized case of "RF correlation", or a case when the "fringe stopping" is ideally performed so that the phase crosses the origin, but the phase slope still remains due to an imperfect "delay tracking".

Then, in view of the shift theorem, the cross–correlation $R_{xy}(\tau)$:

$$R_{xy}(\tau) = \frac{1}{2\pi} \int\limits_{-\infty}^{\infty} S_{xy}(\omega)\, e^{i\omega\tau}\, d\omega,$$

should have a similar form as given in equation (43) or (44), but argument τ is replaced by $\tau - \tau_d$. Thus, cross–correlation coefficient is given by

$$r_{xy}(\tau) \;=\; \frac{R_{xy}(\tau)}{\sqrt{R_{xx}(0)\, R_{yy}(0)}}$$

$$= \begin{cases} \rho \dfrac{\sin[\pi B\,(\tau - \tau_d)]}{\pi B\,(\tau - \tau_d)} \cos\left[2\pi\left(n + \dfrac{1}{2}\right) B\,(\tau - \tau_d)\right] & \text{(passband)}, \\[4mm] \rho \dfrac{\sin[2\pi B\,(\tau - \tau_d)]}{2\pi B\,(\tau - \tau_d)} & \text{(baseband)}, \end{cases} \tag{51}$$

31

where ρ is the maximum cross–correlation coefficient:

$$\rho = \frac{R_{xy}(\tau_d)}{\sqrt{R_{xx}(0)\,R_{yy}(0)}}. \tag{52}$$

Note that the cross–correlation coefficient in the case of the passband spectrum with $n \neq 0$ again shows the "white–fringe" form with the cosine "fringe pattern" enclosed within the sinc function envelope of "bandwidth pattern".
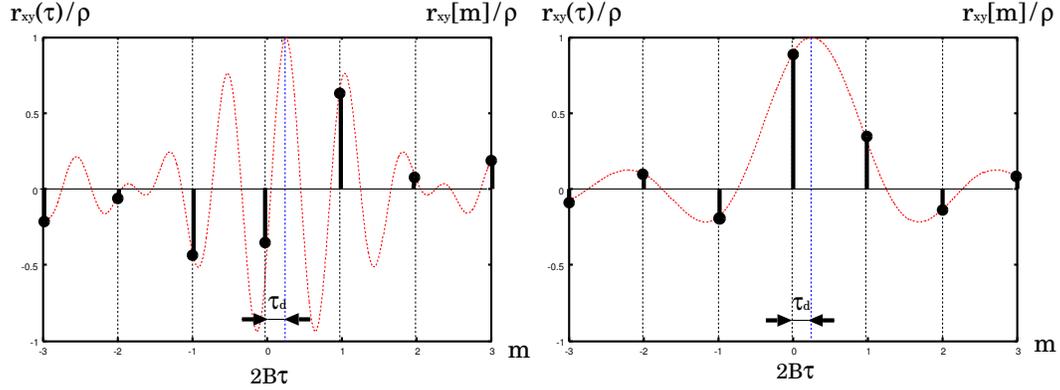
Figure 14: Cross–correlation coefficient $r_{xy}(\tau)$ of jointly stationary continuous–time processes $x(t)$ and $y(t)$ with rectangular passband spectrum when $n = 2$ (left: dotted line) and baseband spectrum with $n = 0$ (right: dotted line), as given by equation (51). τ_d is a time delay between correlated processes $x(t)$ and $y(t)$. Also shown by bars is cross–correlation coefficient $r_{xy}[m] = r_{xy}(mT)$ of time samples $x[i] = x(iT)$ and $y[j] = y(jT)$ sampled with the Nyquist interval $T = 1/(2B)$, where B is a bandwidth of the rectangular spectrum (equation (53)). Horizontal axis shows the time difference τ normalized by the sampling interval $T = 1/(2B)$. Vertical axis is the cross–correlation coefficient $r_{xy}(\tau)$ normalized by its maximum value: $\rho = R_{xy}(\tau_d)/\sqrt{R_{xx}(0)\,R_{yy}(0)}$, where $R_{xy}(\tau)$, $R_{xx}(\tau)$, and $R_{yy}(\tau)$ are cross–correlation and autocorrelations of $x(t)$ and $y(t)$, correspondingly.

Therefore, if we sample $x(t)$ and $y(t)$ with the Nyquist interval $T = 1/(2B)$, cross–correlation coefficient of the time samples is

$$r_{xy}[m] \quad = \quad r_{xy}(mT)$$

32

$$
= \begin{cases} \rho \dfrac{\sin\left[\frac{\pi}{2}\left(m - \frac{\tau_d}{T}\right)\right]}{\frac{\pi}{2}\left(m - \frac{\tau_d}{T}\right)} \cos\left[\pi\left(n + \frac{1}{2}\right)\left(m - \frac{\tau_d}{T}\right)\right] & \text{(passband)}, \\[4mm] \rho \dfrac{\sin\left[\pi\left(m - \frac{\tau_d}{T}\right)\right]}{\pi\left(m - \frac{\tau_d}{T}\right)} & \text{(baseband)}. \end{cases} \tag{53}
$$

Relationship between the cross–correlation coefficient of the original continuous–time data and that of the sampled data is illustrated in Figure 14.

The cross–correlation coefficient $r_{xy}[m]$ of the time samples given by equation (53) no longer has the symmetric "white–noise" form, as shown in equation (49) and in the bottom panel of Figure 13, due to the parallel shift of the cross–correlation coefficient of the continuous–time data along the horizontal axis which is caused by the delay τ_d. Also, it now depends upon n, i.e. upon location of the passband spectrum on the frequency axis, since the "fringe pattern" in the "white–fringe" depends on the location.

Thus, in the cross–correlation coefficient, the simple "white–noise" form and the independence of sample points is obtained only when the delay τ_d is reduced to zero ($\tau_d = 0$) by a suitable compensating operation, such as the "delay tracking" in the interferometry.

1.1.15 S/N Ratio of Correlator Output of Sampled Data

Let us now imagine a "semi–analog" correlator (non–existing in reality), which would multiply and integrate (i.e. time–average) sampled but not quantized (not clipped) data streams from two antennas of an interferometer. We will estimate here a signal–to–noise ratio of such a correlator, before examining actual digital correlators which deal with sampled **and** quantized data.

Let us assume that the two sampled data streams $x[i]$ and $y[i]$ are time samples of jointly stationary continuous–time random processes $x(t)$ and $y(t)$, which obey the second–order Gaussian probability distribution, as we assumed in the signal–to–noise–ratio discussion in Chapter 3. We further assume that $x(t)$ and $y(t)$ have identical rectangular passband spectra with bandwidth B, as given in equation (42), and they are sampled with the Nyquist interval $T = 1/(2B)$. Then, we have $x[i] = x(i\,T)$ and $y[i] = y(i\,T)$. Also, we assume that the delay tracking and the fringe stopping are perfectly performed beforehand, so that the two input data of exactly the same wave front are being correlated.

In this case, "correlator output" $\mathcal{R}_s$ of the sampled data streams is an

average of products of time samples over a certain number N:

$$\mathcal{R}_s = \frac{1}{N} \sum_{i=1}^{N} x[i]\, y[i]. \tag{54}$$

Expectation of this correlator output is nothing but the cross–correlation $R_{xy}[0]$ of $x[i]$ and $y[j]$ at zero argument, since

$$\langle \mathcal{R}_s \rangle = \frac{1}{N} \sum_{i=1}^{N} \langle x[i]\, y[i] \rangle = \frac{1}{N} \sum_{i=1}^{N} \langle x(i\,T)\, y(i\,T) \rangle = R_{xy}(0) = R_{xy}[0]. \tag{55}$$

On the other hand, dispersion of this correlator output σ_s^2 is given by

$$\sigma_s^2 = \langle \mathcal{R}_s^2 \rangle - \langle \mathcal{R}_s \rangle^2, \tag{56}$$

as we saw in Chapter 3. $\langle \mathcal{R}_s^2 \rangle$ is described through a double sum of the fourth statistical momentum in view of equation (54). The fourth statistical momentum is decomposed into a sum of products of second statistical momenta (correlations), as we discussed in Chapter 3, since $x[i] = x(i\,T)$ and $y[j] = y(j\,T)$ obey the joint Gaussian probability distribution. Thus, we have

$$
\begin{aligned}
\langle \mathcal{R}_s^2 \rangle &= \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \langle x[i]\, y[i]\, x[j]\, y[j] \rangle \\
&= \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \big\{\, \langle x[i]\, y[i] \rangle \langle x[j]\, y[j] \rangle \\
&\quad + \langle x[i]\, x[j] \rangle \langle y[i]\, y[j] \rangle \;+\; \langle x[i]\, y[j] \rangle \langle y[i]\, x[j] \rangle \,\big\} \\
&= \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \big\{\, R_{xy}^2[0] + R_{xx}[i-j]\, R_{yy}[i-j] + R_{xy}[i-j]\, R_{xy}[j-i] \,\big\} \\
&= \langle \mathcal{R}_s \rangle^2 + \frac{1}{N} R_{xx}[0]\, R_{yy}[0] + \frac{1}{N} R_{xy}^2[0] \\
&= \langle \mathcal{R}_s \rangle^2 + \frac{1}{N} R_{xx}[0]\, R_{yy}[0]\, (1 + \rho^2), \tag{57}
\end{aligned}
$$

where $\rho = \dfrac{R_{xy}[0]}{\sqrt{R_{xx}[0]\, R_{yy}[0]}} = \dfrac{R_{xy}(0)}{\sqrt{R_{xx}(0)\, R_{yy}(0)}}$ is the maximum cross–correlation coefficient, given in equation (52), in our assumed case with $\tau_d = 0$. In deriving last two lines of equation (57), we used the "white–noise" relations for autocorrelations:

$$R_{xx}[i-j] = R_{xx}[0]\, \delta_{ij}, \text{ and } R_{yy}[i-j] = R_{yy}[0]\, \delta_{ij}, \tag{58}$$

in view of equation (49), and for cross–correlation:

$$R_{xy}[i - j] = R_{xy}[0]\,\delta_{ij}, \tag{59}$$

which is also satisfied since we assumed $\tau_d = 0$.

Therefore, the dispersion of the correlator output in equation (56) is now given by

$$\sigma_s^2 = \frac{1}{N}\,R_{xx}[0]\,R_{yy}[0]\,(1 + \rho^2), \tag{60}$$

and we obtain the signal to noise ratio SNR:

$$SNR = \frac{\langle \mathcal{R}_s \rangle}{\sigma_s} = \frac{R_{xy}[0]}{\sqrt{R_{xx}[0]\,R_{yy}[0]\,(1 + \rho^2)}}\,\sqrt{N} = \frac{\rho}{\sqrt{1 + \rho^2}}\,\sqrt{N}. \tag{61}$$

In the expression of the maximum cross–correlation coefficient ρ, the auto-correlations $R_{xx}(0)$ and $R_{yy}(0)$ are usually dominated by system noise contributions from antenna–receiver systems of a radio interferometer, while the cross–correlation $R_{xy}(0)$ contains contribution of the signal from a radio source only, as we discussed in Chapter 3. Thus, when we observe a continuum spectrum source, ρ is approximately given by

$$\rho = \sqrt{\frac{T_{A_1}\,T_{A_2}}{T_{S_1}\,T_{S_2}}}, \tag{62}$$

as we saw in Chapter 3, where T_{A_1}, T_{A_2} are antenna temperatures, which are assumed constant throughout the frequency band B in the case of the continuum spectrum source, and T_{S_1}, T_{S_2} are system noise temperatures, of antenna 1 and antenna 2.

For most of radio sources, $T_A \ll T_S$, and, therefore, $\rho \ll 1$. In this case, equation (61) is reduced to

$$SNR = \frac{\langle \mathcal{R}_s \rangle}{\sigma_s} = \rho\,\sqrt{N} = \sqrt{\frac{T_{A_1}\,T_{A_2}}{T_{S_1}\,T_{S_2}}}\,\sqrt{N}. \tag{63}$$

If we denote an integration time of the correlation processing as τ_a, the number of samples N with Nyquist interval $1\,/\,(2\,B)$ is equal to

$$N = 2\,B\,\tau_a. \tag{64}$$

Therefore, equation (63) for the continuum spectrum source is reduced to

$$SNR = \sqrt{\frac{T_{A_1}\,T_{A_2}}{T_{S_1}\,T_{S_2}}}\,\sqrt{2\,B\,\tau_a}, \tag{65}$$

35

which is just identical with what we derived for correlator output of continuous–time voltages in Chapter 3.

This means that the Nyquist sampling does not cause any loss of signal–to–noise ratio of the correlator output, compared with the continuous–time case, as expected from the sampling theorem. This also means that there is no room for the oversampling, with a sampling rate higher than the Nyquist rate $(T < 1/2B)$, in improving the signal–to–noise ratio, despite increased number of data points. Thus, the Nyquist sampling is really optimum for the radio interferometry.

Note that equation (63) can be interpreted as showing $\sqrt{N}$–fold improvement of the signal–to–noise ratio after repeating and averaging N "measurements" of a power (product of two data streams, in our case). This means that measurements of a power made at the Nyquist interval are independent of each other, in the case of the rectangular passband spectra. This is a consequence of the independence of time samples themselves discussed earlier.

1.1.16 Nyquist Theorem and Nyquist Interval

The Nyquist theorem (Nyquist, 1928), which we saw in Chapter 2, says that thermal noise power W_ν per unit bandwidth emitted by a resistor in a thermal equilibrium with a temperature T is equal to

$$W_\nu = k\,T,\tag{66}$$

in the classical limit $h\nu \ll kT$, where k and h are the Boltzmann and rhe Planck constants, respectively. Therefore, energy E emitted within a rectangular band with a bandwidth B during a time interval t is

$$E = B\,t\,k\,T.\tag{67}$$

Since energy per one degree of freedom is equal to $\dfrac{1}{2}\,k\,T$ under the thermal equilibrium, number of degrees of freedom in this energy must be $N_F = 2\,B\,t$.

On the other hand, we have $N_I = 2\,B\,t$ Nyquist intervals during the time t, for the bandwidth B. In the case of the rectangular band, one Nyquist interval contains one independent sample, as we saw earlier. Therefore, we have N_I independent samples in the emitted energy during the time t.

The equality $N_I = N_F = 2\,B\,t$ means that one independent sample (Nyquist interval) in the information theory corresponds to one degree of freedom in the physics, in the thermal noise.

36

1.1.17 Higher–Order Sampling in VLBI Receiving Systems

In digital data processings as applied to radio astronomy, sampling of received voltage signals has been traditionally done at the basebands (or the video–bands), containing DC (zero frequency) as the lowest frequency, after frequency conversions. This was the safest way for reliable sampling, when clock rates of sampler circuits were not high enough, and not very stable.
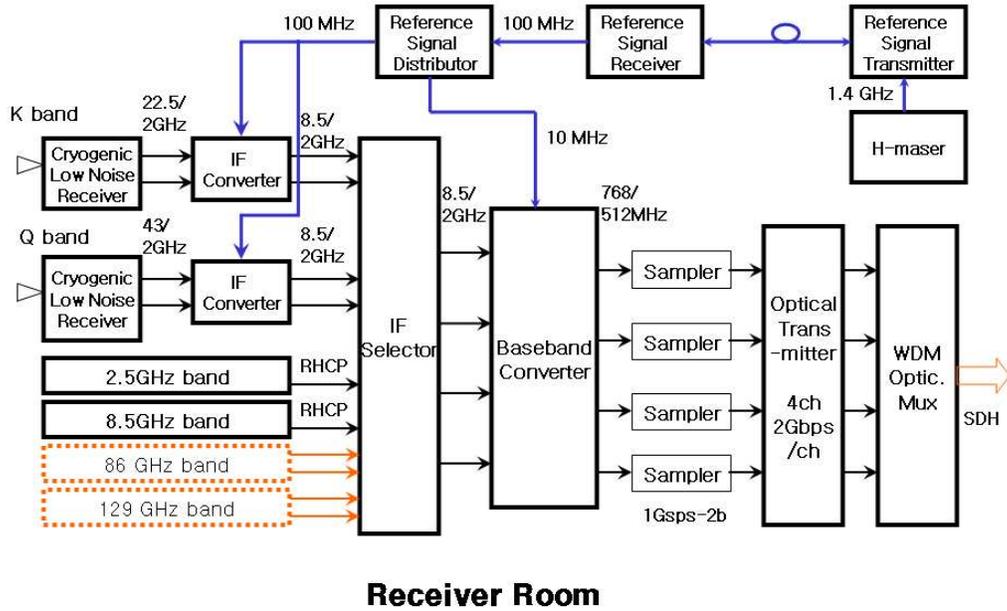
Figure 15: Diagram of receiving system in KVN (Korean VLBI Network) adopting the higher–order sampling technique (figure brought from KVN webpage *http://www.trao.re.kr/ kvn/*).

However, it is not easy, in existing analog filtering technology, to implement a good enough lowpass filter with sharp rectangular edges. This situation has often resulted in rather poor frequency characteristics of the baseband spectra, and made it difficult to achieve high signal–to–noise ratio, close to the one expected from an ideally rectangular spectrum. Also, because of this difficulty, high quality baseband converters tend to be expensive, especially when wide frequency bands are required.

Recently, Noriyuki Kawaguchi and his colleagues successfully applied so–called "higher–order sampling" technique to a number of VLBI systems, including Japanese VERA (VLBI Exploration of Radio Astrometry). The higher–order sampling is the sampling at a passband with $n > 0$, discussed earlier. In general, it is easier to design good analog bandpass filters, with nearly rectangular band shapes, when the ratio B/ν_0 of bandwidth B to central freuency ν_0 is smaller. Therefore, it is easier to make a nearly rectangular wideband filter for a passband, than for a baseband. In fact, the higher–order sampling technique has been effective in wideband receiving systems with typical bandwidth of 512 MHz or wider, for realizing better frequency characteristics and higher signal–to–noise–ratio (Iguchi and Kawaguchi, 2002).

Figure 15 shows a diagram of KVN (Korean VLBI Network) receiving system which adopts the higher–order sampling technique. A "baseband converter" cuts off a 512 MHz band from a 2 GHz–wide first IF signal with 8.5 GHz center frequency (i.e. 7.5 – 9.5 GHz band), and converts it to a 512 MHz–wide passband signal with 768 MHz central frequency (i.e. 512–1024 MHz band), which is then sampled by a high–speed sampler.

Note here that, for the bandwidth $B = 512$ MHz, the passband $nB \leq \nu < (n+1)B$ with $nB = 512$ MHz means an odd number of n (i.e. $n = 1$), where ν is the frequency and n is an integer. Therefore, spectrum of sampled signal is inverted with respect to the spectrum of original continuous–time signal, as we saw in Figure 8. In order to avoid possible inconveniences with the inverted spectra, LO (local oscillator) frequency of the "baseband converter" is chosen so that the passband spectrum is obtained in the lower sideband (LSB), i.e. the spectrum is inverted with respect to the first IF spectrum. In this way, one can obtain a spectrum of the sampled signal which is identical with the one contained in the first IF signal.

It seems worthwhile to mention an interesting question here. The sampling theorem says that the optimal sampling rate for a passband $nB \leq \nu < (n+1)B$ is $2B$. Does this mean that we can use an inexpensive low–speed 4 Msps (mega sample per second) sampler for sampling a signal in a high–frequency passband, say, 10.000 – 10.002 GHz? The answer is "NO", as N. Kawaguchi clearly explains. Although the required sampling interval is really $1/(2B)$, sampling **timings** must be controlled with much greater accuracy, better than $1/\nu_0$, where ν_0 is the central frequency of the passband (the "carrier" frequency). Otherwise, we will get all chaotically "jittered" data at each of VLBI stations, from which we will never find any good fringe. Therefore, a required sampler must be as accurate as, and as stable as, a 20 Gsps sampler, say. Consequently, the high–speed sampling technology is indispensable for successful application of the wideband higher–order sampling technique.

1.1.18 Clipping (or Quantization) of Analog Data

The sampling replaces data that are continuous in time with those discrete in time. This is undoubtedly a big step towards digitizing analog data. However, the sampling alone still leaves values of time samples analog, which may vary arbitrarily from sample to sample. For a complete analog to digital (A/D) conversion, we need to replace each continuously variable value with an element of a finite set of discrete values expressible by a certain number of bits. This step is called the "clipping" or the "quantization".

Number of discrete values expressible by a given number of bits determines "number of levels of quantization". Therefore, 1–bit, 2–bit, $\cdots$, n–bit quantizations usually correspond to 2–level, 4–level, $\cdots$, 2^n–level quantizations, respectively. There have been exceptions of the n–bit, 2^n–level law of quantization, such as 2–bit, 3–level quantization. However, such exceptional quantization schemes are rarely used in present–day VLBI systems.

We will denote henceforth a discrete–time process with quantized values as $\hat{x}[i]$. If number of quantization levels is m, with discrete values x_1, x_2, $\cdots$, x_m, then $\hat{x}[i]$ must take one of these m values as illustrated in Figure 16. A quantized process $\hat{x}[i]$ is supposed to be related in a prescribed way to an original discrete–time process $x[i]$ with analog values before clipping.
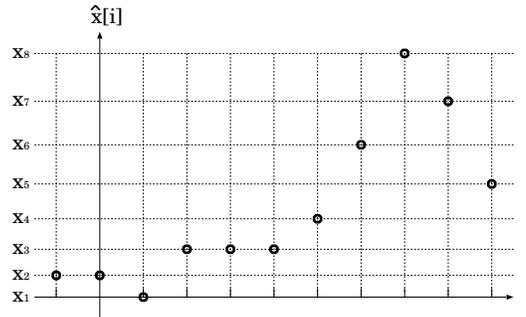
Figure 16: An image of a quantized discrete–time process.

The larger the number of quantization levels (i.e. bits), the more information can remain after the clipping. For reducing data size and increasing processing speed, however, smaller number of bits is preferable. Thus one has to choose an optimal number of quantization levels for one's particular purpose. In VLBI, 1–bit (2–level) and 2–bit (4–level) quantizations are mostly used.

Figure 17 and Table 1 show how quantized values are related to original analog values in cases of VLBI 1–bit (left panel)and 2–bit (right panel)

quantizations.

In 1–bit quantization scheme, clipped quantity takes only one of two values: $+1$ and -1, depending upon if original analog value is positive or negative, respectively. It is generally accepted that bit 0 is assigned to -1 state and bit 1 is assigned to $+1$ state.
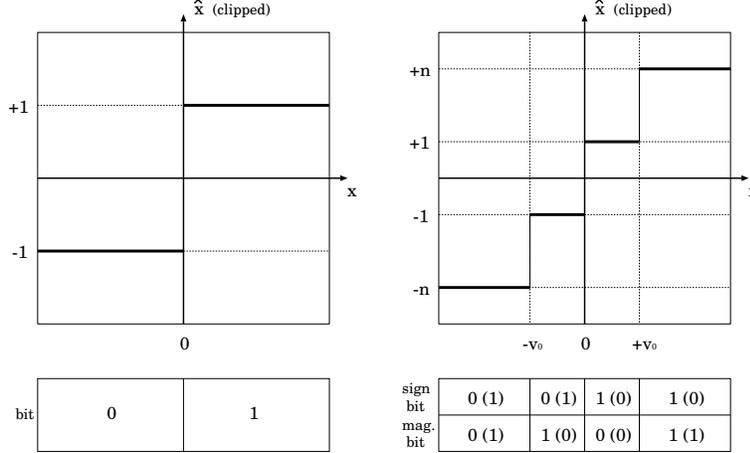
Figure 17: Relations between analog (unquantized) values x and clipped (quantized) values $\hat{x}$ for 1–bit (left) and 2–bit (right) quantizations, respectively. Bit assignments are shown in bottom panels. In case of 2–bit quantization (left), a representative bit assingment for data recorder is shown along with that for a correlator chip given in parentheses.

Quantization	1–bit (2–level)		2–bit (4–level)			
Analog value	$x < 0$	$0 \leq x$	$x < -v_0$	$-v_0 \leq x < 0$	$0 \leq x < v_0$	$v_0 \leq x$
Clipped value	$\hat{x} = -1$	$\hat{x} = +1$	$\hat{x} = -n$	$\hat{x} = -1$	$\hat{x} = +1$	$\hat{x} = +n$
Recorder bit ass.	0	1	s 0, m 0	s 0, m 1	s 1, m 0	s 1, m 1
Correlator bit ass.	0	1	s 1, m 1	s 1, m 0	s 0, m 0	s 0, m 1

Table 1: Clipping criteria and bit assingments of VLBI 1–bit and 2–bit quantizations. In the bit assignments for the 2–bit quantization scheme, "s" and "m" stand for sign bit and magnitude bit, respectively.

In 2–bit, 4–level quantization scheme, 4 quantization states are separated by three threshold values: $-v_0$, 0, and v_0. They correspond to 4 clipped values: $-n$, -1, $+1$, and $+n$, as shown in left panel of Figure 17 and Table 1. Values of parameters v_0 and n are chosen so that signal–to–noise ratio

of the correlator output is maximized. Note that n thus determined is not necessarily an integer.

Bit assignments to the 4 quantization states are not uniquely standardized yet in the 2–bit quantization scheme. Existing VLBI data recording systems mostly adopt 00, 01, 10, and 11 assignments for $-n$, -1, $+1$, and $+n$ states, where first and second bits stand for sign and magnitude bits, respectively, but a widely used 2–bit quantization correlator chip adopts 11, 10, 00, and 01 assignments, as shown in Figure 17 and Table 1. Therefore, recorded bits should be rearranged before the correlation, when we use a correlator with the chip.

Sign bits in the 2–bit quantized data are equivalent to 0 and 1 bits in the 1–bit quantized data. Therefore, it is usually not difficult to cross–correlate 1–bit quantized and 2–bit quantized data, which are obtained in different stations in the same VLBI observation with the same sampling rate, using a 1–bit correlation mode, if we sacrify some of information in the 2–bit quantized data. In this case, the only necessary thing is to pick up sign bits from the 2–bit qunatized data and cross–correlate them with the 1–bit quantized data. Of course, direct cross–correlation of data with different quantization schemes, such as 1–bit and 2–bit, can be performed without loosing information (see, for example, Hagen and Farley, 1973), provided that a special logical circuit for this purpose is built in a correlator.
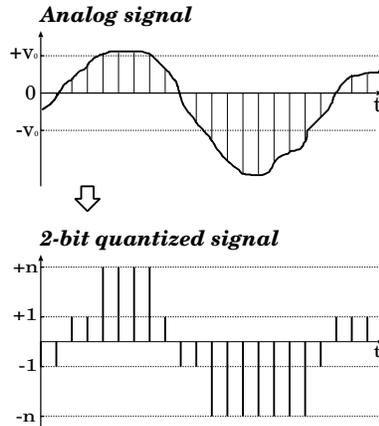
Figure 18: A schematic view of time variation of original analog data (top) and 2–bit quantized data (bottom).

In the 1–bit quantization scheme, only sign information of the original analog data remains in the clipped data, and no information on magnitude (amplitude) is left at all, as illustrated in Figure 3. In the 2–bit quantized

41

data, some information on the magnitude (amplitude) of the analog data is left, as illustrated in Figure 18, but the information is very vague.

Then, how much can we restore scientific information contained in the original analog data of VLBI after clipping them with 1–bit or 2–bit quantization scheme, which looks quite rough at least at the first glance?

The clipping theorem gives a surprising answer. The theorem was originally developed by J.H. van Vleck in a study of radar–jamming during the World War II conducted by the Radio Research Laboratory of Harvard University (Report No.51 on July 21, 1943), and was made public more than 20 years later by van Vleck and Middleton (1966).

1.1.19   Probability Distribution of Clipped Data

Before proceeding to the clipping theorem, we will examine how we can describe probability distribution of clipped data.

As we discussed earlier, signals from astronomical radio sources, as well as system noises produced in antenna–receiving systems and in environments, are well approximated by Gaussian random processes. Therefore, let us consider the data as jointly stationary continuous–time random processes $x(t)$ and $y(t)$, which obey the second–order Gaussian probability density, introduced in Chapter 3. Here we assume a zero–mean case (i.e. expectations of $x(t)$ and $y(t)$ are equal to zero), and use notations suited to the current discussions, slightly changing those adopted in Chapter 3. Also, we assume that both $x(t)$ and $y(t)$ are real processes.

Then, we describe the zero–mean second–order Gaussian probability density of jointly stationary continuous–time random processes $x(t+\tau)$ and $y(t)$ as:

$$f(x,\,y;\,\tau) \equiv f(x,\,y;\,t+\tau,\,t)$$

$$= \frac{1}{2\pi\sigma_x\sigma_y\,\sqrt{1-r_{xy}^2(\tau)}}\, e^{-\frac{1}{2[1-r_{xy}^2(\tau)]}\left[\frac{x^2}{\sigma_x^2} - 2\,r_{xy}(\tau)\,\frac{x\,y}{\sigma_x\,\sigma_y} + \frac{y^2}{\sigma_y^2}\right]}, \quad (68)$$

where we introduced notations: $\sigma_x^2 \equiv R_{xx}(0)$, $\sigma_y^2 \equiv R_{yy}(0)$, for dispersions of $x(t)$ and $y(t)$, and

$$r_{xy}(\tau) \equiv \frac{R_{xy}(\tau)}{\sqrt{R_{xx}(0)\,R_{yy}(0)}},$$

for cross–correlation coefficient of $x(t+\tau)$ and $y(t)$. Here, $R_{xy}(\tau)$, $R_{xx}(\tau)$, and $R_{yy}(\tau)$ are cross–correlation and autocorrelations of $x(t)$ and $y(t)$, as before. We first assume that the cross–correlation coefficient $r_{xy}(\tau)$ is smaller than

1 in absolute value (i.e. $r_{xy}^2(\tau) < 1$) in order to avoid possible singularity in our calculations which may ocur when $r_{xy}^2(\tau) = 1$.
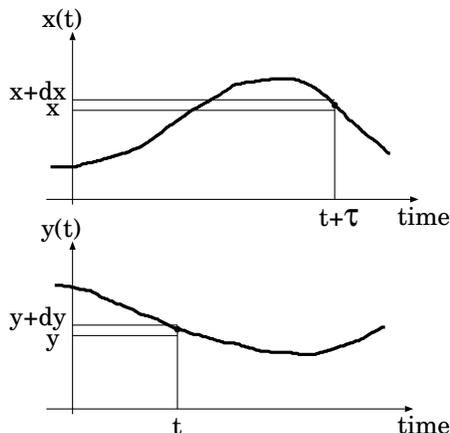
Figure 19: In case of continuous–time processes, Joint probability density describes probability of the processes to take values contained within infinitesimal ranges at certain periods of time.

Of course, this joint probability density satisfies the general definition of the cross–correlation:

$$R_{xy}(\tau) = \langle x(t+\tau)\, y(t) \rangle = \int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} x\, y\, f(x,\, y;\, \tau)\, dx\, dy, \qquad (69)$$

as we saw in Chapter 3 (note that in our zero–mean case the cross–covariance $C_{xy}(\tau)$ is just equal to the cross–correlation, i.e. $C_{xy}(\tau) = R_{xy}(\tau)$).

It is worth to recall here that the joint probability means

"$f(x,\, y;\, \tau)\, dx\, dy$ is a probability for $x(t+\tau)$ and $y(t)$ to be within ranges:

$$\begin{aligned} x &\leq\; x(t+\tau) \;< x + dx, \\ y &\leq\;\;\;\; y(t) \;\;\;\; < y + dy, \end{aligned}$$

for any t (see Figure 19). The condition "for any t" corresponds to our case of the stationary random processes.

Now if we consider discrete–time processes $x[i]$ and $y[i]$, which are time samples of the above jointly stationary real continuous–time random processes $x(t)$ and $y(t)$, i.e. $x[i] = x(iT)$ and $y[i] = y(iT)$, where T is a sampling

interval, the time samples $x[i]$ and $y[i]$ are also jointly stationary as we saw earlier, and their cross–correlation:

$$R_{xy}[m] = \langle x[n+m]\,y[n]\rangle = R_{xy}(mT) = \langle x(nT+mT)\,y(nT)\rangle$$

is described by the same joint probability density of the continuous–time processes $f(x,\,y;\,\tau)$ as

$$R_{xy}[m] = R_{xy}(mT) = \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} x\,y\,f(x,\,y;\,mT)\,dx\,dy. \qquad (70)$$

Let us then consider that we clip the time samples $x[i]$ and $y[i]$, and obtain clipped discrete–time processes which we denote as $\hat{x}[i]$ and $\hat{y}[i]$. They now take only discrete values of a finite number N (this means N–level quantization) $x_1,\,x_2,\,\cdots,\,x_N$ and $y_1,\,y_2,\,\cdots,\,y_N$.
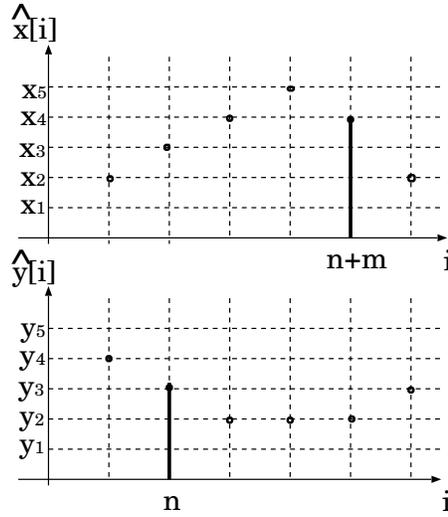
Figure 20: Joint probability $P(x_i, y_j; m)$ of clipped processes $\hat{x}[n+m]$ and $\hat{y}[n]$ describes probability for them to take certain discrete values x_i and y_j on quantization levels (here 5–level case is shown).

In this case, their cross–correlation $R_{\hat{x}\hat{y}}[m] = \langle \hat{x}[n+m]\,\hat{y}[n]\rangle$ is decribed by an equation:

$$R_{\hat{x}\hat{y}}[m] = \sum_{i=1}^{N}\sum_{j=1}^{N} x_i\,y_j P(x_i,\,y_j;\,m), \qquad (71)$$

where $P(x_i,\,y_j;\,m)$ is a joint probability for $\hat{x}[n+m]$ and $\hat{y}[n]$ to take values:

$$\hat{x}[n+m] = x_i,$$
$$\hat{y}[n] = y_j,$$

44

for any n (See Figure 20).

This cross–correlation is what should be yielded by a digital correlator. In fact, in view of the ergodicity, the cross–correlation is well approximated by a time–average of products of $\mathcal{N}$ quantized data $\hat{x}[n+m]$ and $\hat{y}[n]$:

$$R_{\hat{x}\hat{y}}[m] \cong \frac{1}{\mathcal{N}} \sum_{n=1}^{\mathcal{N}} (\hat{x}[m+n]\,\hat{y}[n]), \qquad (72)$$

if $\mathcal{N}$ is sufficiently large; and the digital correlator is nothing but a machine which time–averages a large number of products of digital (i.e. sampled and clipped) data.

At the first glance, equations (70) and (71) look quite different, and it appears difficult to relate $R_{\hat{x}\hat{y}}[m]$ with $R_{xy}[m]$. However, if the quantization condition is clearly specified, we can calculate $P(x_i,\,y_j;\,m)$ rather easily from $f(x,\,y;\,mT)$ (van Vleck and Middleton, 1966).

In case of the 1–bit, 2–level quantization, the condition is

$$\hat{x}[i] = \begin{cases} +1 & \text{for} \quad x[i] = x(iT) \geq 0, \\ -1 & \text{for} \quad x[i] = x(iT) < 0, \end{cases}$$

$$\hat{y}[i] = \begin{cases} +1 & \text{for} \quad y[i] = y(iT) \geq 0, \\ -1 & \text{for} \quad y[i] = y(iT) < 0. \end{cases} \qquad (73)$$

Therefore, the probability for $\hat{x}[i]$ to be +1, for example, is nothing but the probability for $x(iT)$ to be $0 \leq x(iT) < +\infty$. Thus we can describe joint probabilities for all combinations of quantization states through equations:

$$
\begin{aligned}
P(+1,\,+1;\,m) &= \int_0^{+\infty} \left\{ \int_0^{+\infty} f(x,\,y;\,mT)\,dx \right\} dy, \\
P(-1,\,-1;\,m) &= \int_{-\infty}^{0} \left\{ \int_{-\infty}^{0} f(x,\,y;\,mT)\,dx \right\} dy, \\
P(+1,\,-1;\,m) &= \int_{-\infty}^{0} \left\{ \int_0^{+\infty} f(x,\,y;\,mT)\,dx \right\} dy, \\
P(-1,\,+1;\,m) &= \int_0^{+\infty} \left\{ \int_{-\infty}^{0} f(x,\,y;\,mT)\,dx \right\} dy.
\end{aligned} \qquad (74)
$$

Integrals in the RHS of these equations are readily calculated, since $f(x,\,y;\,mT)$ is given by the joint Gaussian probability density in equation (68).

1.1.20 Cross–Correlation of 1–Bit Quantized Data: van Vleck Relationship

In case of the 1–bit quantization, we have

$$N = 2, \quad \text{and} \quad \begin{cases} x_1 = -1, \quad x_2 = +1, \\ y_1 = -1, \quad y_2 = +1. \end{cases} \tag{75}$$

Therefore, cross–correlation $R_{\hat{x}\hat{y}}$ of clipped data $\hat{x}[i]$ and $\hat{y}[i]$ is given by

$$
\begin{aligned}
R_{\hat{x}\hat{y}}[m] &= \sum_{i=1}^{2}\sum_{j=1}^{2} x_i\, y_j\, P(x_i,\, y_j;\, m) \\
&= (+1)\cdot(+1)\cdot P(+1,\, +1;\, m) + (-1)\cdot(-1)\cdot P(-1,\, -1;\, m) \\
&+ (+1)\cdot(-1)\cdot P(+1,\, -1;\, m) + (-1)\cdot(+1)\cdot P(-1,\, +1;\, m) \\
&= P(+1,\, +1;\, m) + P(-1,\, -1;\, m) \\
&- P(+1,\, -1;\, m) - P(-1,\, +1;\, m).
\end{aligned} \tag{76}
$$

Because of symmetric properties of the joint Gaussian probability density shown in equation (68), the joint probabilities given in equation (74) satisfy

$$
\begin{aligned}
P(+1,\, +1;\, m) &= P(-1,\, -1;\, m), \\
P(+1,\, -1;\, m) &= P(-1,\, +1;\, m).
\end{aligned} \tag{77}
$$

Also, by definition of the probability, sum of probabilities of all possible cases must be equal to 1, and hence

$$
\begin{aligned}
P(+1,\, +1;\, m) &+ P(-1,\, -1;\, m) + P(+1,\, -1;\, m) + P(-1,\, +1;\, m) \\
&= 2\,P(+1,\, +1;\, m) + 2\,P(+1,\, -1;\, m) = 1.
\end{aligned} \tag{78}
$$

Then, from equations (76), (77), and (78), we obtain

$$
\begin{aligned}
R_{\hat{x}\hat{y}}[m] &= P(+1,\, +1;\, m) + P(-1,\, -1;\, m) \\
&- P(+1,\, -1;\, m) - P(-1,\, +1;\, m) \\
&= 2\,P(+1,\, +1;\, m) - 2\,P(+1,\, -1;\, m) \\
&= 4\,P(+1,\, +1;\, m) - 1.
\end{aligned} \tag{79}
$$

Substituting the explicit form of the joint Gaussian probability density in equation (68) into equation (74), we calculte $4\,P(+1,\, +1;\, m)$:

$$4\,P(+1,\, +1;\, m) = 4 \int_{0}^{+\infty} \left\{ \int_{0}^{+\infty} f(x,\, y;\, mT)\, dx \right\} dy$$

$$
= \int\limits_{0}^{+\infty} \left\{ \int\limits_{0}^{+\infty} \frac{2}{\pi \sigma_x \sigma_y \sqrt{1 - r_{xy}^2}} \, e^{ -\dfrac{1}{2(1 - r_{xy}^2)} \left(\dfrac{x^2}{\sigma_x^2} - 2\, r_{xy} \dfrac{x\,y}{\sigma_x \, \sigma_y} + \dfrac{y^2}{\sigma_y^2} \right) } dx \right\} dy,
$$

$$(80)$$

where T is the sampling interval, as before, and we denoted the cross–correlation coefficient $r_{xy}[m] = r_{xy}(mT)$ as r_{xy}, for simplicity.

Let us introduce new variables ζ and ϕ, which satisfy

$$
\begin{aligned}
x &= \sigma_x \, \zeta \, \cos \phi, \\
y &= \sigma_y \, \zeta \, \sin \phi.
\end{aligned}
$$

$$(81)$$

Then the above integral is reduced to

$$
\begin{aligned}
4\, P(+1,\, +1;\, m) &= \frac{2}{\pi \sqrt{1 - r_{xy}^2}} \int\limits_{0}^{\frac{\pi}{2}} d\phi \int\limits_{0}^{\infty} e^{ -\dfrac{\zeta^2 (1 - r_{xy} \sin 2\phi)}{2(1 - r_{xy}^2)} } \zeta \, d\zeta \\
&= \frac{2}{\pi} \int\limits_{0}^{\frac{\pi}{2}} \frac{\sqrt{1 - r_{xy}^2}}{1 - r_{xy} \sin 2\phi} \, d\phi.
\end{aligned}
$$

$$(82)$$

If we further introduce another new variable θ, which satisfies

$$
\tan \theta = \frac{\tan \phi - r_{xy}}{\sqrt{1 - r_{xy}^2}}, \quad \text{and, therefore,} \quad \theta = \arctan \left(\frac{\tan \phi - r_{xy}}{\sqrt{1 - r_{xy}^2}} \right), \quad (83)
$$

then we have

$$
\frac{d\theta}{d\phi} = \frac{1}{1 + \left(\dfrac{\tan \phi - r_{xy}}{\sqrt{1 - r_{xy}^2}} \right)^2} \frac{1}{\sqrt{1 - r_{xy}^2}} \frac{1}{\cos^2 \phi} = \frac{\sqrt{1 - r_{xy}^2}}{1 - r_{xy} \sin 2\phi}. \quad (84)
$$

Note that this has the same form as the integrand of equation (82). The limits of the integration $\phi = 0$ and $\phi = \dfrac{\pi}{2}$ now correspond to

$\theta = \theta_0 \equiv -\arctan \left(\dfrac{r_{xy}}{\sqrt{1 - r_{xy}^2}} \right)$ and $\theta = \dfrac{\pi}{2}$, respectively. Terefore, equation (82) becomes

$$
4\, P(+1,\, +1;\, m) = \frac{2}{\pi} \int\limits_{\theta_0}^{\frac{\pi}{2}} d\theta = 1 - \frac{2}{\pi} \theta_0 = 1 + \frac{2}{\pi} \arctan \left(\frac{r_{xy}}{\sqrt{1 - r_{xy}^2}} \right). \quad (85)
$$

Denoting the cross–correlation coefficient r_{xy} through a sine function:

$$r_{xy} = \sin\xi, \quad \text{and, therefore,} \quad \frac{r_{xy}}{\sqrt{1 - r_{xy}^2}} = \tan\xi,$$

we obtain

$$4\,P(+1,\,+1;\,m) = 1 + \frac{2}{\pi}\arctan(\tan\xi) = 1 + \frac{2}{\pi}\xi = 1 + \frac{2}{\pi}\arcsin(r_{xy}). \quad (86)$$

We must specify here a range of $\arcsin(r_{xy})$, since, in general, arcsine is a multi–value function, while the probability $P(+1,\,+1;\,m)$ is certainly not. In view of the general property of the probability, $P(+1,\,+1,\,m)$ must be confined within a range:

$$0 \le P(+1,\,+1,\,m) \le \frac{1}{2}.$$

Indeed, the upper limit corresponds to a case of the complete correlation (identical data), for which

$$P(+1,\,+1;\,m) = P(-1,\,-1;\,m) = \frac{1}{2},$$

because

$$P(+1,\,-1;\,m) = P(-1,\,+1;\,m) = 0,$$

while the lower limit corresponds to a case of the complete anti–correlation (identical data but with different signs), for which

$$P(+1,\,+1;\,m) = P(-1,\,-1;\,m) = 0.$$

Since the cross–correlation coefficient r_{xy} of the original unclipped data must be 1 in the complete correlation, and -1 in the complete anti–correlation, the arcsine function in equation (86) must be confined within a range:

$$-\frac{\pi}{2} \le \arcsin(r_{xy}) \le \frac{\pi}{2}. \quad (87)$$

Finally, from equation (79), we obtain

$$\begin{aligned}
R_{\hat{x}\hat{y}}[m] &= 4P(+1,\,+1;\,m) - 1 = \frac{2}{\pi}\arcsin(r_{xy}[m]) \\
&= \frac{2}{\pi}\arcsin\left(\frac{R_{xy}(mT)}{\sqrt{R_{xx}(0)\,R_{yy}(0)}}\right).
\end{aligned} \quad (88)$$

Since $\hat{x}[i]\,\hat{x}[i] = 1$ and $\hat{y}[i]\,\hat{y}[i] = 1$ for any i, and therefore $R_{\hat{x}\hat{x}}[0] = 1$ and $R_{\hat{y}\hat{y}}[0] = 1$, for the 1–bit quantized data, their cross–correlation coefficient is equal to their cross–correlation:

$$r_{\hat{x}\hat{y}}[m] = \frac{R_{\hat{x}\hat{y}}[m]}{\sqrt{R_{\hat{x}\hat{x}}[0]\,R_{\hat{y}\hat{y}}[0]}} = R_{\hat{x}\hat{y}}[m]. \tag{89}$$

Therefore, equation (88) is described also as

$$r_{\hat{x}\hat{y}}[m] = \frac{2}{\pi}\,\arcsin(r_{xy}[m]). \tag{90}$$

In a particular case of a small cross–correlation coefficient $\mid r_{xy}[m] \mid \ll 1$, which is usually the case in radio interferometry, we have an approximate linear equation:

$$r_{\hat{x}\hat{y}}[m] = \frac{2}{\pi}\,r_{xy}[m]. \tag{91}$$

Although we derived equations (88) and (90) assuming that $r_{xy}^2 < 1$, it is worth to confirm here that the resultant equations are valid in the limiting cases of the complete correlation ($r_{xy} = 1$ and $r_{\hat{x}\hat{y}} = 1$) and the complete anti–correlation ($r_{xy} = -1$ and $r_{\hat{x}\hat{y}} = -1$), too.

Equation (88) or (90) is generally called the "van Vleck relationship". This is indeed a surprising result which shows that the cross–correlation coefficient $r_{xy}[m]$ of the original analog data is almost completely restored from the cross–correlation coefficient $r_{\hat{x}\hat{y}}[m]$ of the 1–bit quantized data by a simple equation:

$$r_{xy}[m] = \sin\left(\frac{\pi}{2}\,r_{\hat{x}\hat{y}}[m]\right). \tag{92}$$

In the case of the small cross–correlation coefficient $\mid r_{xy}[m] \mid \ll 1$, we have

$$r_{xy}[m] = \frac{\pi}{2}\,r_{\hat{x}\hat{y}}[m]. \tag{93}$$

As we saw before, cross–correlation $R_{\hat{x}\hat{y}}[m]\,(= r_{\hat{x}\hat{y}}[m])$ of digital data is readily obtained from a digital correlator. Therefore, equations (92) and (93) mean that, **we can completely derive functional form and, therefore, spectral shape of the original cross–correlation coefficient** from an output $R_{\hat{x}\hat{y}}[m]$ of a digital correlator of the 1–bit quantized data, though the amplitude is reduced by a factor of $\cong 2/\pi$ (see, for example, Fifure 21).

It is natural that the cross–correlation $R_{xy}(mT)$ itself of the original analog data cannot be directly obtained from the 1–bit quantized data alone which carries sign information only. Nevertheless, we can restore the cross–correlation $R_{xy}(mT)$ of the original analog data, if their dispersions $R_{xx}(0)$
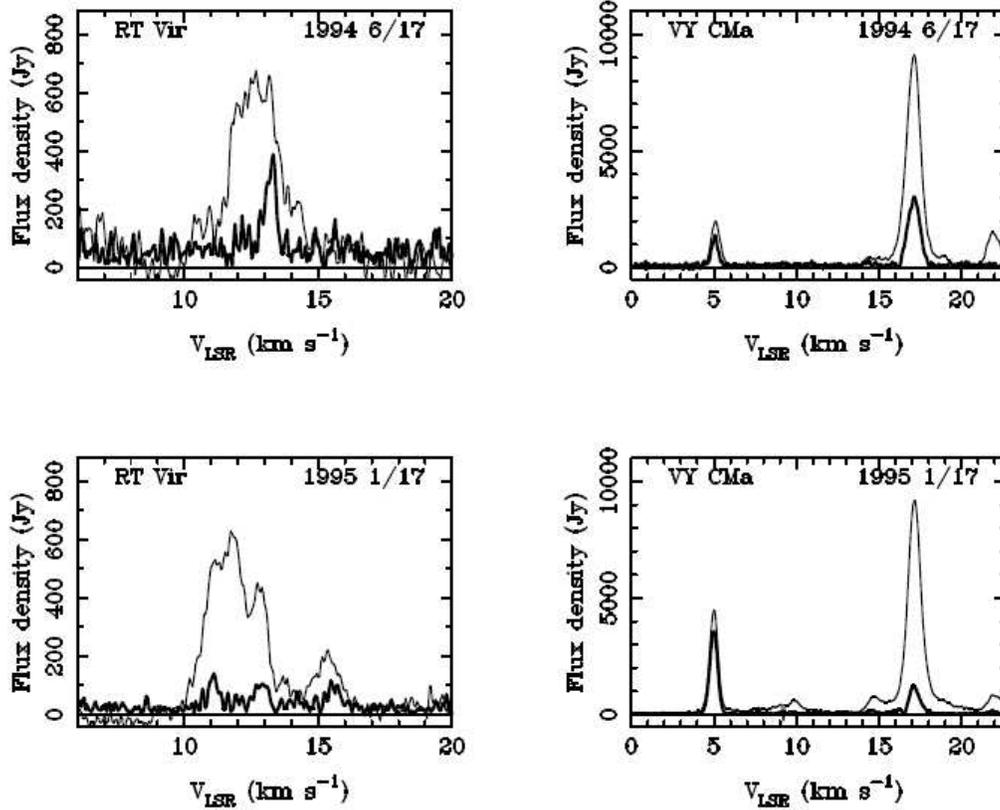
Figure 21: Spectra of water maser lines derived from 1–bit quantized data. Right and left panels show water maser lines of Asymptotic Giant Branch (AGB) stars RT Vir and VY CMa, respectively, obtained at two epochs (top and bottom panels). Thin lines show total–power (i.e. autocorrelation) spectra obtained with Mizusawa 10m antenna. Thick lines show VLBI cross-power spectra obtained with Mizusawa 10m - Kagoshima 6m baseline. The cross–power spectra show lower flux densities compared with the total–power spectra because maser features are slightly extended and hence partially resolved in the VLBI baseline. (Figure courtesy of Imai et al., *Astron. Astrophys*, **317**, L67-L70, 1997.)

and $R_{yy}(0)$ are known, for example, from suitable system–noise measurements.

Derivation of this relationship owes to the clever use of the probability relations in equation (74) by van Vleck. Note that J.H. van Vleck is a scientist who in 1977 received Nobel Prize in Physics in his major field of research: the "fundamental theoretical investigations of the electronic structure of magnetic and distorted systems".

1.1.21 van Vleck Relationship in Autocorrelation

Although we examined the derivation of the van Vleck relationship for cross–correlation, there is no restriction in applying the same logic to autocorrelation of a zero–mean stationary random process $x(t)$. As a matter of fact, van Vleck originally derived his relationship for the autocorrelation. The only difference is that we have to use, instead of equation (68), second–order Gaussian probability density for values of $x(t)$ taken at times $t + \tau$ and t:

$$f(x_1,\, x_2;\, \tau) = \frac{1}{2\pi\sigma_x^2\sqrt{1 - r_{xx}^2(\tau)}}\, e^{-\frac{1}{2(1 - r_{xx}^2(\tau))}\left(\frac{x_1^2 - 2r_{xx}(\tau)\,x_1 x_2 + x_2^2}{\sigma_x^2}\right)},$$

(94)

where $\sigma_x^2 = R_{xx}(0)$ is the dispersion and $r_{xx}(\tau) = \dfrac{R_{xx}(\tau)}{R_{xx}(0)}$ is the correlation coefficient of $x(t)$.

Exactly the same logic as in the cross–correlation case leads to an equation:

$$r_{\hat{x}\hat{x}}[m] = R_{\hat{x}\hat{x}}[m] = \frac{2}{\pi}\, \arcsin(r_{xx}[m]) = \frac{2}{\pi}\, \arcsin\left(\frac{R_{xx}(mT)}{R_{xx}(0)}\right),$$

(95)

for the correlation coefficient of the 1–bit quantized data $\hat{x}[i]$, which is the van Vleck relationship in the autocorrelation.

1.1.22 Spectra of Clipped Data

The derivation of the van Vleck relationship, as we saw above, is quite general and applicable to any data obeying joint Gaussian probability density. In particular, the relationship is well valid for continuous–time data, though we have dealt with only sampled data in our discussion of digitization of analog data. Thus, we generally have

$$R_{\hat{x}\hat{y}}(\tau) = \frac{2}{\pi}\, \arcsin(r_{xy}(\tau)),$$

(96)

51

for the cross–correlation, and

$$R_{\hat{x}\hat{x}}(\tau) = \frac{2}{\pi} \arcsin(r_{xx}(\tau)), \quad \text{and} \quad R_{\hat{y}\hat{y}}(\tau) = \frac{2}{\pi} \arcsin(r_{yy}(\tau)), \qquad (97)$$

for the autocorrelations, with arbitrary time interval τ.

Therefore, we can calculate cross–power spectrum $S_{\hat{x}\hat{y}}(\omega)$, and power spectra $S_{\hat{x}\hat{x}}(\omega)$ and $S_{\hat{y}\hat{y}}(\omega)$ of clipped data in terms of the ordinary Fourier transformations:

$$S_{\hat{x}\hat{y}}(\omega) = \int_{-\infty}^{\infty} R_{\hat{x}\hat{y}}(\tau)\, e^{-i\omega\tau}\, d\tau \;=\; \frac{2}{\pi} \int_{-\infty}^{\infty} \arcsin(r_{xy}(\tau))\, e^{-i\omega\tau}\, d\tau,$$

$$S_{\hat{x}\hat{x}}(\omega) = \int_{-\infty}^{\infty} R_{\hat{x}\hat{x}}(\tau)\, e^{-i\omega\tau}\, d\tau \;=\; \frac{2}{\pi} \int_{-\infty}^{\infty} \arcsin(r_{xx}(\tau))\, e^{-i\omega\tau}\, d\tau,$$

$$S_{\hat{y}\hat{y}}(\omega) = \int_{-\infty}^{\infty} R_{\hat{y}\hat{y}}(\tau)\, e^{-i\omega\tau}\, d\tau \;=\; \frac{2}{\pi} \int_{-\infty}^{\infty} \arcsin(r_{yy}(\tau))\, e^{-i\omega\tau}\, d\tau. \qquad (98)$$

Let us consider a case when analog data have rectangular spectra. This is an important case from a practical point of view, since rectangular bandpass filters are widely adopted in radio astronomy. We saw earlier that correlation coefficients of analog data take sinc function forms when their spectra are rectangular. Then, how spectra of the data will look like after clipping?

Figure 22 shows a power spectrum $S_{\hat{x}\hat{x}}(\omega)$ of 1–bit quantized data (solid line in bottom panel) which is derived from analog data with a rectangular power spectrum $S_{xx}(\omega)$ of bandwidth B (broken line in bottom panel). The power spectrum of the analog data $S_{xx}(\omega)$ is normalized by the dispersion $R_{xx}(0)$ here. In other words, the "power spectrum" in this Figure is the Fourier transform of the correlation coefficient $r_{xx}(\tau)$. Therefore, areas under the spectra (i.e. powers) of the analog and clipped data are equal. Also shown are (1) correlation coefficient of the analog data $r_{xx}(\tau)$ having a sinc function form (upper left panel), and (2) correlation coefficient of the clipped data $r_{\hat{x}\hat{x}}(\tau) = R_{\hat{x}\hat{x}}(\tau) = \frac{2}{\pi} \arcsin(r_{xy}(\tau))$ (upper right panel).

Although the bulk of the spectrum after clipping still remains nearly band–limited, peak amplitude is somewhat reduced and a low level–skirt appears outside the original band which spreads over a wide range of frequency. Thus, the original Nyquist interval (1 / (2B)) is no longer strictly optimum for sampling the clipped data.

On the other hand, Figure 23 shows a cross–power spectrum $S_{\hat{x}\hat{y}}(\omega)$ of 1–bit quantized data $\hat{x}$ and $\hat{y}$ (solid line in bottom panel) which are derived from
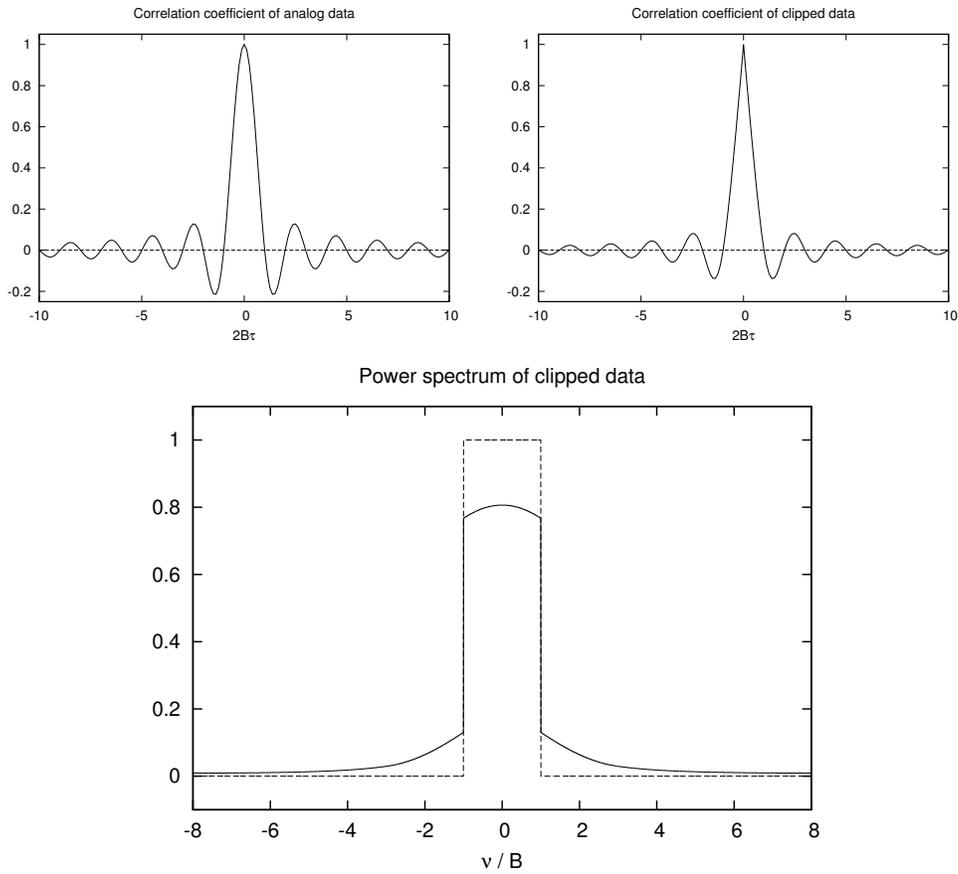
Figure 22: Original rectangular power spectrum $S_{xx}(\omega)$ with bandwidth B (broken line in bottom panel) of analog data, and power spectrum $S_{\hat{x}\hat{x}}(\omega)$ (solid line in bottom panel) of clipped data which is derived from the analog data by means of the 1–bit quantization. Horizontal axis of the bottom panel shows frequency ν normalized by the bandwidth B. Upper panels show correlation coefficient $r_{xx}(\tau)$ of the original analog data having a sinc function form (left), and correlation coefficient $r_{\hat{x}\hat{x}}(\tau)$ of the clipped data (right). Horizontal axes of the upper panels show delay τ normalized by the Nyquist interval $(1 / (2\,B))$ of the original analog data.
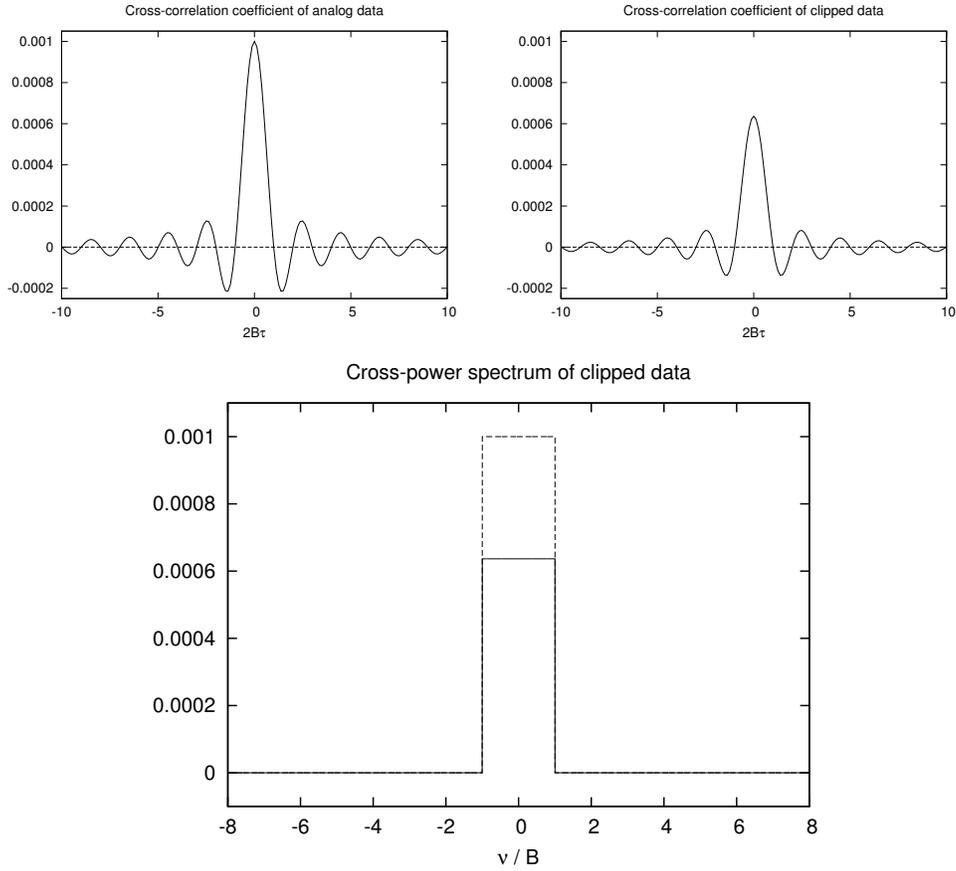
Figure 23: Original rectangular cross–power spectrum $S_{xy}(\omega)$ with bandwidth B (broken line in bottom panel) of analog data, and cross–power spectrum $S_{\hat{x}\hat{y}}(\omega)$ (solid line in bottom panel) of clipped data which are derived from the analog data by means of the 1–bit quantization. Horizontal axis of the bottom panel shows frequency ν normalized by the bandwidth B. Upper panels show cross–correlation coefficient $r_{xy}(\tau)$ of the original analog data having a sinc function form (left), and cross–correlation coefficient $r_{\hat{x}\hat{y}}(\tau)$ of the clipped data (right). The maximum cross–correlation coefficient of the analog data is assumed to be 0.001. Horizontal axes of the upper panels show delay τ normalized by the Nyquist interval $(1 \,/\, (2\,B))$ of the original analog data.

analog data x and y having a real rectangular cross–power spectrum $S_{xy}(\omega)$ of bandwidth B (broken line in bottom panel). The cross–power spectrum of the analog data $S_{xy}(\omega)$ in this Figure is normalized by the geometric mean of the dispersions $\sqrt{R_{xx}(0)\,R_{yy}(0)}$, and, therefore, is the Fourier transform of the cross–correlation coefficient $r_{xy}(\tau)$. We assume here that the maximum cross–correlation coefficient of the analog data is small, and, for definiteness, set it to be 0.001. Upper panels of the Figure show (1) cross–correlation coefficient of the analog data $r_{xy}(\tau)$ having a sinc function form (left), and (2) cross–correlation coefficient of the clipped data $r_{\hat{x}\hat{y}}(\tau) = R_{\hat{x}\hat{y}}(\tau) \cong \dfrac{2}{\pi}\, r_{xy}(\tau)$ (right).

Unlike in the power spectrum case, the spectrum after clipping remains rectangular, though peak amplitude is reduced by a factor of $2/\pi \cong 0.6366$ compared with the original analog one, in this cross–power spectrum case. This is because the cross–correlation coefficient of the clipped data is just proportional to the cross–correlation coefficient of the original analog data with a proportionality coefficient of $2/\pi$, in the approximation of the small cross–correlation coefficient $\mid r_{xy}(\tau)\mid \ll 1$, which we assumed here.

1.1.23 Price's Theorem

Now we proceed to the 2–bit quantization case. Although we can derive the cross–correlation of the 2–bit quantized data by evaluating probabilities of quantization states, as we learned in the 1–bit quantization case, use of Price's theorem (Price, 1958) offers a simpler solution.

The theorem states the following.

Suppose we have zero–mean stationary random processes $x(t)$ and $y(t)$ which satisfy joint Gaussian probability density:

$$f(x,\, y) \equiv f(x,\, y;\, \tau) = \frac{1}{2\pi\sigma_x\sigma_y\,\sqrt{1-r^2}}\, e^{-\dfrac{1}{2(1-r^2)}\left(\dfrac{x^2}{\sigma_x^2} - 2\,r\,\dfrac{x\,y}{\sigma_x\,\sigma_y} + \dfrac{y^2}{\sigma_y^2}\right)},$$

where $\sigma_x^2 = R_{xx}(0)$ and $\sigma_y^2 = R_{yy}(0)$ are dispersions and $r \equiv r_{xy}(\tau) = \dfrac{R_{xy}(\tau)}{\sigma_x\,\sigma_y}$ is a cross–correlation coefficient of the processes $x(t)$ and $y(t)$.

Then, *for an arbitrary function $g(x,\, y)$ with expectation*

$$\langle g(x,\, y)\rangle = \int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} g(x,\, y) f(x,\, y)\, dx\, dy, \tag{99}$$

55

we have

$$\frac{1}{\sigma_x^n \, \sigma_y^n} \frac{\partial^n \langle g(x,\,y) \rangle}{\partial \, r^n} = \left\langle \frac{\partial^{2n} g(x,\,y)}{\partial x^n \, \partial y^n} \right\rangle, \qquad (100)$$

(Price, 1958).

Auxiliary Formula

In order to prove Price's theorem, we first derive an auxiliary formula (Papoulis, 1984):

$$\frac{1}{\sigma_x^n \, \sigma_y^n} \frac{\partial^n f(x,\,y)}{\partial r^n} = \frac{\partial^{2n} f(x,\,y)}{\partial x^n \, \partial y^n}, \quad \text{or} \quad \frac{\partial^n f(x',\,y')}{\partial r^n} = \frac{\partial^{2n} f(x',\,y')}{\partial x'^n \, \partial y'^n}, \quad (101)$$

where

$$x' = \frac{x}{\sigma_x}, \quad y' = \frac{y}{\sigma_y},$$

$$f(x',\,y') = \frac{1}{2\pi \, \sigma_x \, \sigma_y \, \sqrt{1-r^2}} \, e^{-\dfrac{x'^2 - 2\,r\,x'\,y' + y'^2}{2(1-r^2)}}. \qquad (102)$$

We verify this formula using the method of mathematical induction.

- For $n = 1$, simple differentiations show

$$\frac{\partial f(x',\,y')}{\partial r} = \frac{r\,(1-r^2) + (r\,x' - y')\,(r\,y' - x')}{2\pi\sigma_x\,\sigma_y\,(1-r^2)^{5/2}} e^{-\dfrac{x'^2 - 2\,r\,x'\,y' + y'^2}{2(1-r^2)}},$$

and

$$\frac{\partial^2 f(x',\,y')}{\partial x' \, \partial y'} = \frac{r\,(1-r^2) + (r\,x' - y')\,(r\,y' - x')}{2\pi \, \sigma_x \, \sigma_y \, (1-r^2)^{5/2}} e^{-\dfrac{x'^2 - 2\,r\,x'\,y' + y'^2}{2(1-r^2)}}.$$

Therefore, $\dfrac{\partial f(x',\,y')}{\partial r} = \dfrac{\partial^2 f(x',\,y')}{\partial x' \, \partial y'}$, i.e. the formula is valid for $n = 1$.

- Now, if the formula is valid for $n = m$:

$$\frac{\partial^m f(x',\,y')}{\partial r^m} = \frac{\partial^{2m} f(x',\,y')}{\partial x'^m \, \partial y'^m},$$

56

then, for $n = m + 1$, we have

$$\frac{\partial^{m+1} f(x', y')}{\partial r^{m+1}} = \frac{\partial}{\partial r} \frac{\partial^m f(x', y')}{\partial r^m} = \frac{\partial}{\partial r} \frac{\partial^{2m} f(x', y')}{\partial x'^m \partial y'^m}$$

$$= \frac{\partial^{2m}}{\partial x'^m \partial y'^m} \frac{\partial f(x', y')}{\partial r} = \frac{\partial^{2m}}{\partial x'^m \partial y'^m} \frac{\partial^2 f(x', y')}{\partial x' \partial y'} = \frac{\partial^{2(m+1)} f(x', y')}{\partial x'^{m+1} \partial y'^{m+1}},$$

i.e., the formula is also valid for $n = m + 1$.

Since the formula is valid for $n = 1$, this means that the formula is valid for arbitrary $n \geq 1$.

Thus, we confirmed the auxiliary formula.

Proof of Price's Theorem

We now prove Price's theorem, using again the method of mathematical induction.

- For $n = 1$, we have

$$\frac{1}{\sigma_x \sigma_y} \frac{\partial \langle g(x, y) \rangle}{\partial r} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g \frac{1}{\sigma_x \sigma_y} \frac{\partial f}{\partial r} dx \, dy = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g \frac{\partial^2 f}{\partial x \, \partial y} dx \, dy$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left[\frac{\partial}{\partial x} \left(g \frac{\partial f}{\partial y} \right) - \frac{\partial}{\partial y} \left(\frac{\partial g}{\partial x} f \right) + \frac{\partial^2 g}{\partial x \, \partial y} f \right] dx \, dy$$

$$= \int_{-\infty}^{\infty} \left(g \frac{\partial f}{\partial y} \right) \bigg|_{x=-\infty}^{x=+\infty} dy - \int_{-\infty}^{\infty} \left(\frac{\partial g}{\partial x} f \right) \bigg|_{y=-\infty}^{y=+\infty} dx + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{\partial^2 g}{\partial x \, \partial y} f \, dx \, dy,$$

where we used the auxiliary formula:

$$\frac{1}{\sigma_x \sigma_y} \frac{\partial f(x, y)}{\partial r} = \frac{\partial^2 f(x, y)}{\partial x \, \partial y}.$$

First two terms in the RHS vanish as long as $g(x, y) e^{-(x^2+y^2)}$ converges to zero at $x = \pm\infty$ and $y = \pm\infty$. Thus,

$$\frac{1}{\sigma_x \sigma_y} \frac{\partial \langle g(x, y) \rangle}{\partial r} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{\partial^2 g(x, y)}{\partial x \, \partial y} f(x, y) \, dx \, dy = \left\langle \frac{\partial^2 g(x, y)}{\partial x \, \partial y} \right\rangle,$$

i.e. the theorem is valid for $n = 1$.

57

- Now, if the theorem is valid for $n = m$:

$$\frac{1}{\sigma_x^m \, \sigma_y^m} \frac{\partial^m \langle g(x, \, y) \rangle}{\partial \, r^m} = \left\langle \frac{\partial^{2m} g(x, \, y)}{\partial x^m \, \partial y^m} \right\rangle,$$

then, for $n = m + 1$, we have

$$\frac{1}{\sigma_x^{m+1} \, \sigma_y^{m+1}} \frac{\partial^{m+1} \langle g(x, \, y) \rangle}{\partial r^{m+1}} = \frac{1}{\sigma_x \, \sigma_y} \frac{\partial}{\partial r} \left(\frac{1}{\sigma_x^m \, \sigma_y^m} \frac{\partial^m \langle g(x, \, y) \rangle}{\partial \, r^m} \right)$$

$$= \frac{1}{\sigma_x \, \sigma_y} \frac{\partial}{\partial r} \left\langle \frac{\partial^{2m} g(x, \, y)}{\partial x^m \, \partial y^m} \right\rangle = \int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} \frac{\partial^{2m} g(x, \, y)}{\partial x^m \, \partial y^m} \frac{1}{\sigma_x \, \sigma_y} \frac{\partial f(x, \, y)}{\partial r} \, dx \, dy$$

$$= \int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} \frac{\partial^{2m} g(x, \, y)}{\partial x^m \, \partial y^m} \frac{\partial^2 f(x, \, y)}{\partial x \, \partial y} \, dx \, dy.$$

Integrating by parts again, we see that the last term is equal to

$$\int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} \frac{\partial^{2(m+1)} g(x, \, y)}{\partial x^{m+1} \, \partial y^{m+1}} f(x, \, y) \, dx \, dy = \left\langle \frac{\partial^{2(m+1)} g(x, \, y)}{\partial x^{m+1} \, \partial y^{m+1}} \right\rangle,$$

as long as $g(x, \, y) \, e^{-(x^2+y^2)}$ converges to zero at $x = \pm\infty$ and $y = \pm\infty$. Therefore, the theorem is valid for $n = m + 1$, and, hence, for any $n \geq 1$.

Thus we proved Price's Theorem.

Price's Theorem in the 1–Bit Quantaization Case

Let us experience usefulness of Price's theorem in the 1–bit quantization case as an example. If we choose $g(x, \, y) = \hat{x} \, \hat{y}$, then we have

$$\langle g(x, \, y) \rangle = R_{\hat{x}\hat{y}} = \langle \hat{x} \, \hat{y} \rangle,$$

with

$$\hat{x}(x) = \begin{cases} +1 : & x \geq 0 \\ -1 : & x < 0, \end{cases} \qquad \hat{y}(y) = \begin{cases} +1 : & y \geq 0 \\ -1 : & y < 0, \end{cases}$$

in the 1–bit quantization case. We again assume that x and y obey the joint Gaussian probability density:

$$f(x, \, y) = \frac{1}{2\pi\sigma_x\sigma_y \sqrt{1 - r^2}} \, e^{-\frac{1}{2(1 - r^2)} \left(\frac{x^2}{\sigma_x^2} - 2\, r \, \frac{x \, y}{\sigma_x \, \sigma_y} + \frac{y^2}{\sigma_y^2} \right)}.$$

Now, Price's theorem says

$$\frac{1}{\sigma_x \, \sigma_y} \, \frac{\partial R_{\hat{x}\hat{y}}}{\partial r} = \left\langle \frac{\partial^2 (\hat{x} \, \hat{y})}{\partial x \, \partial y} \right\rangle = \left\langle \frac{\partial \hat{x}}{\partial x} \, \frac{\partial \hat{y}}{\partial y} \right\rangle,$$

and we have

$$\frac{\partial \hat{x}}{\partial x} = 2 \, \delta(x), \quad \frac{\partial \hat{y}}{\partial y} = 2 \, \delta(y),$$

in the 1–bit quantization case. Therefore

$$\left\langle \frac{\partial \hat{x}}{\partial x} \, \frac{\partial \hat{y}}{\partial y} \right\rangle = \int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} 4 \, \delta(x) \, \delta(y) \, f(x, \, y) \, dx \, dy = \frac{2}{\pi \, \sigma_x \, \sigma_y \, \sqrt{1 - r^2}}.$$

Thus we have

$$\frac{\partial R_{\hat{x}\hat{y}}}{\partial r} = \frac{2}{\pi \, \sqrt{1 - r^2}},$$

and hence

$$R_{\hat{x}\hat{y}} = \frac{2}{\pi} \int\limits_{0}^{r} \frac{dr'}{\sqrt{1 - r'^2}} = \frac{2}{\pi} \, \arcsin(r),$$

where the limits of the integration are chosen to satisfy $R_{\hat{x}\hat{y}} = 0$ at $r = 0$.

Thus, Price's theorem allows us to derive the van Vleck relationship in the really straightforward way.

1.1.24 Cross–Correlation of the 2-Bit Quantized Data

Now we consider, for definiteness, cross–correlation $R_{\hat{x}\hat{y}}(\tau)$ of 2–bit quantized data $\hat{x}(t)$ and $\hat{y}(t)$, though logics below are equally applicable to autocorrelation as well. Let clipping criteria for the 2–bit quantization are those given in Figure 17 and Table 1. We assume a simple case where dispersions of the original analog data $x(t)$ and $y(t)$ are identical, and denote the common value as σ, i.e. $\sigma_x = \sigma_y = \sigma$. Then, the joint probability density is given by

$$f(x, \, y) \equiv f(x, \, y; \, \tau) = \frac{1}{2\pi\sigma^2 \, \sqrt{1 - r^2}} \, e^{-\dfrac{x^2 - 2 \, r \, x \, y + y^2}{2\sigma^2 \, (1 - r^2)}}, \qquad (103)$$

where $r \equiv r_{xy}(\tau)$ is the cross–correlation coefficient of the analog data.

According to Price's theorem, the cross–correlation of the clipped data $R_{\hat{x}\hat{y}}(\tau)$ satisfies $\quad \dfrac{1}{\sigma^2} \, \dfrac{\partial R_{\hat{x}\hat{y}}}{\partial r} = \left\langle \dfrac{\partial^2 (\hat{x} \, \hat{y})}{\partial x \, \partial y} \right\rangle = \left\langle \dfrac{\partial \hat{x}}{\partial x} \, \dfrac{\partial \hat{y}}{\partial y} \right\rangle,$ where the derivatives

are now given by

$$\frac{\partial \hat{x}}{\partial x} = (n-1)\,\delta(x+v_0) + 2\,\delta(x) + (n-1)\,\delta(x-v_0),$$

$$\frac{\partial \hat{y}}{\partial y} = (n-1)\,\delta(y+v_0) + 2\,\delta(y) + (n-1)\,\delta(y-v_0).$$

Using these derivatives, we can calculate

$$\frac{\partial R_{\hat{x}\hat{y}}}{\partial r} = \sigma^2 \left\langle \frac{\partial \hat{x}}{\partial x} \frac{\partial \hat{y}}{\partial y} \right\rangle = \sigma^2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{\partial \hat{x}}{\partial x} \frac{\partial \hat{y}}{\partial y}\, f(x,\, y)\, dx\, dy.$$

After simple manipulations, we obtain

$$\frac{\partial R_{\hat{x}\hat{y}}}{\partial r} = \frac{1}{\pi\sqrt{1-r^2}} \left\{ (n-1)^2 \left[e^{-\frac{v_0^2}{\sigma^2(1+r)}} + e^{-\frac{v_0^2}{\sigma^2(1-r)}} \right] + 4(n-1)e^{-\frac{v_0^2}{2\sigma^2(1-r^2)}} + 2 \right\}.$$

Therefore, the solution which satisfies a condition $R_{\hat{x}\hat{y}} = 0$ when $r = 0$, is

$$R_{\hat{x}\hat{y}} = \frac{1}{\pi} \int_0^r \frac{1}{\sqrt{1-r'^2}} \left\{ (n-1)^2 \left[e^{-\frac{v_0^2}{\sigma^2(1+r')}} + e^{-\frac{v_0^2}{\sigma^2(1-r')}} \right] \right.$$
$$\left. + 4(n-1)e^{-\frac{v_0^2}{2\sigma^2(1-r'^2)}} + 2 \right\} dr'. \qquad (104)$$

In the limiting case where $|\,r\,| \ll 1$, the cross–correlation is given by

$$R_{\hat{x}\hat{y}} = \frac{r}{\pi} \left[2\,(n-1)^2\, e^{-\frac{v_0^2}{\sigma^2}} + 4\,(n-1)\, e^{-\frac{v_0^2}{2\sigma^2}} + 2 \right] = \frac{2\,r}{\pi} \left[(n-1)\,E + 1 \right]^2, \quad (105)$$

where we introduced a notation

$$E \equiv e^{-\frac{v_0^2}{2\sigma^2}}. \qquad (106)$$

Note that the cross–correlation of the clipped data $R_{\hat{x}\hat{y}}(\tau)$ is proportional to the analog cross–correlation coefficient $r = r_{xy}(\tau)$, in this case.

1.1.25 Cross–Correlation Coefficient of the 2–Bit Quantized Data

Now we would like to calculate cross–correlation coefficient $r_{\hat{x}\hat{y}}(\tau)$ of the 2–bit quantized data. For this purpose, we need their dispersions: $\sigma_{\hat{x}\hat{x}}^2 = R_{\hat{x}\hat{x}}(0)$ and $\sigma_{\hat{y}\hat{y}}^2 = R_{\hat{y}\hat{y}}(0)$, other than the cross–correlation $R_{\hat{x}\hat{y}}(\tau)$.

We assumed here that analog data $x(t)$ and $y(t)$ have equal dispersions: $R_{xx}(0) = R_{yy}(0) = \sigma^2$, and therefore they obey identical Gaussian probability densities:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{x^2}{2\sigma^2}}, \quad \text{and} \quad f(y) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{y^2}{2\sigma^2}}. \tag{107}$$

Since 4 quantization states in the 2–bit clipped data take values $-n$, -1, $+1$, and $+n$, the joint probabilities of the quantization states of $\hat{x}[n]$ and $\hat{x}[n]$ (or $\hat{y}[n]$ and $\hat{y}[n]$), taken at the same time, satisfy

$$P(-n,\ -n;\ 0) + P(-1,\ -1;\ 0) + P(+1,\ +1;\ 0) + P(+n,\ +n;\ 0) = 1, \tag{108}$$

and the dispersion: $R_{\hat{x}\hat{x}}(0) = \langle \hat{x}(t)\,\hat{x}(t)\rangle = \sum\limits_{i=1}^{4} x_i\, x_i\, P(x_i,\ x_i;\ 0)$, must be

$$\begin{aligned} R_{\hat{x}\hat{x}}(0) = n^2\, P(-n,\ -n;\ 0) &+ P(-1,\ -1;\ 0) \\ &+ P(+1,\ +1;\ 0) + n^2\, P(+n,\ +n;\ 0). \end{aligned} \tag{109}$$

Combining equations (108) and (109), and repeating the same thing for $\hat{y}$, we obtain

$$R_{\hat{x}\hat{x}}(0) = R_{\hat{y}\hat{y}}(0) = \Phi + n^2\,(1-\Phi), \tag{110}$$

where we introduced a notation:

$$\Phi = P(-1,\ -1;\ 0) + P(+1,\ +1;\ 0) = \frac{1}{\sigma\sqrt{2\pi}} \int\limits_{-v_0}^{+v_0} e^{-\frac{\xi^2}{2\sigma^2}}\, d\xi. \tag{111}$$

We used here the explicit form of the probability density in equation (107).

Thus, the cross–correlation coefficient of the 2–bit quantized data $r_{\hat{x}\hat{y}}(\tau)$ is given by

$$\begin{aligned} r_{\hat{x}\hat{y}}(\tau) &= \frac{R_{\hat{x}\hat{y}}(\tau)}{\sqrt{R_{\hat{x}\hat{x}}(0)\,R_{\hat{y}\hat{y}}(0)}} = \frac{R_{\hat{x}\hat{y}}(\tau)}{\Phi + n^2\,(1-\Phi)} \\ &= \frac{1}{\pi\,[\Phi + n^2\,(1-\Phi)]} \int\limits_{0}^{r_{xy}(\tau)} \frac{1}{\sqrt{1-r'^2}} \left\{ (n-1)^2 \left[e^{-\frac{v_0^2}{\sigma^2(1+r')}} + e^{-\frac{v_0^2}{\sigma^2(1-r')}} \right] \right. \\ &\qquad \left. + 4(n-1)e^{-\frac{v_0^2}{2\sigma^2(1-r'^2)}} + 2 \right\} dr', \end{aligned} \tag{112}$$

as a function of the cross–correlation coefficient of the original analog data $r = r_{xy}(\tau)$. This is an analogue of the van Vleck relationship in the 2–bit quantization case.

61

In the limiting case of $|\ r_{xy}(\tau)\ | \ll 1$, equations (105) and (110) yield

$$r_{\hat{x}\hat{y}}(\tau) = \frac{2\left[(n-1)\,E+1\right]^2}{\pi\left[\Phi + n^2\,(1-\Phi)\right]}\,r_{xy}(\tau). \tag{113}$$

This is an analogue of equation (91) in the 2–bit quantization case.
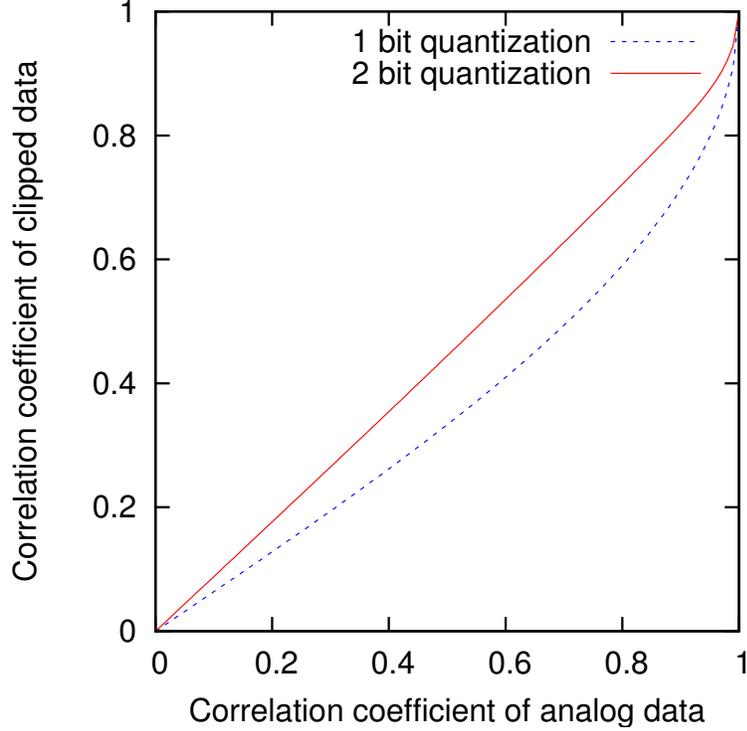
Figure 24: Correlation coefficients of 1–bit (dotted line) and 2–bit (solid line) quantized data with $n = 3$ and $v_0 = 0.996\,\sigma$ as functions of the analog correlation coefficient.

Exactly the same logic leads to the same functional forms as equations (112) and (113) for a relationship between the correlation coefficient of the 2–bit quantized data $r_{\hat{x}\hat{x}}(\tau)$ and that of the original analog data $r_{xx}(\tau)$. Thus the correlation coefficient of the 2–bit quantized data is given by

$$r_{\hat{x}\hat{x}}(\tau) = \frac{1}{\pi\left[\Phi + n^2\,(1-\Phi)\right]} \int_0^{r_{xx}(\tau)} \frac{1}{\sqrt{1-r'^2}} \left\{ (n-1)^2 \left[e^{-\frac{v_0^2}{\sigma^2(1+r')}} + e^{-\frac{v_0^2}{\sigma^2(1-r')}} \right] \right.$$

$$\left. +4(n-1)e^{-\frac{v_0^2}{2\sigma^2(1-r'^2)}} + 2 \right\}\,dr', \tag{114}$$

for a general case, and

$$r_{\hat{x}\hat{x}}(\tau) = \frac{2\left[(n-1)E+1\right]^2}{\pi\left[\Phi + n^2(1-\Phi)\right]} r_{xx}(\tau), \tag{115}$$

for the limiting case of $\mid r_{xx}(\tau) \mid \ll 1$.

Figure 24 shows the relationship between the correlation coefficient, which could be either (auto–)correlation coefficient or cross–correlation coefficient, of the clipped data and that of the original analog data. The dotted line shows the van Vleck relationship for the 1–bit quantization case as given by equation (90) or (95). The solid line shows the relationship for the 2–bit quantization case as given by equation (112) or (114) for a particular set of parameters $n = 3$ and $v_0 = 0.996\,\sigma$.

1.1.26 Power and Cross–Power Spectra of 2-Bit Quantized Data

We can now calculate power spectrum $S_{\hat{x}\hat{x}}(\omega)$ and cross–power spectrum $S_{\hat{x}\hat{y}}(\omega)$ of the 2–bit quantized data, which are normalized by the dispersion $\sigma_{\hat{x}}^2$ and geometric mean of the dispersions $\sigma_{\hat{x}}\,\sigma_{\hat{y}}$, respectively, by Fourier transforming the correlation coefficient $r_{\hat{x}\hat{x}}(\tau)$ and the cross–correlation coefficient $r_{\hat{x}\hat{y}}(\tau)$ given in equation (112),

$$S_{\hat{x}\hat{x}}(\omega) = \int_{-\infty}^{\infty} r_{\hat{x}\hat{x}}(\tau)\,e^{-i\omega\tau}\,d\tau = \int_{-\infty}^{\infty} F(r_{xx}(\tau))\,e^{-i\omega\tau}\,d\tau,$$

$$S_{\hat{x}\hat{y}}(\omega) = \int_{-\infty}^{\infty} r_{\hat{x}\hat{y}}(\tau)\,e^{-i\omega\tau}\,d\tau = \int_{-\infty}^{\infty} F(r_{xy}(\tau))\,e^{-i\omega\tau}\,d\tau, \tag{116}$$

where the function $F(r)$ is given by

$$F(r) = \frac{1}{\pi\left(\Phi + n^2(1-\Phi)\right)} \int_0^r \frac{1}{\sqrt{1-r'^2}} \left\{ (n-1)^2 \left[e^{-\frac{v_0^2}{\sigma^2(1+r')}} + e^{-\frac{v_0^2}{\sigma^2(1-r')}} \right] \right.$$

$$\left. + 4(n-1)e^{-\frac{v_0^2}{2\sigma^2(1-r'^2)}} + 2 \right\} dr'.$$

Let us again consider the case when analog data have rectangular spectra. Figure 25 shows the power spectrum $S_{\hat{x}\hat{x}}(\omega)$ of the 2–bit quantized data in the case with $n = 3$ and $v_0 = 0.996\,\sigma$ (solid line in bottom panel) and that of the original analog data $S_{xx}(\omega)$ with the rectangular shape of bandwidth B (broken line in bottom panel). Similarly to the 1–bit case, areas under the spectra (i.e. powers) of the analog and clipped data are equal. Also shown are (1) correlation coefficient of the analog data $r_{xx}(\tau)$ having a sinc function

63
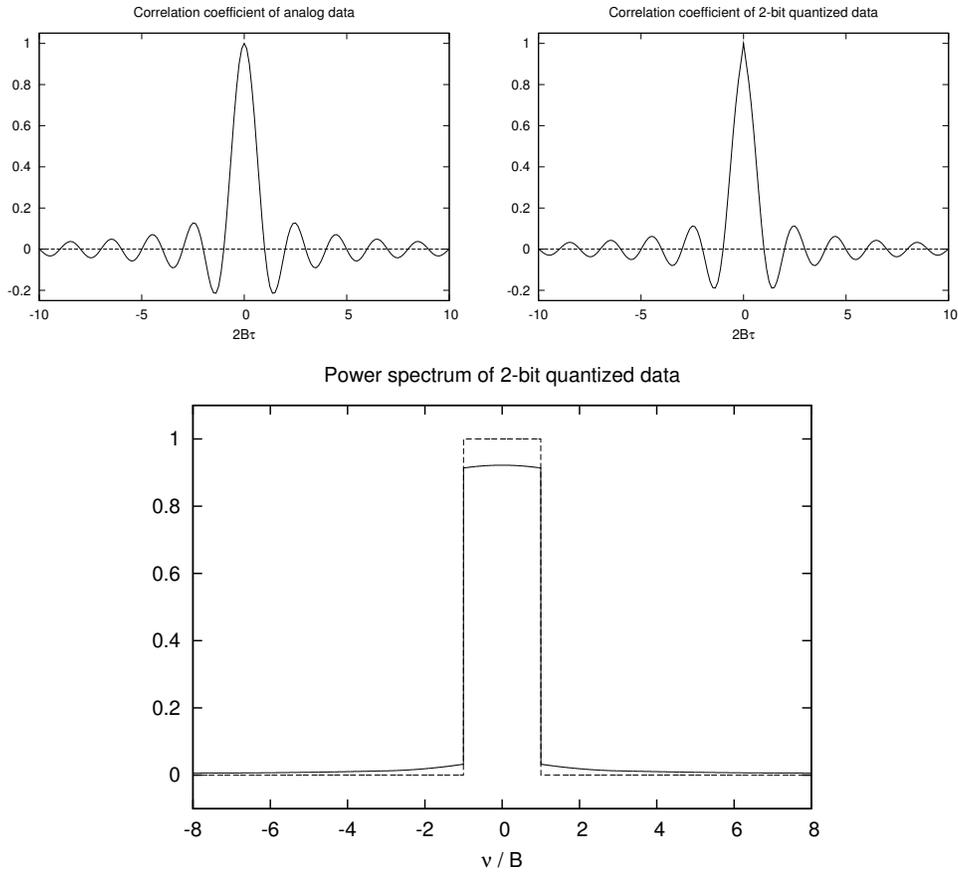
Figure 25: Original rectangular power spectrum $S_{xx}(\omega)$ with bandwidth B (broken line in bottom panel) of the analog data, and power spectrum $S_{\hat{x}\hat{x}}(\omega)$ (solid line in bottom panel) of the 2–bit quantized data with $n = 3$ and $v_0 = 0.996\,\sigma$. Horizontal axis of the bottom panel shows frequency ν normalized by the bandwidth B. Upper panels show correlation coefficient $r_{xx}(\tau)$ of the original analog data having a sinc function form (left), and correlation coefficient $r_{\hat{x}\hat{x}}(\tau)$ of the 2–bit quantized data (right). Horizontal axes of the upper panels show delay τ normalized by the Nyquist interval $(1\,/\,(2\,B))$ of the original analog data.
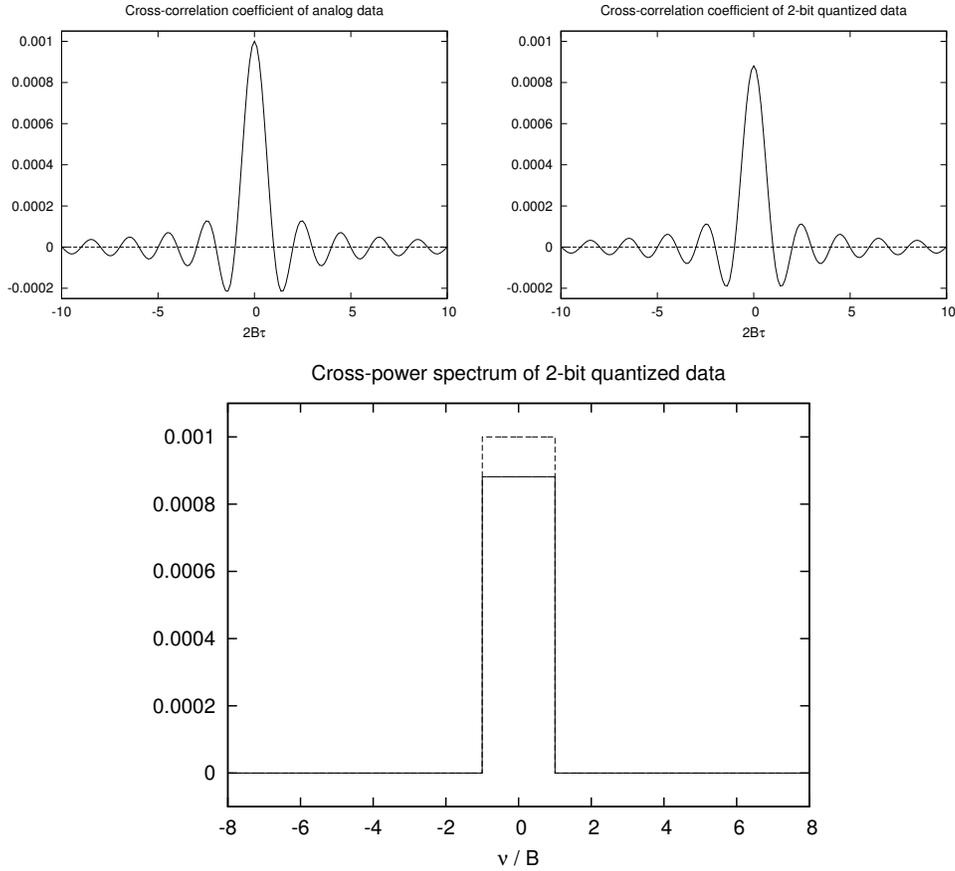
Figure 26: Original rectangular cross–power spectrum $S_{xy}(\omega)$ with bandwidth B (broken line in bottom panel) of the analog data, and cross–power spectrum $S_{\hat{x}\hat{y}}(\omega)$ (solid line in bottom panel) of the 2–bit quantized data with $n = 3$ and $v_0 = 0.996\,\sigma$. Horizontal axis of the bottom panel shows frequency ν normalized by the bandwidth B. Upper panels show cross–correlation coefficient $r_{xy}(\tau)$ of the original analog data having a sinc function form (left), and cross–correlation coefficient $r_{\hat{x}\hat{y}}(\tau)$ of the 2–bit quatized data (right). The maximum cross–correlation coefficient of the analog data is assumed to be 0.001. Horizontal axes of the upper panels show delay τ which is normalized by the Nyquist interval $(1\,/\,(2\,B))$ of the original analog data.

65

form (upper left panel), and (2) correlation coefficient of the 2–bit quantized data $r_{\hat{x}\hat{x}}(\tau)$ (upper right panel).

The spectrum after clipping again shows somewhat reduced peak amplitude and a low–level skirt extending over a wide range of frequency. Thus, the original Nyquist interval ($1 / (2B)$) is not strictly optimum for sampling the clipped data again, though to a smaller extent compared with the 1–bit case.

On the other hand, Figure 26 shows a cross–power spectrum $S_{\hat{x}\hat{y}}(\omega)$ of the 2–bit quantized data $\hat{x}$ and $\hat{y}$ in the case with $n = 3$ and $v_0 = 0.996\,\sigma$ (solid line in bottom panel) and that of the original analog data x and y, having a real rectangular cross–power spectrum $S_{xy}(\omega)$ of bandwidth B (broken line in bottom panel). We assumed here again that the maximum cross–correlation coefficient of the analog data is as small as 0.001, for definiteness. Upper panels of the Figure show (1) cross–correlation coefficient of the analog data $r_{xy}(\tau)$ having a sinc function form (left), and (2) cross–correlation coefficient of the clipped data $r_{\hat{x}\hat{y}}(\tau)$ (right). Note that the spectrum after clipping remains again rectangular, because of the approximate linearity between the analog and clipped cross–correlation coefficients.

1.1.27   Dispersion of Digital Correlator Output

Let us now examine the signal–to–noise ratio of an output of a digital correlator. How does the S/N ratio differ from the one expected from the analog or "semi–analog" correlator, which we saw earlier? For this purpose, we first consider dispersion of the digital correlator output.

The output $\mathcal{R}$ of a digital correlator, which averages products of $\mathcal{N}$ samples of clipped data $\hat{x}[i]$ and $\hat{y}[i]$ from two antennas in a radio interferometer, is

$$\mathcal{R} = \frac{1}{\mathcal{N}} \sum_{i=1}^{\mathcal{N}} \hat{x}[i]\,\hat{y}[i]. \tag{117}$$

We again assume here that the delay tracking and the fringe stopping are completely performed before the multiplication, so that signal parts in the two data streams $\hat{x}[i]$ and $\hat{y}[i]$ are perfectly aligned.

Under the assumption of the stationary random processes, the expectation $\langle \mathcal{R} \rangle$ and the dispersion $\sigma_{\mathcal{R}}^2$ of the output $\mathcal{R}$ are given by

$$\langle \mathcal{R} \rangle = \langle \hat{x}[i]\,\hat{y}[i] \rangle = R_{\hat{x}\hat{y}}[0], \tag{118}$$

$$\sigma_{\mathcal{R}}^2 = \langle \mathcal{R}^2 \rangle - \langle \mathcal{R} \rangle^2 = \frac{1}{\mathcal{N}^2} \sum_{i=1}^{\mathcal{N}} \sum_{j=1}^{\mathcal{N}} \langle \hat{x}[i]\,\hat{y}[i]\,\hat{x}[j]\,\hat{y}[j] \rangle - R_{\hat{x}\hat{y}}^2[0]. \tag{119}$$

Note that usually $\mathcal{N}$ is a huge number. For example, if we integrate data sampled with a rate of 4 Msps (mega sample per second) for 1 second, then $\mathcal{N} \approx 4,000,000$.

Let us assume that the original analog data $x(t)$ and $y(t)$ have rectangular baseband spectra with bandwith B, and the data are sampled with Nyquist rate $2B$. In such a case, different sample pairs are independent in the analog data, as we discussed earlier, i.e. $R_{xy}[m]$, $R_{xx}[m]$, and $R_{yy}[m]$ are all zero, if $m \neq 0$. The same statement is valid for the 1–bit quantized and 2–bit quantized data, since correlation coefficients, and correlations, too, of these clipped data are equal to zero when those of the original analog data are zero, as we can easily see in equations (90), (95), and (112). Thus, $R_{\hat{x}\hat{y}}[m]$, $R_{\hat{x}\hat{x}}[m]$, and $R_{\hat{y}\hat{y}}[m]$ are also all zero, if $m \neq 0$. In addition, we assume that the cross–correlation $R_{\hat{x}\hat{y}}[0]$ is much smaller than the autocorrelations $R_{\hat{x}\hat{x}}[0]$ and $R_{\hat{y}\hat{y}}[0]$, as usually so in radio interferometers.

Then in the double sum of equation (119):

$$\sum_{i=1}^{\mathcal{N}} \sum_{j=1}^{\mathcal{N}} \langle \hat{x}[i]\,\hat{y}[i]\,\hat{x}[j]\,\hat{y}[j] \rangle = \sum_{i=1}^{\mathcal{N}} \langle \hat{x}[i]\,\hat{y}[i]\,\hat{x}[i]\,\hat{y}[i] \rangle + \sum_{i=1}^{\mathcal{N}} \sum_{j \neq i}^{\mathcal{N}} \langle \hat{x}[i]\,\hat{y}[i]\,\hat{x}[j]\,\hat{y}[j] \rangle,$$
(120)

dominating terms will be

$$\sum_{i=1}^{\mathcal{N}} \sum_{j=1}^{\mathcal{N}} \langle \hat{x}[i]\,\hat{y}[i]\,\hat{x}[j]\,\hat{y}[j] \rangle \cong \sum_{i=1}^{\mathcal{N}} \langle \hat{x}[i]\,\hat{x}[i] \rangle \langle \hat{y}[i]\,\hat{y}[i] \rangle + \sum_{i=1}^{\mathcal{N}} \sum_{j \neq i}^{\mathcal{N}} \langle \hat{x}[i]\,\hat{y}[i] \rangle \langle \hat{x}[j]\,\hat{y}[j] \rangle$$

$$= \mathcal{N}\,R_{\hat{x}\hat{x}}[0]\,R_{\hat{y}\hat{y}}[0] + \mathcal{N}\,(\mathcal{N}-1)\,R_{\hat{x}\hat{y}}^2[0] \cong \mathcal{N}\,R_{\hat{x}\hat{x}}[0]\,R_{\hat{y}\hat{y}}[0] + \mathcal{N}^2\,R_{\hat{x}\hat{y}}^2[0],$$
(121)

where we neglected 1 compared with the large number $\mathcal{N}$.

Now equations (119) and (121) yeild an approximate formula for the dispersion of the digital correlator output:

$$\sigma_{\mathcal{R}}^2 \cong \frac{1}{\mathcal{N}}\,R_{\hat{x}\hat{x}}[0]\,R_{\hat{y}\hat{y}}[0] + R_{\hat{x}\hat{y}}^2[0] - R_{\hat{x}\hat{y}}^2[0] = \frac{1}{\mathcal{N}}\,R_{\hat{x}\hat{x}}[0]\,R_{\hat{y}\hat{y}}[0]. \qquad (122)$$

1.1.28 S/N Ratio of Digital Correlator Output

We are now ready to calculate the signal $\langle \mathcal{R} \rangle$ to noise $\sigma_{\mathcal{R}}$ ratio SNR of the digital correlator output, using equation (122). We obtain

$$SNR = \frac{\langle \mathcal{R} \rangle}{\sigma_{\mathcal{R}}} = \sqrt{\mathcal{N}}\,\frac{R_{\hat{x}\hat{y}}[0]}{\sqrt{R_{\hat{x}\hat{x}}[0]\,R_{\hat{y}\hat{y}}[0]}} = \sqrt{\mathcal{N}}\,r_{\hat{x}\hat{y}}[0] = \sqrt{\mathcal{N}}\,\frac{r_{\hat{x}\hat{y}}[0]}{r_{xy}[0]}\,\rho, \qquad (123)$$

where $\rho \equiv r_{xy}[0]$ is the maximum cross–correlation coefficient of the original analog data under our assumption of the perfect delay tracking and fringe

67

stopping. As we saw earlier, the maximum cross–correlation coefficient ρ for a continuum spectrum source is approximately given by

$$\rho = \sqrt{\frac{T_{A_1} T_{A_2}}{T_{S_1} T_{S_2}}},$$

where T_{A_1}, T_{A_2} and T_{S_1}, T_{S_2} are antenna temperatures and system noise temperatures, respectively, at antennas 1 and 2, which are both assumed constant throuout the frequency band B. On the other hand, if the integration time is τ_a, the number of samples $\mathcal{N}$ with the Nyquist interval $1/(2B)$ is equal to

$$\mathcal{N} = 2 B \tau_a.$$

Therefore, equation (123) is reduced to

$$SNR = \frac{r_{\hat{x}\hat{y}}[0]}{r_{xy}[0]} \sqrt{\frac{T_{A_1} T_{A_2}}{T_{S_1} T_{S_2}}} \sqrt{2 B \tau_a} = \frac{r_{\hat{x}\hat{y}}[0]}{r_{xy}[0]} SNR_{analog}, \tag{124}$$

where

$$SNR_{analog} \equiv \sqrt{\frac{T_{A_1} T_{A_2}}{T_{S_1} T_{S_2}}} \sqrt{2 B \tau_a},$$

is the signal–to–noise ratio of an analog correlator output for the same continuum spectrum source received with the same antenna–receiver systems, as we saw in equation (65).

Thus the so–called "coherence factor" η_c, which is defined as

$$\eta_c \equiv \frac{SNR}{SNR_{analog}}, \tag{125}$$

is given by

$$\eta_c = \frac{r_{\hat{x}\hat{y}}[0]}{r_{xy}[0]}, \tag{126}$$

in this case.

For the case of the small cross–correlation coefficient $|\, r_{xy}[0]\,| \ll 1$, which is usually the case in radio interferometry, $r_{\hat{x}\hat{y}}[0]$ of the clipped data is proportional to $r_{xy}[0]$ of the original analog data. Then, in view of equations (91) and (113), we have

$$\eta_c = \frac{2}{\pi} \cong 0.64, \tag{127}$$

for the 1–bit quantization case, and

$$\eta_c = \frac{2 \left[(n-1) E + 1\right]^2}{\pi \left[\Phi + n^2 (1 - \Phi)\right]}, \tag{128}$$

for the 2–bit quantization case.

With the coherence factor η_c, equation (124) is reduced to

$$SNR = \eta_c \sqrt{\frac{T_{A_1} T_{A_2}}{T_{S_1} T_{S_2}}} \sqrt{2 B \tau_a}, \qquad (129)$$

which is an important formula for estimating sensitivity of a radio interfer-
ometer observing a continuum spectrum source. Note that the coherence
factor as given by equation (126) still takes into account only the loss due
to effects of clipping. Usually, the factor is further reduced in view of losses
occuring through digital processing in hardware correlators.

1.1.29 Optimum Parameters v_0 and n in the 2–Bit Quantization

We mentioned earlier that the two parameters in the 2–bit quantization, the
threshold v_0 and the higher–level value n, are chosen so that the signal–to–
noise ratio of the digital correlator output is maximized. This condition is
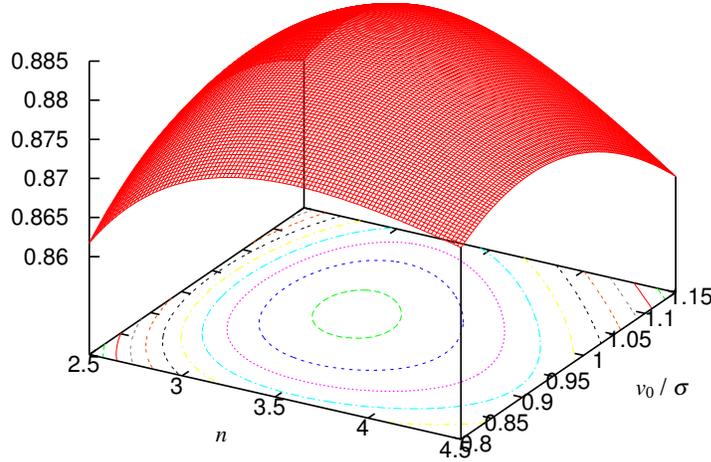
Figure 27: Coherence factor of the 2–bit quantized data as a function of
normalized threshold v_0/σ and higher–level value n in the limiting case of
the small cross–correlation coefficient.

fulfilled when the coherence factor η_c takes the maximum. In the limiting case of the small cross–correlation coeficient $\mid r_{xy}[0] \mid \ll 1$, which is common in radio interferometry, we can calculate the coherence factor η_c using equation (128) with explicit forms for the functions E and Φ given in equations (106) and (111).

Figure 27 shows the coherece factor as a function of the two parameters, v_0 normalized by the analog dispersion σ, and n, in the small cross–correlation coefficient case.

As we see from the Figure, the maximum value of the coherence factor $\eta_c = 0.883$ is obtained when $n = 3.34$ and $v_0 = 0.982\,\sigma$. However, from a practical point of view, designing of digital circuitry for 2–bit hardware correlators can be more easily implemented when n is an integer. Therefore, $n = 3$, $v_0 = 0.996\,\sigma$ with $\eta_c = 0.881$, and $n = 4$, $v_0 = 0.942\,\sigma$ with $\eta_c = 0.880$, are often used in existing VLBI 2–bit quantization systems. This is why we adopted $n = 3$, $v_0 = 0.996\,\sigma$ in Figures 24, 25, and 26.

Accordingly, in the limiting case of the small cross–correlation coefficient $\mid r_{xy}[0] \mid \ll 1$, the coherence factor, which takes into acount the effect of clipping only, is given by

$$\eta_c \cong 0.64, \tag{130}$$

for the 1–bit quantization case, and

$$\eta_c \cong 0.88, \tag{131}$$

for the 2–bit quantization case.

1.1.30 Effect of Oversampling in S/N Ratio of Clipped Data

As we saw earlier, power spectra of clipped data show low–level but broad skirt beyond band edges at $\nu = \pm B$ of the original rectangular spectra of corresponding analog data (Figures 22 and 25). Therefore, the Nyquist sampling for the original analog data with $2B$ rate is no longer optimum for the clipped data.

Sampling with a rate higher than the Nyquist rate for the analog data $2B$, which is a little incorrectly called the "oversampling", improves the signal–to–noise ratio of the clipped data. In this case, different sample points are no longer independent any more, and we have to take into account contributions of autocorrelations between different sample points, i.e. $R_{xx}[m]$ and $R_{yy}[m]$ with $m \neq 0$, when we calculate the signal–to–noise ratio.

Let us consider that we sample our data with a rate which is β times as fast as the Nyquist rate of the original band–limited analog data $2B$. Then, in the calculation of the dispersion of the digital correlator output $\mathcal{R}$ shown

70

in equations (119) – (122), we must leave autocorrelation terms with non–zero arguments. Specifically, the second term in the RHS of equation (120) now must be

$$\sum_{i=1}^{\mathcal{N}} \sum_{j\neq i}^{\mathcal{N}} \langle \hat{x}[i]\,\hat{y}[i]\,\hat{x}[j]\,\hat{y}[j]\rangle \cong \sum_{i=1}^{\mathcal{N}} \sum_{j\neq i}^{\mathcal{N}} [\langle \hat{x}[i]\,\hat{y}[i]\rangle\,\langle \hat{x}[j]\,\hat{y}[i]\rangle + \langle \hat{x}[i]\,\hat{x}[j]\rangle\,\langle \hat{y}[i]\,\hat{y}[i]\rangle],$$

where the products of the autocorrelations in the second term in the RHS could be well larger than the products of the cross–correlations in the first term in this oversampling case. Thus the dispersion now becomes

$$\sigma_{\mathcal{R}}^2 = \frac{1}{\mathcal{N}^2} \left[\sum_{i=1}^{\mathcal{N}} \langle \hat{x}[i]\,\hat{y}[i]\,\hat{x}[i]\,\hat{y}[i]\rangle + \sum_{i=1}^{\mathcal{N}} \sum_{j\neq i}^{\mathcal{N}} \langle \hat{x}[i]\,\hat{y}[i]\,\hat{x}[j]\,\hat{y}[j]\rangle \right] - R_{\hat{x}\hat{y}}^2[0]$$

$$\cong \frac{1}{\mathcal{N}} R_{\hat{x}\hat{x}}[0]\,R_{\hat{y}\hat{y}}[0] + \frac{1}{\mathcal{N}^2} \sum_{i=1}^{\mathcal{N}} \sum_{j\neq i}^{\mathcal{N}} R_{\hat{x}\hat{x}}[i-j]\,R_{\hat{y}\hat{y}}[i-j], \qquad (132)$$

where the second term in the RHS stands for the largest contribution of the autocorrelations between different i–th and j–th sample points. Since number of combinations of i and j having the same diffference $k = i - j$ is $\mathcal{N} - |k|$, and autocorrelations are likely to be large enough for small $|k| \ll \mathcal{N}$ only, we have

$$\sigma_{\mathcal{R}}^2 \cong \frac{1}{\mathcal{N}} R_{\hat{x}\hat{x}}[0]\,R_{\hat{y}\hat{y}}[0] + \frac{2}{\mathcal{N}^2} \sum_{k=1}^{\mathcal{N}-1} (\mathcal{N}-k) R_{\hat{x}\hat{x}}[k]\,R_{\hat{y}\hat{y}}[k]$$

$$\cong \frac{1}{\mathcal{N}} R_{\hat{x}\hat{x}}[0]\,R_{\hat{y}\hat{y}}[0]\,(1 + 2\sum_{k=1}^{\infty} r_{\hat{x}\hat{x}}[k]\,r_{\hat{y}\hat{y}}[k]). \qquad (133)$$

Thus, for a continuum spectrum source, the signal–to–noise ratio of the digital correlator output of $\mathcal{N} = 2\,\beta\,B\,\tau_a$ oversampled data is approximately given by

$$SNR = \frac{\langle \mathcal{R} \rangle}{\sigma_{\mathcal{R}}} = \frac{r_{\hat{x}\hat{y}}[0]}{r_{xy}[0]} \sqrt{\frac{T_{A_1}\,T_{A_2}}{T_{S_1}\,T_{S_2}}} \frac{\sqrt{2\,\beta\,B\,\tau_a}}{\sqrt{1 + 2\sum_{k=1}^{\infty} r_{\hat{x}\hat{x}}[k]\,r_{\hat{y}\hat{y}}[k]}}. \qquad (134)$$

Therefore, the coherence factor in the case of the oversampling is given by

$$\eta_c = \frac{SNR}{SNR_{analog}} = \frac{r_{\hat{x}\hat{y}}[0]}{r_{xy}[0]} \frac{\sqrt{\beta}}{\sqrt{1 + 2\sum_{k=1}^{\infty} r_{\hat{x}\hat{x}}[k]\,r_{\hat{y}\hat{y}}[k]}}. \qquad (135)$$

We can estimate this factor, using equations (95) and (114) which describe correlation coefficients of the clipped data $r_{\hat{x}\hat{x}}[k]$ and $r_{\hat{y}\hat{y}}[k]$ as functions of analog correlation coefficients $r_{xx}[k]$ and $r_{yy}[k]$, which we denote as $r_{\hat{x}\hat{x}}[k] = f_{cl}(r_{xx}[k])$ and $r_{\hat{y}\hat{y}}[k] = f_{cl}(r_{yy}[k])$. Here f_{cl} is the clipping function which is given in equation (95) and equation (114) for the 1–bit and 2–bit quatization cases, respectively.

When the original analog data have rectangular baseband spectra with bandwidth B, as we have assumed in the present discussion, the analog correlation coefficients $r_{xx}(\tau)$ and $r_{yy}(\tau)$ have sinc function forms:

$$r_{xx}(\tau) = r_{yy}(\tau) = \frac{\sin(2\pi B \tau)}{2\pi B \tau},$$

as given in equation (46). Therefore, correlation coefficients of their time samples $x[k]$ and $y[k]$ with the sampling interval $T_s = 1/(2 B \beta)$ are given by

$$r_{xx}[k] = r_{yy}[k] = r_{xx}(k T_s) = r_{yy}(k T_s) = \frac{\sin\left(\frac{\pi k}{\beta}\right)}{\frac{\pi k}{\beta}}. \tag{136}$$
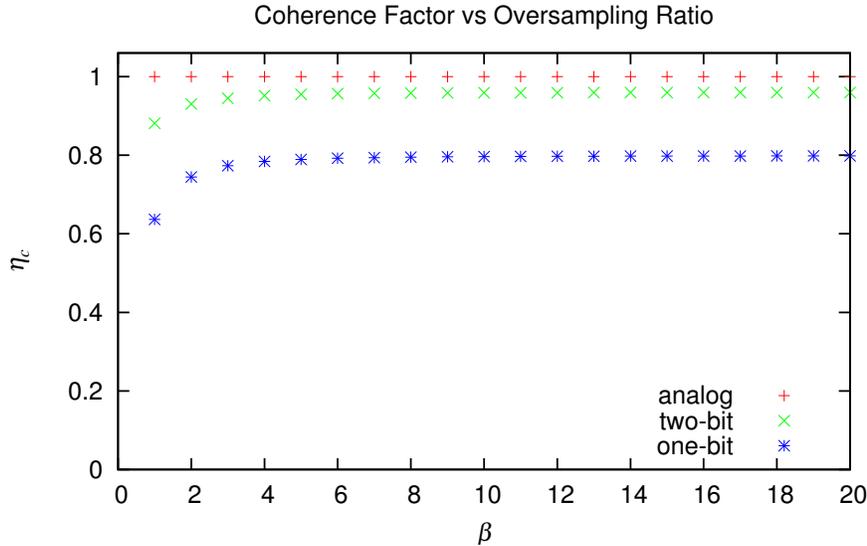
Figure 28: Coherence factor η_c for the analog (+), 2–bit quantized ($\times$), and 1–bit quantized ($*$) data as functions of the oversampling ratio β.

Thus, the signal–to–noise ratio and the coherence factor of the oversampled clipped data is given by

$$SNR = \eta_c \sqrt{\frac{T_{A_1} T_{A_2}}{T_{S_1} T_{S_2}}} \sqrt{2 B \tau_a}, \tag{137}$$

$$\eta_c = \frac{r_{\hat{x}\hat{y}}[0]}{r_{xy}[0]} \frac{\sqrt{\beta}}{\sqrt{1 + 2\sum_{k=1}^{\infty} f_{cl}^2 \left(\frac{\sin\left(\frac{\pi k}{\beta}\right)}{\frac{\pi k}{\beta}} \right)}}. \tag{138}$$

Figure 28 shows the coherence factor η_c for the analog, 2–bit quantized, and 1–bit quantized data as functions of the oversampling ratio β, calculated by means of equation (138). The coherence factor of the clipped data improves gradually with increasing β, approaching to some constant values. On the other hand, the coherence factor of the analog data is always 1, irrespective of the oversampling ratio β, in accordance with the sampling theorem.

In a particular case of the "double–speed sampling" $\beta = 2$, we have $\eta_c = 0.744$ for the 1–bit quantized data, and $\eta_c = 0.930$ for the 2–bit quantized data.

1.1.31 Coherence Factor and Sensitivity with Given Bit Rate

Coherence factor η_c with different clipping and oversampling is summarized in Table 2. It is obvious that the coherence factor increases as we increase

Number of	Number of	coherence factor η_c	
bits N_b	quantization levels	$\beta = 1$	$\beta = 2$
1	2	0.64	0.74
2	4	0.88	0.93

Table 2: Relationship of coherence factor η_c with number of bits N_b and oversampling ratio β.

the number of bits N_b and the oversampling ratio β.

However, actual observations are usually limited by maximum bit–rate ν_b of data streams allowed by correlators or recorders. If the maximum bit–rate ν_b is fixed, the maximum permissible sampling rate, which is β times as large as the Nyquist rate $2B$ with the analog bandwidth B, is $\nu_b/N_b = 2\beta B$. Therefore, the maximum allowable bandwidth is limited by the bit–rate as $B = \nu_b/(2\beta N_b)$. Hence, in view of equation (137), the maximum signal–to–noise ratio, i.e. the sensitivity of the interferometric observation, for a continuum spectrum source, is proportional to

$$SNR \propto \eta_c \sqrt{B} \propto \frac{\eta_c}{\sqrt{\beta N_b}}. \tag{139}$$

Number of bits N_b	Number of quantization levels	$\eta_c/\sqrt{\beta\,N_b}$	
		$\beta = 1$	$\beta = 2$
1	2	0.64	0.52
2	4	0.62	0.47

Table 3: Factor $\eta_c/\sqrt{\beta\,N_b}$, which determines the signal–to–noise ratio, estimated for clipped and oversampled data with a fixed bit–rate.

Table 3 shows the factor $\eta_c/\sqrt{\beta\,N_b}$, which determines the signal–to–noise ratio for the continuum spectrum source. When the bit–rate is given, the sensitivity turns out to be largest in the simplest case of the 1–bit quantization ($N_b = 1$) with the Nyquist sampling ($\beta = 1$)! This is the reason why many VLBI observations still use the 1–bit quantization scheme with the Nyquist sampling.

However, if we observe a line spectrum source, such as an astronomical maser source, the frequency range containing the signal from the radio source is confined within a limited spectral profile of the source where we often see many narrow lines. Thus, in order to estimate the signal–to–noise ratio of a spectral line, we must replace in equation (137) the bandwidth B of our receiving system by the width of the spectral line. The line width is intrincic to the radio source and constant irrespective of system–dependent parameters, such as the number of bits N_b, or the oversampling ratio β. Therefore, the larger the coherence factor η_c, the higher is the sensitivity in this case. Consequently, modern VLBI systems tend to adopt the 2–bit quantization scheme for better performance in line–spectrum–source observations. Note also that the 2–bit quantization scheme ($N_b = 2$) with the Nyquist sampling ($\beta = 1$) offers almost the same sensitivity as the 1–bit quantization scheme with the Nyquist sampling, i.e. 0.62 against 0.64, for a continuum source, as evident from Table 3. Figure 29 shows an example of the maser source spectra derived from 2–bit quantized data.
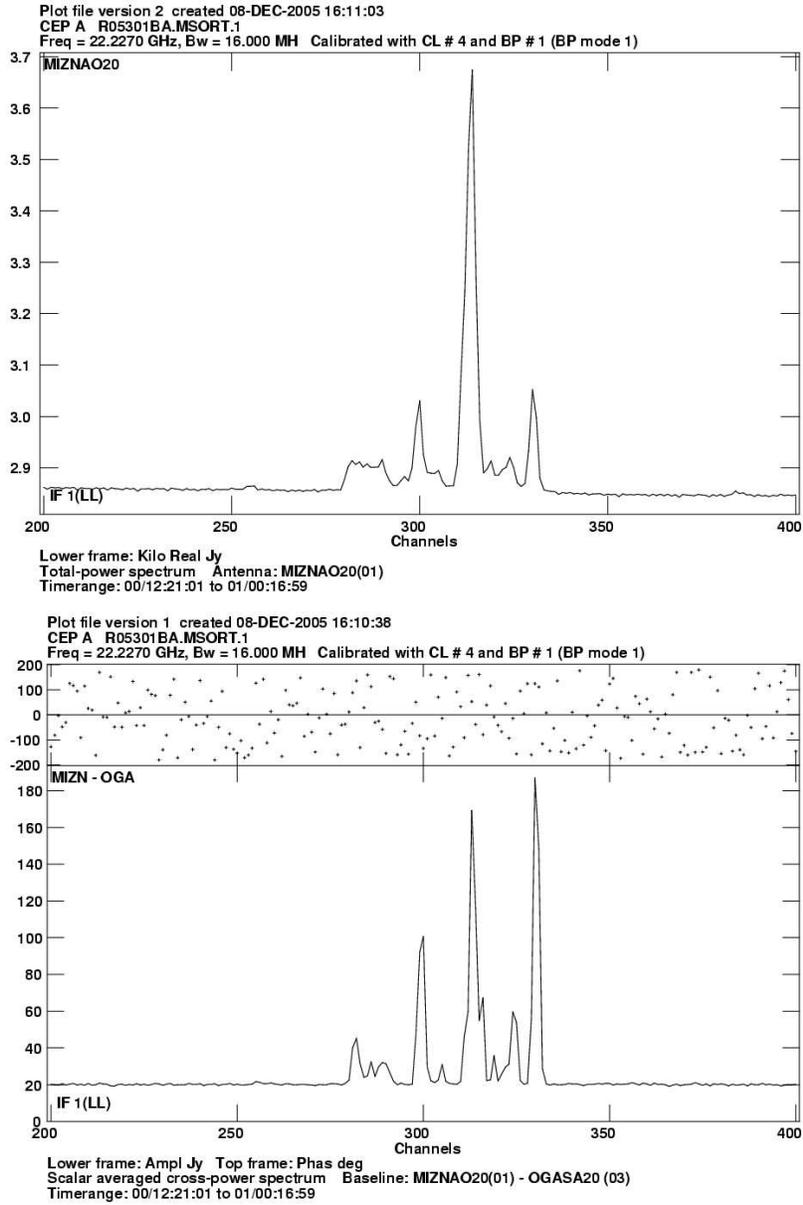
Figure 29: Spectra of the water maser source Cep A in an active region of star formation derived from 2–bit quantized data. Upper panel shows total–power spectrum with Mizusawa 20m antenna, and lower panel shows phase and amplitude of cross–power spectrum with Mizusawa–Ogasawara baseline of the VERA. Profiles of the two spectra are fairly different which indicate effects of partial resolution of maser features in the VLBI baseline. Note that amplitude scales are different in the two panels. (T. Hirota, private communication in 2005.)

75

1.2 Frequency Standard

1.2.1 VLBI Requires "Absolute" Frequency Stability
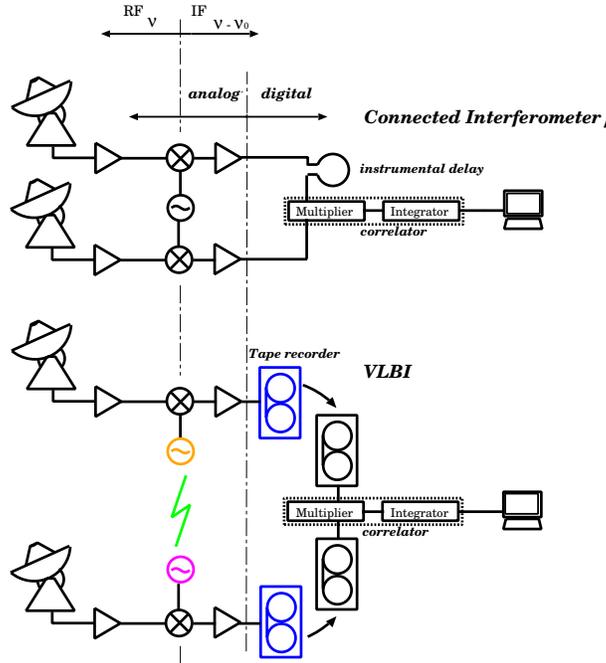
Figure 30: Connected–Element Interferometer vs. VLBI system.

Development of the highly stable frequency standard was indispensable for realization of VLBI.

In the case of the connected–element interferometer, a common frequency standard is used to generate local oscillator (LO) reference signals for frequency conversion in all antennas (see upper panel of Figure 30). Therefore, any phase noise, due to instability in the frequency standard, is common in the signals from all antennas. Such common phase noise from any given antenna is always compensated by the same phase noise from another antenna, in the correlation processing.

In fact, let us assume that a signal from a radio source, which we assume to be point–like for simplicity, received by antenna A is:

$$v_A(t) \propto \sin[\omega t - \phi(t)],$$

and the same signal received by antenna B is:

$$v_B(T) \propto \sin[\omega(t - \tau_g) - \phi(t)],$$

76

where ω is a frequency of the radio signal, τ_g is the geometric delay between the two antennas, and $\phi(t)$ is the **common** phase noise due to the frequency standard.

After multiplication of these signals, we have

$$v_A(t)\, v_B(t) \propto \frac{1}{2}\{\cos(\omega\tau_g) - \cos[2\,\omega t - \omega\tau_g - 2\,\phi(t)]\},$$

and after integration (averaging), we have

$$\langle v_A(t)\, v_B(t)\rangle \propto \frac{1}{2}\cos(\omega\tau_g), \qquad (140)$$

since rapidly oscillating second term with frequency $2\,\omega$ is averaged out.

Thus, no effect of the phase noise remains in the correlator output! Consequently, **correlation results in connected–element interferometers are almost unaffected by the instability of the frequency standard**. Then, we readily obtain the almost pure fringe pattern $\propto \cos(\omega\tau_g)$ in the crorrelation results as shown in equation (140), as far as we perform sufficiently long integration to suppress the thermal noise occuring in receiving systems and in the environment, and achieve high enough signal–to–noise ratio. This is why very high stability of frequency standards is not necessarily required in connected–element interferometers.

How about VLBI, then?

Each antenna in VLBI uses its own independent frequency standard to generate the LO reference signal (see lower panel of Figure 30).

Then, a signal from a point–like radio source received by antenna A is given by
$$v_A(t) \propto \sin[\omega t - \phi_A(t)],$$
and the same signal received by antenna B is:

$$v_B(t) \propto \sin[\omega(t - \tau_g) - \phi_B(t)],$$

where $\phi_A(t)$ and $\phi_B(t)$ are phase noises in **independent** frequency standards.

After multiplication, we have

$$v_A(t)\, v_B(t) \propto \frac{1}{2}\{\cos[\omega\tau_g - \phi_A(t) + \phi_B(t)] - \cos[2\,\omega t - \omega\tau_g - \phi_A(t) - \phi_B(t)]\},$$

and after integration (averaging),

$$\langle v_A(t)\, v_B(t)\rangle \propto \frac{1}{2}\cos[\omega\tau_g - \phi_A(t) + \phi_B(t)]. \qquad (141)$$

Therefore, no compensation of the phase noises is expected in the correlator output of VLBI. In other words, **correlation results in VLBI are always directly affected by the instability of the frequency standards!**

The noise in the fringe phase (the argument of the sinusoidal fringe pattern in equation (141)) gives rise to two difficulties in VLBI.

First, the fringe phase, which is the important observable in radio interferometry, is contaminated by the phase noise.

Second, the phase noise severely limits the sensitivity of VLBI. Indeed, it becomes impossible for us to completely stop the oscillation of the sinusoidal fringe pattern if the phase noise varies in time, even when we ideally compensate the geometric delay τ_g by applying accurate enough delay tracking and fringe stopping. Then, if we further integrate (time–average) the correlator output, hoping to get higher signal–to–noise ratio, the phase–noise–induced oscillation of the fringe pattern results in smaller amplitude of the averaged signal. The thermal noise contribution in the correlator output must be surely suppressed by the integration, but the averaged signal itself could be reduced even more rapidly due to the oscillation. Depending on the ratio between the integration time and the timescale of the oscillation, the integration may not improve the signal–to–noise ratio at all, but even degrade it. Such an effect is called the "coherence loss". Therefore, the integration time must be short enough not to cause large coherence loss, but this implies that the signal–to–noise ratio must be limited by the short integration time.

This is why we need "absolute" stability of frequency standards in VLBI, in order to ease these difficulties.

Of cource, not only the noise due to the frequency standards, but any other phase noise, due for example to the propagation delay through the turbulent atmosphere, causes similar difficulties, as we will see later.

Now we would like to discuss the way to quantitatively describe the frequency stability, and to estimate effects of the incomplete stability.

1.2.2 How to Describe Frequency Stability?

Let us consider a reference signal $v(t)$ which is expected to have a form in an ideal case:

$$v(t) = v_0 \cos(\omega_0 t), \tag{142}$$

with a nominal frequency ω_0. In actuality, however, any real reference signal has a phase noise $\phi(t)$, and therefore has a form:

$$v(t) = v_0 \cos(\omega_0 t + \phi(t)). \tag{143}$$

In this actual case, instantaneous frequency $\omega_a(t)$ will be

$$\omega_a(t) = \omega_0 + \frac{d\phi(t)}{dt}. \tag{144}$$

We then introduce a concept of the "**fractional frequency deviation (FFD)**" $y(t)$, which is defined by

$$y(t) = \frac{\delta\omega(t)}{\omega_0} = \frac{\omega_a(t) - \omega_0}{\omega_0} = \frac{1}{\omega_0}\frac{d\phi(t)}{dt}, \tag{145}$$

as a measure of the frequency stability.

Let us assume that $\phi(t)$ and $y(t)$ are stationary random processes. Then, their autocorrelations are functions of the time difference τ:

$$R_{\phi\phi}(\tau) = \langle \phi(t+\tau)\,\phi(t)\rangle, \quad \text{and} \quad R_{yy}(\tau) = \langle y(t+\tau)\,y(t)\rangle, \tag{146}$$

and their power spectra are given by Fourier transform relations:

$$S_{\phi\phi}(\omega) = \int_{-\infty}^{\infty} R_{\phi\phi}(\tau)\,e^{-i\omega\tau}\,d\tau, \quad R_{\phi\phi}(\tau) = \frac{1}{2\pi}\int_{-\infty}^{\infty} S_{\phi\phi}(\omega)\,e^{i\omega\tau}\,d\omega,$$

$$S_{yy}(\omega) = \int_{-\infty}^{\infty} R_{yy}(\tau)\,e^{-i\omega\tau}\,d\tau, \quad R_{yy}(\tau) = \frac{1}{2\pi}\int_{-\infty}^{\infty} S_{yy}(\omega)\,e^{i\omega\tau}\,d\omega. \tag{147}$$

In view of equation (145), the autocorrelations of $y(t)$ and $\phi(t)$ are mutually related by

$$R_{yy}(t,t') = \langle y(t)\,y(t')\rangle = \frac{1}{\omega_0^2}\left\langle \frac{d\phi(t)}{dt}\frac{d\phi(t')}{dt'}\right\rangle = \frac{1}{\omega_0^2}\frac{\partial^2}{\partial t\,\partial t'}R_{\phi\phi}(t,t'). \tag{148}$$

In our case of the stationary random processes, this is reduced to

$$R_{yy}(\tau) = -\frac{1}{\omega_0^2}\frac{d^2}{d\tau^2}R_{\phi\phi}(\tau), \tag{149}$$

where $\tau = t - t'$.

From equations (147) and (149), we have

$$R_{yy}(\tau) = -\frac{1}{\omega_0^2}\frac{d^2}{d\tau^2}\frac{1}{2\pi}\int_{-\infty}^{\infty} S_{\phi\phi}(\omega)\,e^{i\omega\tau}\,d\omega = \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{\omega^2}{\omega_0^2} S_{\phi\phi}(\omega)\,e^{i\omega\tau}\,d\omega,$$

and, therefore, the power spectra of $y(t)$ and $\phi(t)$ are related to each other by a relation:

$$S_{yy}(\omega) = \frac{\omega^2}{\omega_0^2} S_{\phi\phi}(\omega). \tag{150}$$

Let us make a comment here about conventions which have been traditionally used for describing power spectra in frequency stability discussions.

Since $y(t)$ and $\phi(t)$ are real functions of time, the autocorrelation $R_{\phi\phi}(\tau)$ and $R_{yy}(\tau)$, and power spectra $S_{\phi\phi}(\omega)$, $S_{yy}(\omega)$ are all real and even functions of τ and ω. Using this property, and using frequency ν instead of angular frequency $\omega = 2\pi\,\nu$, we can describe the power spectra in the "single–sided forms" $\mathsf{S}_{\phi\phi}(\nu)$ and $\mathsf{S}_{yy}(\nu)$, which have been widely used in the frequency stability discussions:

$$\mathsf{S}_{\phi\phi}(\nu) = 4 \int\limits_0^\infty R_{\phi\phi}(\tau)\,\cos(2\pi\,\nu\,\tau)\,d\tau, \quad R_{\phi\phi}(\tau) = \int\limits_0^\infty \mathsf{S}_{\phi\phi}(\nu)\,\cos(2\pi\,\nu\,\tau)\,d\nu,$$

$$\mathsf{S}_{yy}(\nu) = 4 \int\limits_0^\infty R_{yy}(\tau)\,\cos(2\pi\,\nu\,\tau)\,d\tau, \quad R_{yy}(\tau) = \int\limits_0^\infty \mathsf{S}_{yy}(\nu)\,\cos(2\pi\,\nu\,\tau)\,d\nu.$$

$$(151)$$

These single–sided spectra $\mathsf{S}_{\phi\phi}(\nu)$ and $\mathsf{S}_{yy}(\nu)$ are related to our double–sided power spectra $S_{\phi\phi}(\omega)$ and $S_{yy}(\omega)$ by:

$$\mathsf{S}_{\phi\phi}(\nu) = 2\,S_{\phi\phi}(2\pi\nu), \quad \text{and} \quad \mathsf{S}_{yy}(\nu) = 2\,S_{yy}(2\pi\nu), \tag{152}$$

for the positive frequency range $\nu \geq 0$. Therefore, their mutual relationship has been often given by:

$$\mathsf{S}_{yy}(\nu) = \frac{\nu^2}{\nu_0^2}\mathsf{S}_{\phi\phi}(\nu), \tag{153}$$

instead of equation (150), where $\nu_0 = \dfrac{\omega_0}{2\pi}$.

However, we will continue to use our double–sided power spectrum forms in equations (147), following our previous discussions.

1.2.3 Types of Phase and Frequency Noises

Measurements have shown that noises in the frequency stability are classified into following types according to the power–law index α of the power spectrum of the FFD $S_{yy}(\omega) = H_\alpha\,\omega^\alpha$ with roughly constant coefficients H_α. Each noise type, or power–law component of the FFD spectrum, has its own characteristic name listed in Table 4.

An oscillator in a frequency standard often shows a combination of all power–law components of Table 4 in various frequency ranges, as shown schematically in Figure 31.

α	Name of Noise Type	$S_{yy}(\omega)$	$S_{\phi\phi}(\omega)$
2	White phase	$H_2\,\omega^2$	$\omega_0^2\,H_2$
1	Flicker phase	$H_1\,\omega^1$	$\omega_0^2\,H_1\,\omega^{-1}$
0	White frequency	H_0	$\omega_0^2\,H_0\,\omega^{-2}$
-1	Flicker frequency	$H_{-1}\,\omega^{-1}$	$\omega_0^2\,H_{-1}\,\omega^{-3}$
-2	Random walk of frequency	$H_{-2}\,\omega^{-2}$	$\omega_0^2\,H_{-2}\,\omega^{-4}$

Table 4: Power–law types of phase and frequency noise spectra.
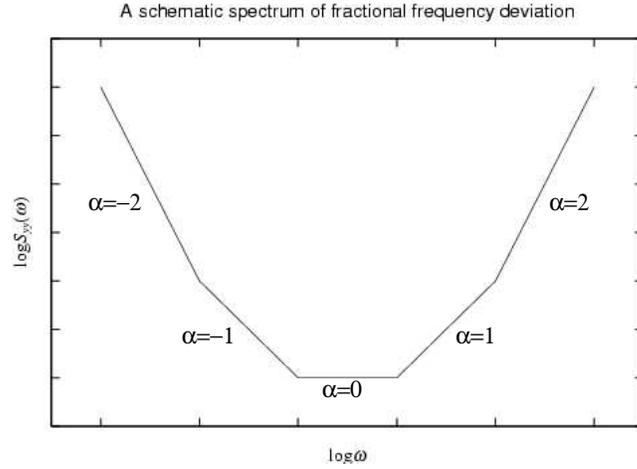
Figure 31: A schematic spectrum of the FFD, $y(t)$, of an oscillator showing all power–law components in Table 4 in various frequency ranges. Horizontal and vertical axes show the angular frequency ω and the power spectrum $S_{yy}(\omega)$, respectively, in log–scales.

81

1.2.4 Time Domain Measurements

It has traditionally been easier to make measurements in the time domain, than in the frequency domain. Hence, frequency stability characteristics are usually given in terms of the time domain measurements.

Suppose that we have N phase values: $\phi(t_1)$, $\phi(t_2)$, $\cdots$, $\phi(t_k)$, $\cdots$, $\phi(t_N)$, measured at equally spaced time points: t_1, t_2, $\cdots$, t_k, $\cdots$, t_N, with time interval $\mathcal{T}$, where $\mathcal{T} = t_{k+1} - t_k$ for any k.

Then, using these phase values measured in the time domain, we form a discrete time series of the "**mean fractional frequency deviation**" $\bar{y}[k]$ which is defined by

$$\bar{y}[k] = \frac{\phi(t_{k+1}) - \phi(t_k)}{\omega_0\,\mathcal{T}} = \frac{1}{\omega_0\,\mathcal{T}} \int\limits_{t_k}^{t_{k+1}} \frac{d\phi(t')}{dt'}\,dt' = \frac{1}{\mathcal{T}} \int\limits_{t_k}^{t_{k+1}} y(t')dt', \qquad (154)$$

and can be described also as

$$\bar{y}[k] = \frac{1}{\mathcal{T}} \int\limits_{t-\frac{\mathcal{T}}{2}}^{t+\frac{\mathcal{T}}{2}} y(t')dt', \qquad (155)$$

where $t = t_k + \mathcal{T}/2$. This is a running mean of the FFD $y(t)$ at the point $t = t_k + \mathcal{T}/2$ over the time interval $\mathcal{T}$. Thus, $\bar{y}[k]$ for any k is given by a linear system of $y(t)$:

$$\bar{y}[k] = y(t) * a(t) = \int\limits_{-\infty}^{\infty} y(t - t')\,a(t')\,dt', \qquad (156)$$

with an impulse response $a(t)$:

$$a(t) = \begin{cases} \frac{1}{\mathcal{T}} & \text{if } -\frac{\mathcal{T}}{2} < t \leq \frac{\mathcal{T}}{2}, \\ 0 & \text{otherwise,} \end{cases} \qquad (157)$$

at $t = t_k + \mathcal{T}/2$, similarly to what we saw in Chapter 3.

1.2.5 "True Variance" and "Allan Variance" of Fractional Frequency Deviation

Using the mean fractional frequency deviation $\bar{y}[k]$, which is obtained by averaging the FFD $y(t)$ for the time interval $\mathcal{T}$, as shown in equation (155),

we introduce the "**true variance of the fractional frequency deviation (TVAR)**" as a function of the time interval $\mathcal{T}$:

$$I^2(\mathcal{T}) = \langle \bar{y}^2[k] \rangle, \tag{158}$$

where, we assumed zero–mean of the mean FFD $\bar{y}[k]$:

$$\langle \bar{y}[k] \rangle = 0.$$

In view of the ergodicity, we can estimate the TVAR $I^2(\mathcal{T})$ by means of a time–average Σ_t of $N-1$ values of the squared mean FFD $\bar{y}^2[k]$:

$$\Sigma_t = \frac{1}{N-1} \sum_{k=1}^{N-1} \bar{y}^2[k]. \tag{159}$$

Generally speaking, there seems no difficulty in calculating the time–average Σ_t, and, therefore, this estimated TVAR appears a good measure of the frequency stability. In actuality, however, Σ_t, as given in equation (159), diverges for some important types of frequency noises, and therefore cannot be used for characterizing the frequency stability as a whole. The assumption of the stationary random process, or the ergodicity, is not likely to be strictly fulfilled in these diverging cases.

In order to overcome this difficulty, David W. Allan proposed to take a difference of successive two samples:

$$\Delta \bar{y}[k] = \bar{y}[k+1] - \bar{y}[k], \tag{160}$$

and calculate an average of the squared difference $(\Delta \bar{y}[k])^2$ divided by two:

$$\Sigma_a = \frac{1}{2\,(N-2)} \sum_{k=1}^{N-2} (\Delta \bar{y}[k])^2, \tag{161}$$

(Allan, 1966). In view of equation (154), the two sample difference $\Delta \bar{y}[k]$ of the mean FFD is easily derived from the measured phase noise $\phi[k]$ by

$$\Delta \bar{y}[k] = \frac{\phi(t_{k+2}) - 2\,\phi(t_{k+1}) + \phi(t_k)}{\omega_0\,\mathcal{T}}. \tag{162}$$

The mean square sum Σ_a gives an estimation of the "two–sample variance" or "**Allan variance (AVAR)**":

$$\sigma_y^2(\mathcal{T}) = \frac{\langle (\Delta \bar{y}[k])^2 \rangle}{2}, \tag{163}$$

83
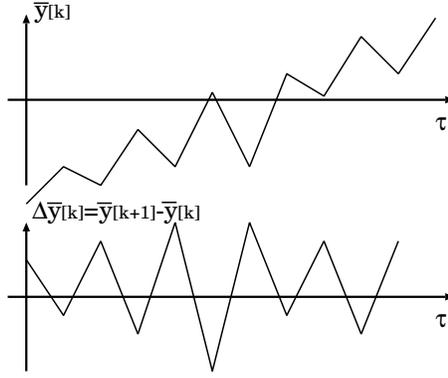
Figure 32: Mean square sum of the two–sample difference of the mean FFD (lower panel) has a better chance to converge than that of the mean FFD itself (upper panel), even when the mean FFD shows a diverging long–time–scale behaviour.

as a function of the time interval $\mathcal{T}$.

Estimation of the AVAR with Σ_a has a better chance to converge than that of the TVAR with Σ_t, as schematically illustrated in Figure 32. Therefore, the estimated AVAR is widely accepted as a good measure of the frequency stability.

In view of equation (154), we can describe the two–sample difference of the mean FFD $\Delta \bar{y}[k]$ through time integrations of the FFD $y(t)$:

$$\Delta \bar{y}[k] = \bar{y}[k+1] - \bar{y}[k] = \frac{1}{\mathcal{T}} \left\{ \int_{t_{k+1}}^{t_{k+2}} y(t') \, dt' - \int_{t_k}^{t_{k+1}} y(t') \, dt' \right\}$$

$$= \frac{1}{\mathcal{T}} \left\{ \int_{t}^{t+\mathcal{T}} y(t') \, dt' - \int_{t-\mathcal{T}}^{t} y(t') \, dt' \right\}, \tag{164}$$

where $t = t_{k+1}$. This is nothing but a linear system of the FFD $y(t)$:

$$\Delta \bar{y}[k] = y(t) * b(t) = \int_{-\infty}^{\infty} y(t - t') \, b(t') \, dt', \tag{165}$$

with an impulse response:

$$b(t) = \begin{cases} \frac{1}{\mathcal{T}} & \text{if } -\mathcal{T} < t \le 0, \\ -\frac{1}{\mathcal{T}} & \text{if } 0 < t \le \mathcal{T}, \\ 0 & \text{otherwise,} \end{cases} \tag{166}$$

84

at time $t = t_{k+1}$.

1.2.6 True Variance and Allan Variance through Power Spectrum of Fractional Frequency Deviation

Equations (156) and (165) describe the discrete–time series $\bar{y}[k]$ (the meam FFD) and $\Delta\bar{y}[k]$ (the two–sample difference of the mean FFD) through the linear systems with impulse responses $a(t)$ and $b(t)$ at times $t = t_k + \mathcal{T}/2$ and $t = t_{k+1}$, respectively. We can formally extend these linear systems to yield continuous–time functions $\bar{y}(t)$ and $\Delta\bar{y}(t)$:

$$\bar{y}(t) = y(t) * a(t) = \int\limits_{-\infty}^{\infty} y(t - t') \, a(t') \, dt',$$

$$\Delta\bar{y}(t) = y(t) * b(t) = \int\limits_{-\infty}^{\infty} y(t - t') \, b(t') \, dt', \tag{167}$$

at arbitrary time t, and introduce their autocorrelations

$$R_{\bar{y}\bar{y}}(\tau) = \langle \bar{y}(t + \tau) \, \bar{y}(t) \rangle,$$
$$R_{\Delta\bar{y}\Delta\bar{y}}(\tau) = \langle \Delta\bar{y}(t + \tau) \, \Delta\bar{y}(t) \rangle, \tag{168}$$

and power spectra $S_{\bar{y}\bar{y}}(\omega)$ and $S_{\Delta\bar{y}\Delta\bar{y}}(\omega)$. Of course, $\bar{y}(t)$ takes the particular value $\bar{y}[k]$ at $t = t_k + \mathcal{T}/2$, and $\Delta\bar{y}(t)$ takes the particular value of $\Delta\bar{y}[k]$ at $t = t_{k+1}$, i.e.

$$\bar{y}(t) = \bar{y}[k] \quad \text{at} \ \ t = t_k + \frac{\mathcal{T}}{2},$$
$$\Delta\bar{y}(t) = \Delta\bar{y}[k] \quad \text{at} \ \ t = t_{k+1}. \tag{169}$$

Then, we can describe the TVAR, $I^2(\mathcal{T})$, through the autocorrelation $R_{\bar{y}\bar{y}}(\tau)$, and then the power spectrum $S_{\bar{y}\bar{y}}(\omega)$:

$$I^2(\mathcal{T}) = \langle (\bar{y}[k])^2 \rangle = R_{\bar{y}\bar{y}}(0) = \frac{1}{2\pi} \int\limits_{-\infty}^{\infty} S_{\bar{y}\bar{y}}(\omega) \, d\omega. \tag{170}$$

Similarly, for the AVAR, $\sigma_y^2(\mathcal{T})$, we have

$$\sigma_y^2(\mathcal{T}) = \frac{\langle (\Delta\bar{y}[k])^2 \rangle}{2} = \frac{R_{\Delta\bar{y}\Delta\bar{y}}(0)}{2} = \frac{1}{4\pi} \int\limits_{-\infty}^{\infty} S_{\Delta\bar{y}\Delta\bar{y}}(\omega) \, d\omega. \tag{171}$$

Since the functions $\bar{y}(t)$ and $\Delta\bar{y}(t)$ are related to the FFD $y(t)$ through the linear system equations (167), we can describe the power spectra $S_{\bar{y}\bar{y}}(\omega)$

85

and $S_{\Delta\bar{y}\Delta\bar{y}}(\omega)$ through the power spectrum $S_{yy}(\omega)$ of the FFD $y(t)$, which we introduced in equation (147). For this purpose, let us introduce system functions $\mathsf{A}(\omega)$ and $\mathsf{B}(\omega)$ of the impulse responses $a(t)$ and $b(t)$, respectively:

$$\mathsf{A}(\omega) = \int\limits_{-\infty}^{\infty} a(t)\, e^{-i\omega t}\, dt = \frac{1}{\mathcal{T}} \int\limits_{-\frac{\mathcal{T}}{2}}^{\frac{\mathcal{T}}{2}} e^{-i\omega t}\, dt = \frac{\sin(\frac{\omega\mathcal{T}}{2})}{(\frac{\omega\mathcal{T}}{2})},$$

$$\mathsf{B}(\omega) = \int\limits_{-\infty}^{\infty} b(t)\, e^{-i\omega t}\, dt = \frac{1}{\mathcal{T}} \left\{ \int\limits_{-\mathcal{T}}^{0} e^{-i\omega t}\, dt - \int\limits_{0}^{\mathcal{T}} e^{-i\omega t}\, dt \right\} = 2\, i\, \frac{\sin^2(\frac{\omega\mathcal{T}}{2})}{(\frac{\omega\mathcal{T}}{2})}.$$

$$(172)$$

From general properties of linear systems, we have

$$S_{\bar{y}\bar{y}}(\omega) = S_{yy}(\omega)\, |\, \mathsf{A}(\omega)\,|^2,$$

and

$$S_{\Delta\bar{y}\Delta\bar{y}}(\omega) = S_{yy}(\omega)\, |\, \mathsf{B}(\omega)\,|^2 .$$

Therefore, equations (170) and (171) are reduced to
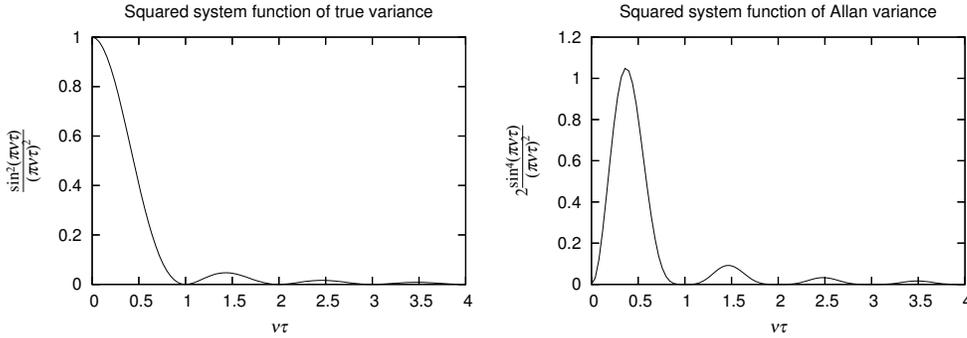
Figure 33: Squared system functions for the TVAR, $|\, \mathsf{A}(\omega)\,|^2$, (left) and the AVAR, $|\, \mathsf{B}(\omega)\,|^2$, (right). Horizontal axes show the product $\nu\mathcal{T}$ of the frequency and the time interval.

$$I^2(\mathcal{T}) = \frac{1}{2\pi} \int\limits_{-\infty}^{\infty} S_{yy}(\omega)\, \frac{\sin^2(\frac{\omega\mathcal{T}}{2})}{(\frac{\omega\mathcal{T}}{2})^2}\, d\omega, \tag{173}$$

$$\sigma_y^2(\mathcal{T}) = \frac{1}{2\pi} \int\limits_{-\infty}^{\infty} S_{yy}(\omega)\, 2\, \frac{\sin^4(\frac{\omega\mathcal{T}}{2})}{(\frac{\omega\mathcal{T}}{2})^2}\, d\omega. \tag{174}$$

These are the equations which relate the TVAR and the AVAR to the power spectrum $S_{yy}(\omega)$ of the FFD $y(t)$.

Figure 33 shows squared system functions $\mid \mathsf{A}(\omega) \mid^2$ and $\mid \mathsf{B}(\omega) \mid^2$ for the TVAR and the AVAR, respectively. It is evident from this Figure that the low frequency noise is effectively suppressed in the AVAR.

1.2.7 Time–Interval Dependence of Allan Variance

We can calculate dependence of the TVAR and the AVAR on the time interval $\mathcal{T}$ according to equations (173) and (174) for each power–law index α of the FFD spectrum $S_{yy}(\omega) \propto \omega^\alpha$ given in Table 4. Results are shown in Table 5, where ω_h in the white phase ($\alpha = 2$) and the flicker phase ($\alpha = 1$) components is a cut–off frequency at the high frequency side of the FFD spectrum.

Noise Type	$S_{yy}(\omega)$	$S_{\phi\phi}(\omega)$	$\sigma_y^2(\mathcal{T})$	$I^2(\mathcal{T})$
White phase	$H_2\,\omega^2$	$\omega_0^2\,H_2$	$\frac{3H_2\omega_h}{\pi\mathcal{T}^2}$	$\frac{2H_2\omega_h}{\pi\mathcal{T}^2}$
Flicker phase	$H_1\,\omega^1$	$\omega_0^2\,H_1\,\omega^{-1}$	$\frac{3H_1\ln(\omega_h\mathcal{T})}{\pi\mathcal{T}^2}$	—
White frequency	H_0	$\omega_0^2\,H_0\,\omega^{-2}$	$\frac{H_0}{\mathcal{T}}$	$\frac{H_0}{\mathcal{T}}$
Flicker frequency	$H_{-1}\,\omega^{-1}$	$\omega_0^2\,H_{-1}\,\omega^{-3}$	$\frac{(2\ln 2)\,H_{-1}}{\pi}$	—
Random walk of frequency	$H_{-2}\,\omega^{-2}$	$\omega_0^2\,H_{-2}\,\omega^{-4}$	$\frac{\mathcal{T}H_{-2}}{3}$	—

Table 5: Time–interval dependence of the AVAR, $\sigma_y^2(\mathcal{T})$, and the TVAR, $I^2(\mathcal{T})$, for each power–law component of the FFD spectrum. ω_h is a cut–off frequency at the high frequency side of a power–law spectrum $S_{yy}(\omega)$.

Figure 34 schematically shows the time–interval dependence of the "Allan standard deviation (ASD)", $\sigma_y(\mathcal{T})$, which is defined as the square root of the AVAR. It is clear from this Figure that, having measured ASD as a function of the time interval between samples $\mathcal{T}$, we can readily distinguish noise types, except for the case of the white phase and the flicker phase. This demonstrates the real usefulness of the ASD (or the AVAR) as a measure of the frequency stability.

Figure 35 shows characteristic performance of various frequency standards in terms of the ASD versus the time interval. ASD's of frequency standards are approximated by different noise types in different time–interval ranges. The active hydrogen maser, which shows ASD $\leq 10^{-15}$ in its flicker frequency regime, exhibits the highest frequency stability in the time–interval range of 1,000 to 10,000 seconds.
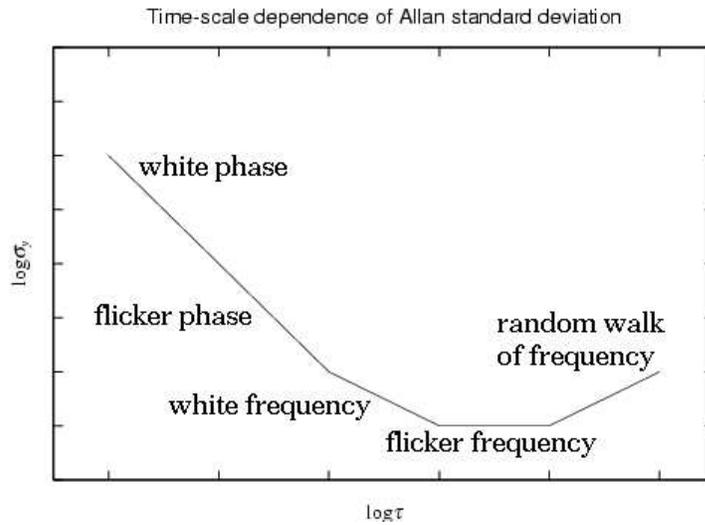
Figure 34: A shematic view of the time–interval $\mathcal{T}$ dependence of the Allan standard deviation $\sigma_y(\mathcal{T})$ in log–scales. Each noise type shows its own gradient in this diagram, though lines of the white phase and the flicker phase are hardly distinguished by the difference of their gradients.
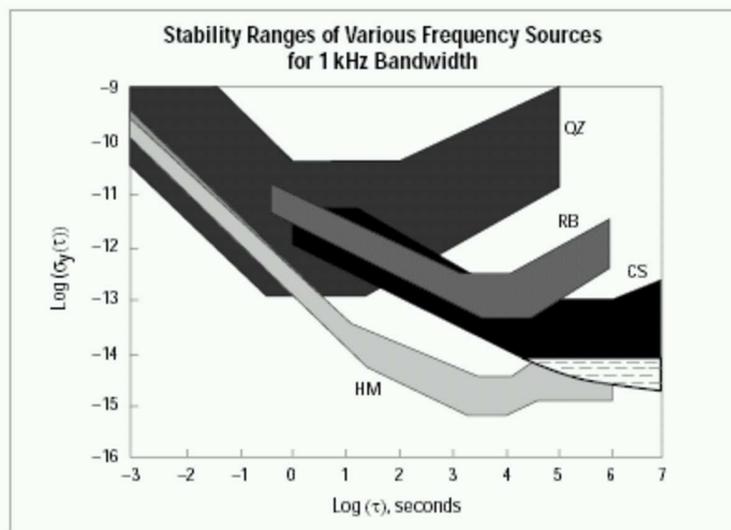
Figure 35: Performance of various frequency standards in terms of the ASD— the time interval between samples. QZ: Quartz, RB: Rubidium, CS: Cesium, HM: Active Hydrogen Maser. (Figure courtesy of HEWLETT PACKARD, Application Note 1289).

1.2.8 True Variance and Allan Variance through Autocorrelation of Phase Noise

We can describe the TVAR and the AVAR also through the autocorrelation $R_{\phi\phi}(\mathcal{T})$ of the phase noise $\phi(t)$. In fact, for the TVAR, equations (154) and (158) yield

$$I^2(\mathcal{T}) = \langle \bar{y}^2[k] \rangle = \frac{\langle [\phi(t_k + \mathcal{T}) - \phi(t_k)]^2 \rangle}{(\omega_0 \mathcal{T})^2} = \frac{2\,[R_{\phi\phi}(0) - R_{\phi\phi}(\mathcal{T})]}{(\omega_0 \mathcal{T})^2}. \quad (175)$$

Also, for the AVAR, equations (162) and (163) lead to

$$\sigma_y^2(\mathcal{T}) = \frac{\langle (\Delta \bar{y}[k])^2 \rangle}{2} = \frac{\langle [\phi(t_k + 2\mathcal{T}) - 2\,\phi(t_k + \mathcal{T}) + \phi(t_k)]^2 \rangle}{2\,(\omega_0 \mathcal{T})^2}$$

$$= \frac{3\,R_{\phi\phi}(0) - 4\,R_{\phi\phi}(\mathcal{T}) + R_{\phi\phi}(2\mathcal{T})}{(\omega_0 \mathcal{T})^2}. \quad (176)$$

Now, from equation (175), we have,

$$I^2(\mathcal{T}) - I^2(2\mathcal{T}) = \frac{2\,[R(0) - R(\mathcal{T})] - \dfrac{1}{2}\,[R(0) - R(2\mathcal{T})]}{(\omega_0 \mathcal{T})^2}$$

$$= \frac{3\,R_{\phi\phi}(0) - 4\,R_{\phi\phi}(\mathcal{T}) + R_{\phi\phi}(2\mathcal{T})}{2\,(\omega_0 \mathcal{T})^2}. \quad (177)$$

Therefore, in cases when the TVAR $I^2(\mathcal{T})$ does not diverge, we have the following relationship between the AVAR and the TVAR:

$$\sigma_y^2(\mathcal{T}) = 2\,[I^2(\mathcal{T}) - I^2(2\mathcal{T})]. \quad (178)$$

1.2.9 Coherence Function

As we saw earlier, the phase noise in the fringe phase causes the serious problem, the coherence loss, for the sensitivity of VLBI.

The coherence loss due to the phase noise $\phi(t)$, after integrating the correlator output for duration T, can be estimated by introducing the "**coherence function**" (Rogers and Moran, 1981), which is defined by

$$C(T) = \left| \frac{1}{T} \int_0^T e^{i\phi(t)}\, dt \right|. \quad (179)$$

Magnitude of the squared coherence function is represented by its dispersion:

$$\langle C^2(T) \rangle = \frac{1}{T^2} \int_0^T \int_0^T \langle e^{i[\phi(t) - \phi(t')]} \rangle\, dt\, dt'. \quad (180)$$

Now, if we assume the Gaussian distribution of the phase noise difference $\Phi = \phi(t) - \phi(t')$:

$$f(\Phi) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{\Phi^2}{2\sigma^2}}, \tag{181}$$

where

$$\sigma^2 = \langle \Phi^2 \rangle = \langle [\phi(t) - \phi(t')]^2 \rangle,$$

then, using the formula

$$\int_{-\infty}^{\infty} e^{-x^2 - iax}\, dx = \sqrt{\pi}\, e^{-\frac{a^2}{4}},$$

we obtain

$$\langle e^{i\Phi} \rangle = \int_{-\infty}^{\infty} f(\Phi)\, e^{i\Phi}\, d\Phi = e^{-\frac{\sigma^2}{2}} = e^{-\frac{\langle \Phi^2 \rangle}{2}}.$$

Hence, we reduce equation (180) to

$$\langle C^2(T) \rangle = \frac{1}{T^2} \int_0^T \int_0^T e^{-\frac{\langle [\phi(t) - \phi(t')]^2 \rangle}{2}}\, dt\, dt' = \frac{1}{T^2} \int_0^T \int_0^T e^{-\frac{D(t,\,t')}{2}}\, dt\, dt', \tag{182}$$

where $D(t,\, t')$ is the so–called "**temporal structure function**" defined by

$$D(t,\, t') = \langle [\phi(t) - \phi(t')]^2 \rangle. \tag{183}$$

Under the assumption of the stationary random phase noise, we have

$$D(t,\, t') = D(\tau),$$

with $\tau = t - t'$. Then we can reduce the double integral in equation (182) to a single integral. In fact, noting that $dz = \frac{1}{\sqrt{2}}\, d\tau$ and $L = \sqrt{2}\,(T - \tau)$ in Figure 36, we obtain

$$\langle C^2(T) \rangle = \frac{2}{T^2} \int_0^T (T - \tau)\, e^{-\frac{D(\tau)}{2}}\, d\tau = \frac{2}{T} \int_0^T \left(1 - \frac{\tau}{T}\right) e^{-\frac{D(\tau)}{2}}\, d\tau. \tag{184}$$

According to equation (154), the mean FFD $\bar{y}[k]$ with the sample interval τ is

$$\bar{y}[k] = \frac{\phi(t_k + \tau) - \phi(t_k)}{\omega_0\, \tau}.$$

Therefore, we can describe the structure function $D(\tau)$ through the TVAR $I^2(\tau)$ as

$$D(\tau) = \omega_0^2\, \tau^2\, \langle \bar{y}^2 \rangle = \omega_0^2\, \tau^2\, I^2(\tau). \tag{185}$$

90
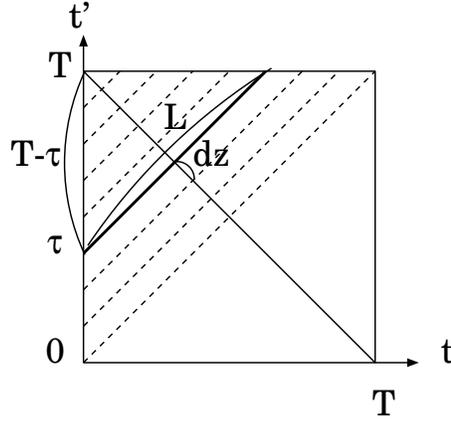
Figure 36: Geometry of integration in equation (182).

This enable us to describe the dispersion of the coherence function through the TVAR:

$$\langle C^2(T) \rangle = \frac{2}{T} \int_0^T \left(1 - \frac{\tau}{T}\right) e^{-\frac{\omega_0^2 \tau^2 I^2(\tau)}{2}} \, d\tau$$

$$= \frac{2}{T} \int_0^T \left(1 - \frac{\tau}{T}\right) e^{-\frac{\omega_0^2 \tau^2}{4\pi} \int_{-\infty}^{\infty} S_{yy}(\omega) \frac{\sin^2(\frac{\omega \tau}{2})}{(\frac{\omega \tau}{2})^2} \, d\omega} \, d\tau. \qquad (186)$$

Thus, when a functional form of the TVAR $I^2(\tau)$ is given, we can calculate the dispersion of the coherence function $\langle C^2(T) \rangle$. Then, we can estimate the "**coherence time** τ_c", i.e. the interval during which we can more or less coherently integrate our signal. As the criterion for the coherence time τ_c, we usually adopt the time interval which gives $\langle C^2(\tau_c) \rangle \sim 0.85$.

1.2.10   Approximate Estimation of Coherence Time

Precise estimation of the coherence time using equation (186) is sometimes impractical when the estimation of the TVAR with equation (159) diverses.

For practical purposes, the coherence time is usually estimated by a simpler way.

Phase noise accumulated during a time T is approximately given by $\simeq \omega_0 \sigma_y T$, where σ_y is "some standard deviation" of the FFD, for which we usually adopt the ASD.

It is obvious that we will not obtain any meaningful correlation result if the phase noise varies more than 2π, during an integration time T. It is the

91

usual practice to require that the accumulated phase noise must not exceed 1 radian:

$$\omega_0 \sigma_y(T) T \leq 1 \ \text{radian.} \tag{187}$$

(Note that σ_y is, in general, a function of the time interval T.) Thus, the coherence time τ_c is estimated as a time which satisfies

$$\omega_0 \sigma_y(\tau_c) \tau_c \simeq 1. \tag{188}$$

If we observe at $\nu = 8$ GHz, and require that $\tau_c = 1000$ sec, say, we need a frequency stability better than

$$\sigma_y(1000 \ \text{sec}) \simeq \frac{1}{2\pi \times 8 \times 10^9 \times 10^3} \simeq 2 \times 10^{-14},$$

which is sometimes described as the **"stability of a clock which would deviate by 1 sec in 5×10^{13} sec, or 1.6 million years!"**

The Active Hydrogen Maser Frequency Standard, which has the frequency stability of $10^{-16} < \sigma_y < 10^{-14}$ at time scales around 1000 sec (see Figure 35), fulfilled the requirement, and is widely used in the world VLBI observations.

1.2.11 Estimation of Time–Averaged Phase Noise

In some applications of VLBI obserrvation, for example in VLBI astrometry, it is meaningful to theoretically estimate the phase noise expected after time–averaging of the correlator output for a duration of time T.

If we have a time series of measured fringe phase with high enough signal–to–noise ratio, the time–averaged phase noise $\bar{\phi}$ is given by a simple model:

$$\bar{\phi} = \frac{1}{T} \int_0^T \phi(t) \, dt, \tag{189}$$

where $\phi(t)$ is the phase noise at time t. This model is good enough for the case of the "vector averaging" of the correlator output, such as shown in equation (179) for the coherence function, as long as the phase noise is kept well smaller than 1 radian. In fact, if $\phi(t) \ll 1$, we have

$$\frac{1}{T} \int_0^T e^{i\phi(t)} \, dt \cong \frac{1}{T} \int_0^T (1 + i\,\phi(t)) \, dt \cong e^{i\bar{\phi}}.$$

Now, we will derive formulae for the dispersion of the residual phase noise from its time average:

$$\Delta\phi = \phi(t) - \bar{\phi} = \phi(t) - \frac{1}{T} \int_0^T \phi(t') dt', \tag{190}$$

92

and that of the time-averaged phase itself, given by equation (189).

The dispersions are

$$
\sigma_{\Delta\phi}^2(T) = \left\langle \frac{1}{T} \int_0^T \left[\phi(t) - \frac{1}{T} \int_0^T \phi(t')dt' \right]^2 dt \right\rangle
$$

$$
= \frac{2}{T^2} \int_0^T (T - \tau)[R_{\phi\phi}(0) - R_{\phi\phi}(\tau)]d\tau = \frac{1}{T^2} \int_0^T (T - \tau)\,\omega_0^2\,\tau^2\,I^2(\tau)d\tau.
$$

(191)

$$
\sigma_{\bar\phi}^2(T) = \left\langle \left[\frac{1}{T} \int_0^T \phi(t)dt \right]^2 \right\rangle = \frac{2}{T^2} \int_0^T (T - \tau)R_{\phi\phi}(\tau)d\tau.
$$

(192)

Therefore, we can calculate these dispersions if we know the TVAR $I^2(\tau)$ and the autocorrelation of the phase noise $R_{\phi\phi}(\tau)$. Note that equation (191) describes the accumulated phase noise around its time average. In particular, equation (186) can be reduced to

$$
\langle C^2(T) \rangle = \frac{2}{T} \int_0^T \left(1 - \frac{\tau}{T} \right) e^{-\frac{\omega_0^2\,\tau^2\,I^2(\tau)}{2}} d\tau
$$

$$
\cong \frac{2}{T} \int_0^T \left(1 - \frac{\tau}{T} \right) [1 - \frac{\omega_0^2\,\tau^2\,I^2(\tau)}{2}]d\tau = 1 - \sigma_{\Delta\phi}^2(T), \quad (193)
$$

as long as $\omega_0^2\,\tau^2\,I^2(\tau)\,/\,2 \ll 1$.

In the simplest case of the white phase noise, we have

$$
I^2(\tau) = \frac{2\,H_2\,\omega_h}{\pi\,\tau^2}, \quad \text{and} \quad \sigma_y^2(\tau) = \frac{3\,H_2\,\omega_h}{\pi\,\tau^2}, \tag{194}
$$

for the TVAR and the AVAR, and

$$
S_{\phi\phi}(\omega) = \omega_0^2\,H_2, \quad \text{and} \quad R_{\phi\phi}(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_{\phi\phi}(\omega)\,e^{i\omega\tau}\,d\omega = \omega_0^2\,H_2\,\delta(\tau),
$$

(195)

for the power spectrum and the autocorrelation of the phase noise. Then, equations (191) and (192) yield

$$
\sigma_{\Delta\phi}^2(T) = \frac{1}{T^2} \int_0^T (T - \tau)\,\omega_0^2\,\tau^2\,I^2(\tau)d\tau = \frac{H_2\,\omega_0^2\,\omega_h}{\pi}, \tag{196}
$$

for the dispersion of the residual phase noise, and

$$
\sigma_{\bar\phi}^2(T) = \frac{2}{T^2} \int_0^T (T - \tau)R_{\phi\phi}(\tau)d\tau = \frac{H_2\,\omega_0^2}{T}, \tag{197}
$$

93

for the dispersion of the time–averaged phase noise.

It is evident that, for the white phase noise, the dispersion of the residual phase noise $\sigma^2_{\Delta\phi}(T)$ does not depend on the averaging time T. This dispersion is readily estimated, if we have measured AVAR $\sigma^2_y(\tau)$, from which we can easily extract the coefficient $H_2\,\omega_h$.

On the other hand, the dispersion of the time–averaged phase noise $\sigma^2_{\phi}(T)$ decreases with increasing averaging time as $\propto 1/T$, just like the thermal noise. In this case, only H_2 figures as the unknown coefficient which is not directly available from the measured AVAR alone.

We can estimate the cut–off frequency ω_h, and then H_2, in the following way. Suppose that actual measurements of the phase noise are performed with a time interval τ_m. If we regard the measured phase noise value as a kind of the running mean for the duration τ_m, the high frequency cut–off should be $\omega_h \cong 1/\tau_m$. Then, we can estimate $\sigma^2_{\phi}(T)$, using this ω_h and $H_2\,\omega_h$ derived from the measured AVAR.

1.3   Time Synchronization

In order to find the white fringe within the coherence interval, the clocks of the element antennas must be synchronized with accuracy τ_{sync}, which should be better than the coherence interval $2/B$, where B is the recorded bandwidth, as we saw in Chapter 3.

Therefore, even in the early days of VLBI observations, with the typical bandwidth of $B = 2$ MHz, we needed high time–synchronization accuracy:

$$\tau_{sync} < 1\ \mu\text{sec for } B = 2\text{MHz},$$

which was not easily available from time transfer systems at that time using surface waves.

Nowadays, the requirement is much more severe since we use the observing bandwidth with $B = 256$ MHz or wider. Then, we need

$$\tau_{sync} < 7.8\ \text{nsec(!) for } B = 256\text{MHz}.$$

In actuality, multi–lag correlators, which we will discuss later, can significantly ease this requirement. But, anyway, we need highly accurate time synchronization better than 100 nsec.

Fortunately, now GPS (Global Positioning System) Satellites are capable of providing time synchronizations at a few tens nsec level. Therefore, right now we do not have any essential problem in the time synchronization technology.

Once VLBI fringe is successfully detected, VLBI itself serves as the best time synchronization device, with 1 nsec level accuracy.

1.4 Recording System

Even in the very early stages, VLBI required a very high data rate and capacity for the recording system. For example, if we wish to record digitized data, with bandwidth $B = 2$ MHz, sampled at Nyquist rate (which equals $2B$ samples/sec) with 1 bit (or 2–level) quantization, for a time duration of 400 sec, we need a data recording rate of 4 Mbit/sec, and a data capacity of at least 1.6 Gbit.

Nowadays, new recording sytems, such as the VERA system, allow 1 hour recording at a 1 Gbit/sec rate, per volume of magnetic tape. Therefore, such a tape records 3.6 Tbit of data.

TABLE 9.5 Characteristics of Some VLBI Tape-Recording Systems

System	Period of Use	Basic Description	Tape Recorder	Sample Rate[a] (10^6 s^{-1})	Tape Time (min)	References
NRAO Mark I[b]	1967–78	IBM computer-compatible format	Ampex TM-12	0.72	3.2	Bare et al. (1967)
NRAO Mark II(A)	1971–78	Digital recording on TV recorder	Ampex VR660C	4	190	Clark (1973)
NRAO Mark II(B)	1976–82	Digital recording on TV recorder	IVC 800	4	64	
NRAO Mark II(C)	1979–	Video cassette recorder	RCA VCT 500	4	246	
Canadian	1971–83	Analog recording on TV recorder	IVC 800	8	64	Broten et al. (1967), Moran (1976)
MIT/NASA Mark III	1977–	Instrumentation recorder	Honeywell 96	112[c]	13.6	Rogers et al. (1983)
MIT/NASA Mark III(A)	1984–	Instrumentation recorder	Honeywell 96[d]	112[c]	164	Clark et al. (1985)
NRAO VLBA	1990–	Instrumentation recorder	Honeywell 96[d]	128[e]	720[f]	Hinteregger et al. (199 Rogers (1995)
MIT/NASA Mark IV	1997–	Instrumentation recorder	Honeywell 96[d]	1024	90	Whitney (1993), Roge
S2 (Canada)	1992–	8 Video cassette recorders		128	256	Wietfeldt et al. (1996) Cannon et al. (1997)
K-4 (Japan)	1990–	Video cassette recorder	Sony DIR-1000	256	63	Kawaguchi (1991)

Figure 37: VLBI systems (from Thompson, Moran, and Swenson, 2001).

Generations of "VLBI systems" were marked by the development of the digital data recording technology, as shown in Figure 37.

The Communications Research Laboratory (CRL, now NICT: National Institute of Information and Communications Technology) has developed the Mark II–compatible K-1, the experimental real–time VLBI system K-2, the Mark III–compatible K-3, and cassette–based K-4 systems, by its own efforts (Takahashi, et al., 2000).

Figures 38, 39, 40 and 41 show the VLBA recorder, S2 recorder, K-4 recorder, and the tape handler for K-4 recorder. Figure 42 shows the S2, K-4, and VLBA (Mark IV) tapes.

95

Figure 38: VLBA recorder (MIT Haystack Observatory/NRAO, USA).

Figure 39: S2 recorder (Center for Research in Earth and Space Technology, Canada).

Figure 40: K-4 recorder (NICT/CRL, Japan).

Figure 41: Automated tape handler for K-4 (NICT/CRL, Japan).

Figure 42: Magnetic tapes used in VLBI: S2 (top left), K-4 (left), and VLBA/Mark IV (right).

2 Overview of the VLBI System

2.1 MK–3 As a Prototype of Modern VLBI Systems

Figure 43 shows a schematic view of the Mark III / K-3 VLBI system, often called "MK–3", which was initially developed by the MIT Haystack Observatory with sponsorship from NASA, and by the CRL (now NICT), in the late 1970's. Although this system has now been almost replaced by newer systems, such as the Mark IV, Mark V, VLBA, K-4, K-5 and VSOP, many important elements of these modern VLBI systems were implemented for the first time in the MK–3 system, and have been further developed in the latest systems. Therefore, we briefly describe here the basic components, and their functions, of the MK–3 system.
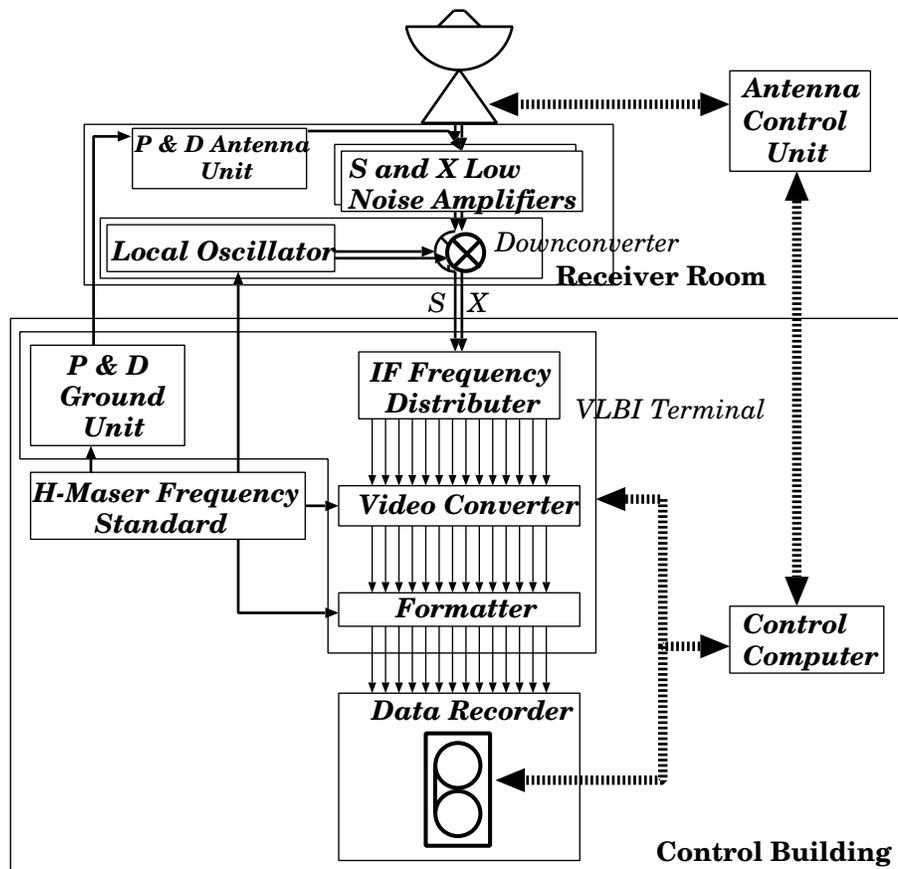
Figure 43: A schematic view of the Mark III / K-3 VLBI system.

The MK–3 system was originally developed mainly for realizing high–

precision VLBI geodesy, although the system has been extensively used for astrophysical radio source imaging observations as well. In order to accurately estimate the group delay, the system adopted a multi–frequency channel design, together with high–speed recording technology; these features have been retained in most of the latest systems.

2.1.1 Dual–Frequency Reception

In geodetic applications, a VLBI antenna receives radio waves from distant extragalactic sources usually at two frequencies simultaneously, in order to correct the effects of frequency–dependent propagation delay in the ionosphere. The two frequencies, S–band (2 GHz) and X–band (8 GHz), are most widely used in global geodetic VLBI observations. The RF frequency bands in the MK–3 system typically covered 200 MHz for S–band (e.g., from 2120 to 2320 MHz), and 420 MHz for X–band (e.g., from 8180 to 8600 MHz). Multi–frequency coaxial feed horns are widely used for dual–frequency reception. Other systems, including FSS (Frequency Selective Surfaces) and spiral arrays (see, e.g., Figure 44), are also used.
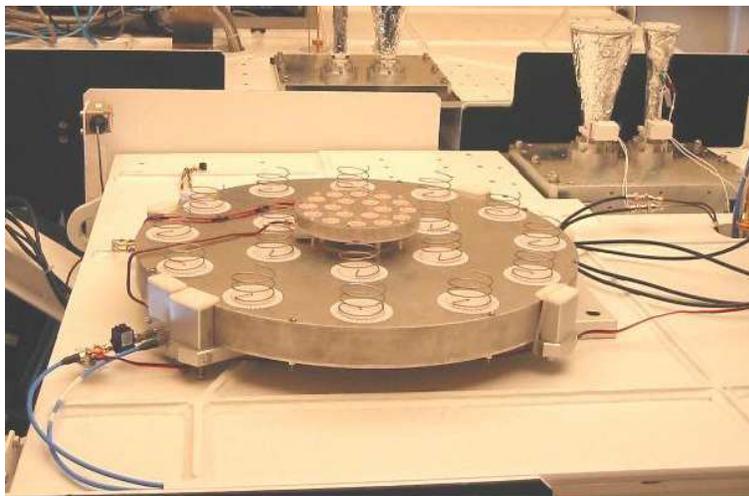
Figure 44: S/X spiral–array feed system, developed by Hosei University, Japan, is used for dual–frequency reception in VERA.

2.1.2 First Frequency Conversion

In the receiver room, usually built in the antenna structure, the RF signals are amplified by the S–band and X–band low–noise amplifiers, and down-

converted to, for example, 100 – 300 MHz (S–band) and 100 – 520 MHz (X–band), respectively, with the local oscillator signals at frequencies of 2020 MHz (S–band) and 8080 MHz (X–band), which are generated from a reference signal provided by the Hydrogen Maser Frequency Standard.

2.1.3 Transmission to Control Building

Then, the IF signals are fed to a so–called MK–3 VLBI terminal rack in a control building, via transmission cables which are usually laid in underground ducts, in order to minimize phase fluctuations due to cable length variations, induced by temperature changes.

2.1.4 Intermediate Frequency Distributer
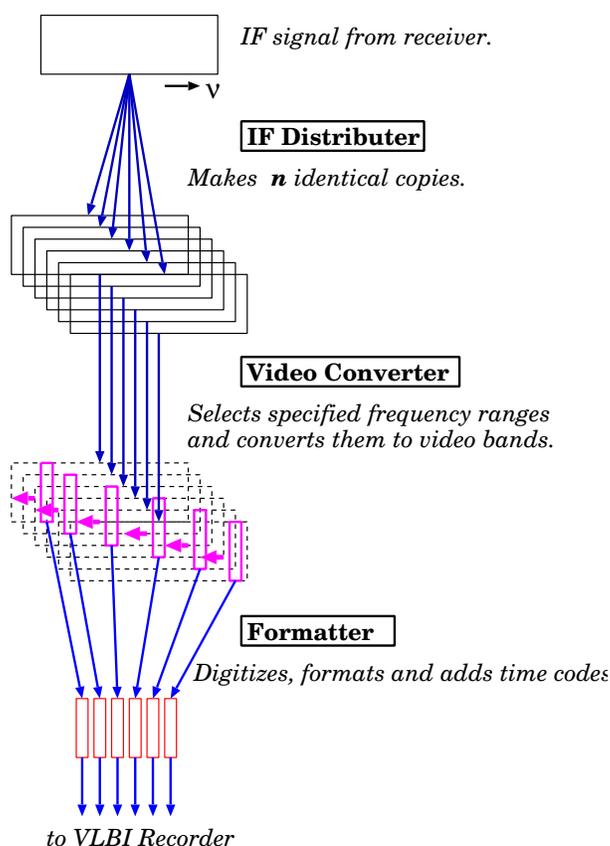
Figure 45: Data flow in the MK–3 VLBI terminal.

In the VLBI terminal rack, the S–band and X–band IF signals are copied

103

over to 6 (S–band) and 8 (X–band) identical signals, respectively, by the so–called "Intermediate Frequency Distributer" unit (Figure 45).

2.1.5 Baseband Conversion

The 6 and 8 identical IF signals are then sent to 14 special units consisting of frequency downconverters, called "Video Converters" (VC's) (or "Baseband Converters", BBC's) located in the same rack. Each Video Converter selects a specified frequency range in the IF band, and downconverts this range into a video band, typically from 0 to 2 MHz. For this purpose, a frequency synthesizer, which is built in the Video Converter, generates an LO frequency at the edge of the selected IF range, using a reference signal provided by the Hydrogen Maser Frequency Standard. The Video Converter is equipped with a sideband–rejection mixer, which is capable of converting both lower and upper sidebands around the selected LO frequency, separately. Thus, in total up to 28 (14 × 2 SB's) baseband frequency channels are generated by the 14 Video Converters. Usually, of course, the baseband channels are selected in such a way that they correspond to different RF frequency ranges. Table 6 and Figure 46 show an example of the LO frequency distribution used in a recent geodetic VLBI observation. This apparently strange distribution of the data among the baseband channels, which are to be subsequently digitized and recorded, are spread over fairly wide ranges of the RF bands; This is specifically designed for better estimation of the group delay observables via a technique called "bandwidth synthesis", which will be discussed later.
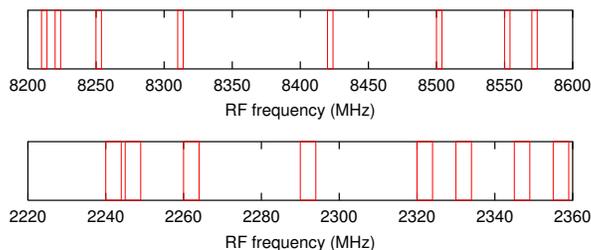
Figure 46: Distributions of baseband channels within the RF frequency ranges, in a sample geodetic VLBI observation, as used in Table 6. Top: X–band, and bottom: S–band.

Ch#	band	LO freequency (MHz)	bandwidth (MHz)
1	X	8209.99	4.00
2	X	8219.99	4.00
3	X	8249.99	4.00
4	X	8309.99	4.00
5	X	8419.99	4.00
6	X	8499.99	4.00
7	X	8549.99	4.00
8	X	8569.99	4.00
9	S	2239.99	4.00
10	S	2244.99	4.00
11	S	2259.99	4.00
12	S	2289.99	4.00
13	S	2319.99	4.00
14	S	2329.99	4.00
15	S	2344.99	4.00
16	S	2354.99	4.00

Table 6: An example of the distribution of LO frequencies in baseband channels. Note that this example is taken from a recent geodetic VLBI observation, which uses the K-4 VLBI system. The K-4 is newer than the MK–3, and has 16 Video Converters (therefore, 16 baseband channels), each with 4 MHz bandwidth.

2.1.6 Formatter

The baseband signals from the 14 Video Converters (possibly containing both USB and LSB channels) are then fed to 14 "Formatter" units in the same rack, each of which converts the analog baseband signal into a digital signal, using a high–speed sampler, with a one–bit (or two–level) quantization scheme, and Nyquist sampling rate (i.e. $2B_{VC}$ samples/sec, where $B_{VC} \leq 2$ MHz is the bandwidth of each video band channel). At the same time, the Formatter periodically generates time mark codes, using the reference signal and the clock pulse provided by the Hydrogen Maser Frequency Standard, and inserts them into the digitized data. These time mark codes play basic and important roles in the later digital correlation processing.

2.1.7 Data Recorder

Up to 28 channels of the digitized, formatted data are sent to a high–speed data recording device, the Honeywell M–96 open–reel digital recorder, in the case of the MK–3 system and its successors (see Figure 38). The formatted multi–channel data are recorded in parallel tracks.

From the above discussion, we see, that the maximum recording rate in the MK–3 system is given by:

$$1 \text{ bit/sample } \times \ 2 \times 2 \text{ M sample/sec (Nyquist rate) } \times$$
$$2 \text{ sidebands } \times \ 14 \text{ Video Converters } = 112 \text{ Mbit/sec.}$$

2.1.8 Phase and Delay Calibration

The 14 Video Converters (VC's) add their own arbitrary initial phases to the converted video band signals. It is necessary, for the accurate estimation of the group delay, to calibrate the phase offsets among the video bands due to the VC initial phases. For that purpose, a special device called the "Phase and Delay Calibrator" is used (Figure 47). This calibrator system, which consists of a Ground Unit in the VLBI terminal rack, and an Antenna Unit in the receiver room, generates so-called "comb–tone" signal, which is nothing but the comb function given in equations (19) and (23), i.e. equally spaced in time delta functions ("pulse series"). As we saw before, Fourier transform of the comb function in the time domain is the comb function in the frequency domain. Therefore, in the frequency domain, the comb–tone signal consists of a large number of sine waves at equally spaced frequencies. Moreover, the sine waves have regularly aligned phases, based on the reference signal from the Hydrogen Maser Frequency Standard. The Antenna unit generates the comb–tones and adds them to the observed signal in the RF band. In the

actual observations, the LO frequencies of the Video Converters are selected in such a way that at least one comb–tone signal falls into each of the video bands (e.g., at a frequency of 10 kHz in the MK–3 system). The comb–tone signals pass through the same units as the observed data do, and, therefore, are affected by the same phase offsets due to the VC initial phases, as the observed data are. Then, the data containing the comb–tones are digitized and recorded. These comb–tones are later detected in the correlation processing, and the phases of the comb–tones in the different video bands are compared with each other, in order to estimate, and then remove from the observed data, the unknown phase offsets due to the VC initial phases.
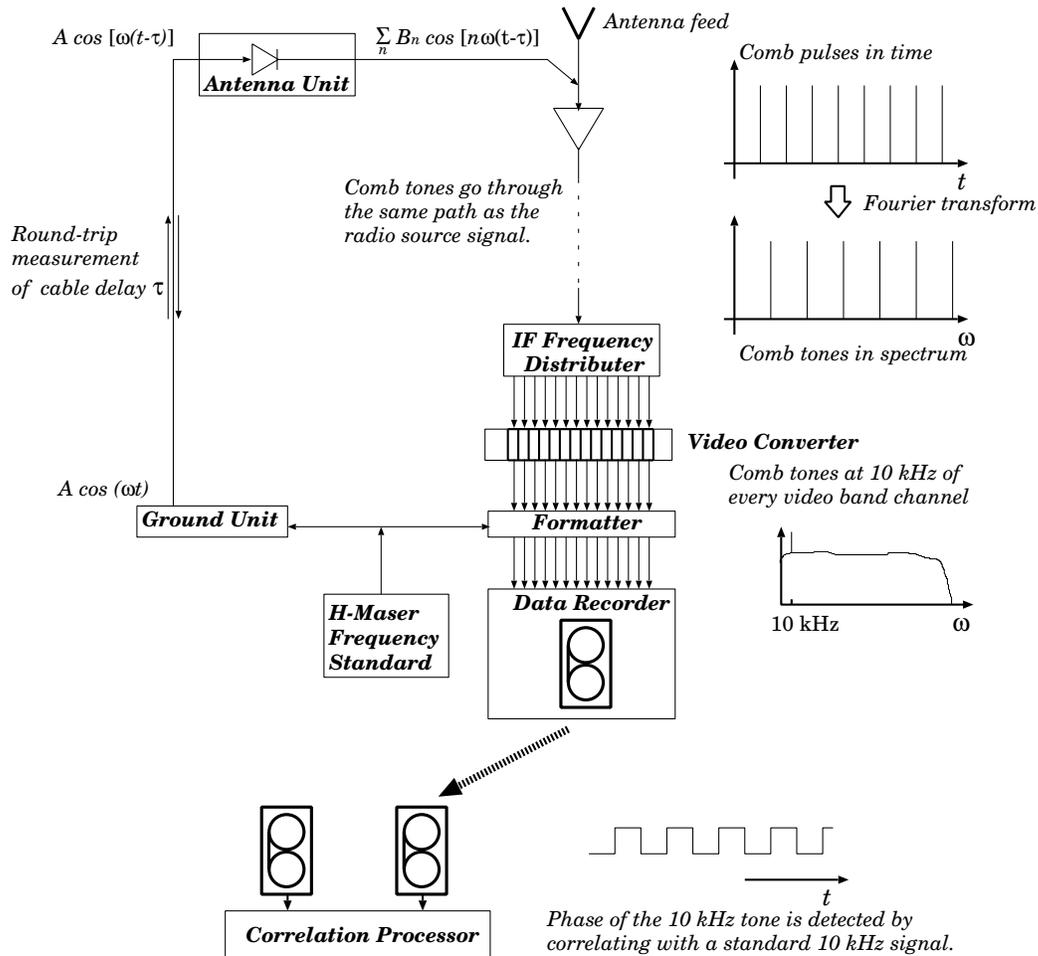
Figure 47: Phase and delay calibration system in the MK–3.

107

2.1.9 Hydrogen Maser Frequency Standard

The Hydrogen Maser Frequency Standard provides very high stability reference signals, usually at 5 MHz or 10 MHz, to the first downconverters in the receiver room, the Video Converters, the Formatters, and to the Phase and Delay Calibrator. It also provides clock pulses, usually at 1 PPS (pulse per second), to many devices, to guarantee their synchronous operation. The Hydrogen Maser Frequency Standard is usually placed in a special magnetically–shielded and temperature–controlled room, to shied against external disturbances as much as possible. Figure 48 shows a Hydrogen Maser Frequency Standard installed in the VERA Mizusawa station.

Figure 48: A Hydrogen Maser Frequency Standard in the VERA Mizusawa station.

2.1.10 Automated Operation by Control Computer

All devices in a VLBI observing station are usually designed to be fully controllable by a single control computer (Figure 49). All commands for setting LO frequencies, sampler modes, recording rates, etc., are remotely controlled. Also, commands for antenna operations, measurements of system noise temperature and meteorological parameters, and so on, are issued from the control computer. In a VLBI observation, a so–called VLBI schedule file is given to the control computer, where all setting information, coordinates of the observed radio sources, and station locations are listed. Also, the detailed time sequence of the observing events, such as repointing the antenna to the next

source, tape start, tape stop, and so on, are given in Universal Times (UT).
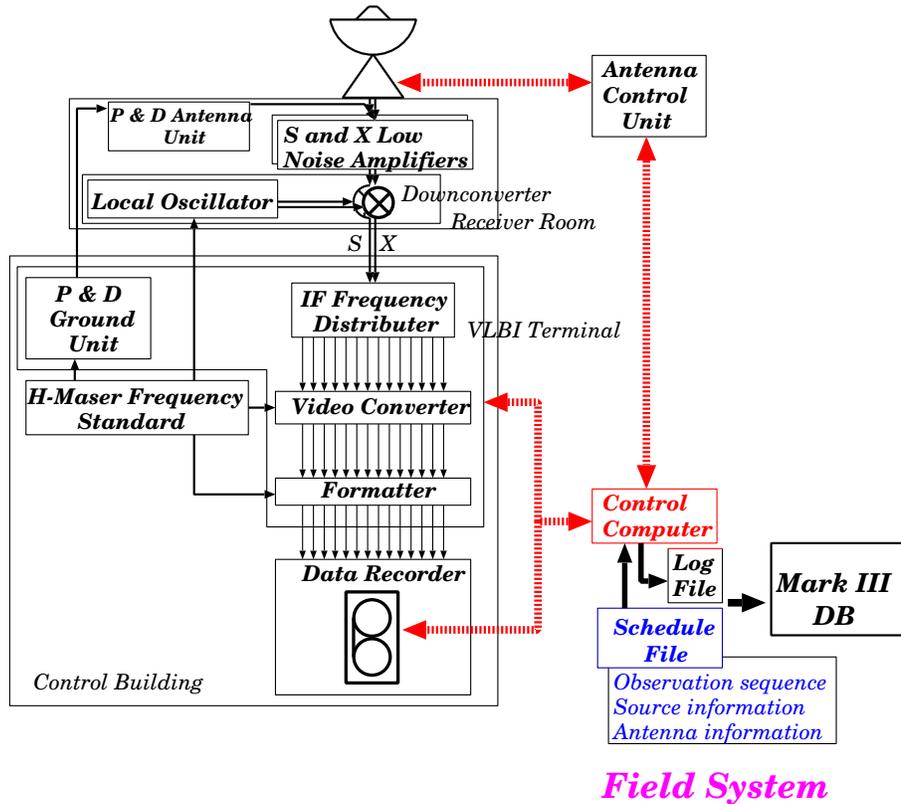
Figure 49: A typical automated VLBI operation system.

The control computer automatically conducts all steps of the observation by issuing commands to the antenna control unit, VLBI terminal, and data recorder, according to the schedule file. As a result, the only remaining task, to be done by an operator in a normal VLBI observation, is to change the recording tapes once an hour or, at most a few times a day, depending on the system design. The most widely used software for automated control of the VLBI equipment, which was developed in the NASA Goddard Space Flight Center (NASA/GSFC), is called the "VLBI Field System".

2.2 Modern VLBI Systems

2.2.1 New Recording and Fiber–Link Systems

The major successors of the MK–3 system are listed in Figure 37. They are the VLBA and Mark IV sytems based on the advanced head controls of the

109

Honeywell M–96 recorder (Figure 38), the S2 system using 8 video cassette recorders in parallel (Figure 39), and the K-4 system based on an ID1 digital video cassette recorder (Figure 40).

Nowadays, modern VLBI recording systems with 1 Gbps recording speed have been developed and some of them are already in use (Table 7). They are the Mark V system (Figure 50) using a hard disk array instead of the magnetic tapes, the K–5 system also based on the hard disk array (Figure 51), the GBR-1000 system based on an HDTV video cassette recorder (Figure 52), the VERA recording system based on a new ID1 digital video cassette recorder (Figure 53), and the fiber–linked e–VLBI.

name	type	bitrate
Mark V	hard disk array	1024 Mbps
K–5	hard disk array	512 & 2048 Mbps
GBR-1000	HDTV	1024 Mbps
VERA	ID1	1024 Mbps
e–VLBI	real–time fiber link	> 2048 Mbps

Table 7: New generation recording and fiber–link systems.
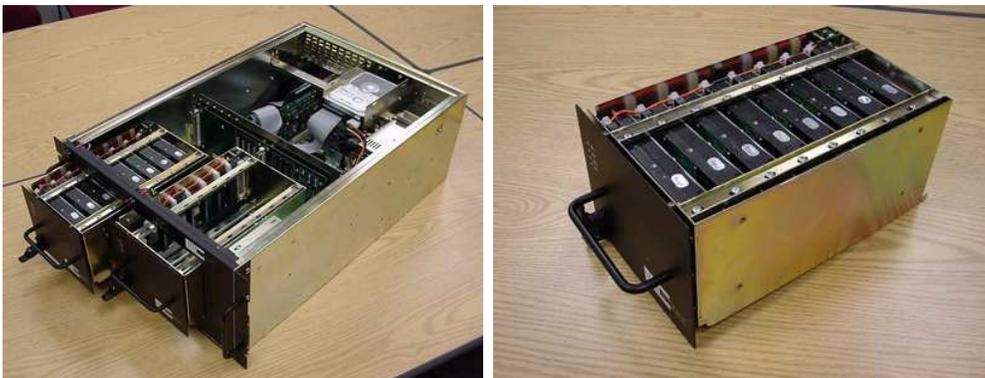
Figure 50: Mark V recording system based on a new hard disk array design.
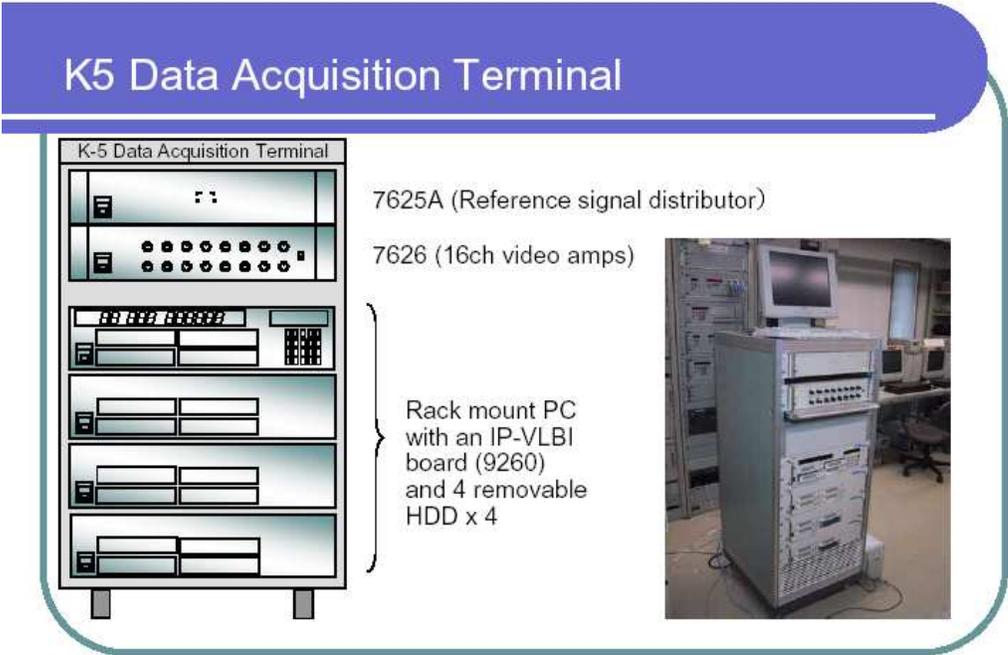
Figure 51: K–5 VLBI terminal.

Figure 52: GBR-1000 1 Gbps VLBI system based on an HDTV video cassette recorder.

Figure 53: VERA recording system based on a new ID1 digital video cassette recorder.

2.2.2 Digital Baseband Converters

Another progress is beeing made in several countries for implementing digital BBC systems based on the digital filtering technology. The idea is first to
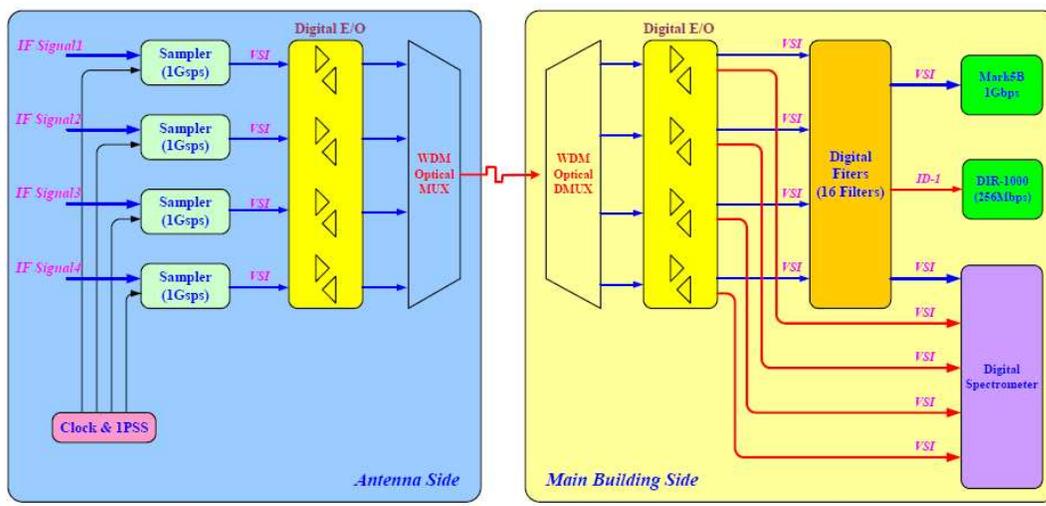
Figure 54: Data Aquisition System of the KVN adopting the digital filter system for the baseband conversion.

digitize a wideband IF data using a high–speed sampler (analog–to–digital converter), and second to cut the digital data into baseband channels by means of a high–speed digital filter. Then the baseband data will no longer suffer from irregular bandpass characteristics of analog BBC's and arbitrary

113

phases added by the LO's of the BBC's. In fact, the digital filter can in principle remove the phase calibration systems discussed above, as far as the LO phase problem only is concerned. Also, the digital filter allows flexible organization of baseband channels meeting different scientific requirements. For example, a single channel ultra–wideband data could be suited to imaging very weak continuum sources, while 16 channel data are necessary for geodetic observations. Such a digital BBC system was developed and successfully utilized in 4 stations of the VERA array.

Figure 54 shows the Data Aqusition System (DAS) of the KVN which uses a digital BBC system for the baseband conversion. The DAS system includes 4 high–speed samplers located in the receiver room. The 4 samplers are equipped for 4 frequency bands to be simultaneously received for mm–wave VLBI observations. Each sampler is capable of digitizing IF data of 512 MHz bandwidth using the higher–order sampling technique discussed earlier with 1 Gsps speed in the 2–bit quantization mode (therefore, the output bit–rate is 2 Gbps). The digital data are transmitted to the control building via optical fiber cables. Then the digital filter forms 1, 2, 4, 8, and 16 baseband channels with bandwidths of 256, 128, 64, 32, and 16 or 8 MHz, correspondingly, out of the 512 MHz / 2–bit / 1 Gsps input, depending on scientific purposes.

2.2.3   e–VLBI

Several groups in the world are now developing VLBI systems based on high–speed data transmission techniques via optical fibers. The idea is to replace the data tapes by ultra–wide–band transmission cables. This technology is now called "e–VLBI". Figures 55 and 56 show the first successful EVN (European VLBI Network) e–VLBI observation conducted by three observatories, in April 2004, and a beautiful image of the gravitational lens object B0218+357 obtained through this coordinated effort.

Of course, this technology by no means transforms VLBI into a connected–element interferometer, since the frequency standards in different stations must remain independent, in view of the essential technical difficulties to transmit reference signals over thousands of kilometers without significant delays or phase fluctuations. Nevertheless, this technology will bring VLBI much closer to the connected–element interferometer, in the sense that the observed data could be correlated and analyzed in real–time, or almost in real time ("near–real-time"). For example, pioneering real–time e–VLBI experiments conducted in the KSP (Key Stone Project, 1995–2001) of the CRL (now NICT) regularly yielded final geodetic results a few minutes after each observation. Moreover, the optical fiber cables offer even higher data trans-
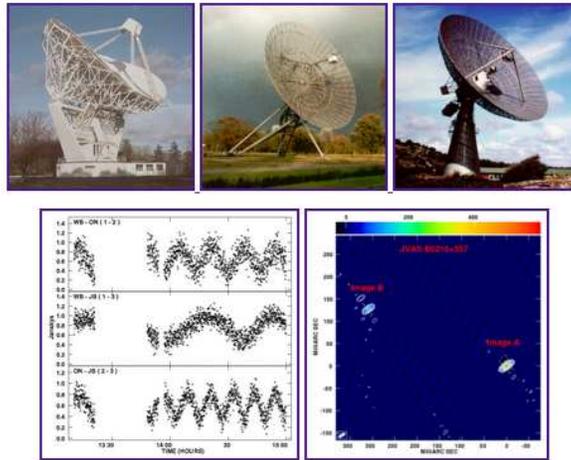
Figure 55: First EVN e–VLBI observation (2004 April) using radio telescopes at Jodrell Bank, UK; Westerbork, the Netherlands; and Onsala, Sweden. The lower panels show beats of the fringe amplitudes produced by two closely spaced sources, and an image map obtained during the observation (from URL: http://www.jive.nl).
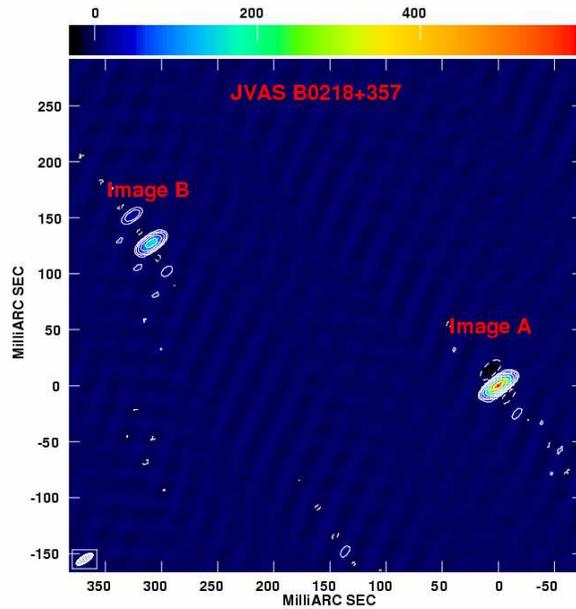
Figure 56: A close–up view of the image of a radio–loud gravitational lens JVAS B0218+357, obtained in the first EVN e–VLBI observation (from URL: http://www.jive.nl).

115

mission rates, and therefore higher sensitivity, than the magnetic tapes, or hard disk arrays. Real–time VLBI experiments with 2.5 Gbps transmission rate have been successfully conducted since 1998 (Figure 57). Even much higher transmission rates are expected in new connected–element interferometer arrays (for example, 96 Gbps / antenna is planned for the Expanded VLA (EVLA), and for the Atacama Large Millimeter and submillimeter Array (ALMA)). High–speed correlators are now being developed to meet these high data rates.

At the same time, less expensive and more widely accessible VLBI data transmission, via broad–band Internet using the IP protocol, is also being intensively studied, and has been successfully tested. This IP–based e–VLBI, or IP–VLBI, seems to be a particularly promising technology which will make VLBI observations much more user–friendly for many astronomers and geophysicists around the world. The K–5 system, developed at the NICT, is designed to realize the IP–VLBI concept (Figure 58).

2.2.4 VLBI Standard Interface (VSI)

Another remarkable example of progress in modern VLBI is the definition of the international "VLBI Standard Interface (VSI)" specifications (http://web.haystack.edu/vsi).

The world VLBI community has long suffered from incompatibily of the various different VLBI systems, shown above, which had been developed in different institutions. In order to cross–correlate data recorded on different tapes by different VLBI systems A and B, say, one had to convert the format of B's data into A's format, copy the converted data to A's tape, and then cross–correlate them with A's correlator, or vice versa.

The purpose of the VSI specifications is to make all VLBI systems in the world compatible, provided only that they obey these standard specifications.

So far, a hardware specification called "VSI–H", and a software specification called "VSI–S", were worked out by an international working group of VLBI specialists. Another specification for e–VLBI, to be called "VSI–E", is now under intensive discussion and development.

The VSI–H hardware specification defines a standard interface between the "data–acquisition system (DAS)" (i.e. VLBI terminals) and the "data–transmission sytem (DTS)" (i.e. tapes or hard disks), as well as between the "data–transmission system" and the "data–processing system (DPS)" (i.e. correlators). Figure 59 shows how the interface could be realized. Every DAS, DTS, and DPS must have common connectors with well specified pin–assignments and data rates. The actual data transmission media (the handler units for tapes or hard–disks) are supposed to be equipped with a "data
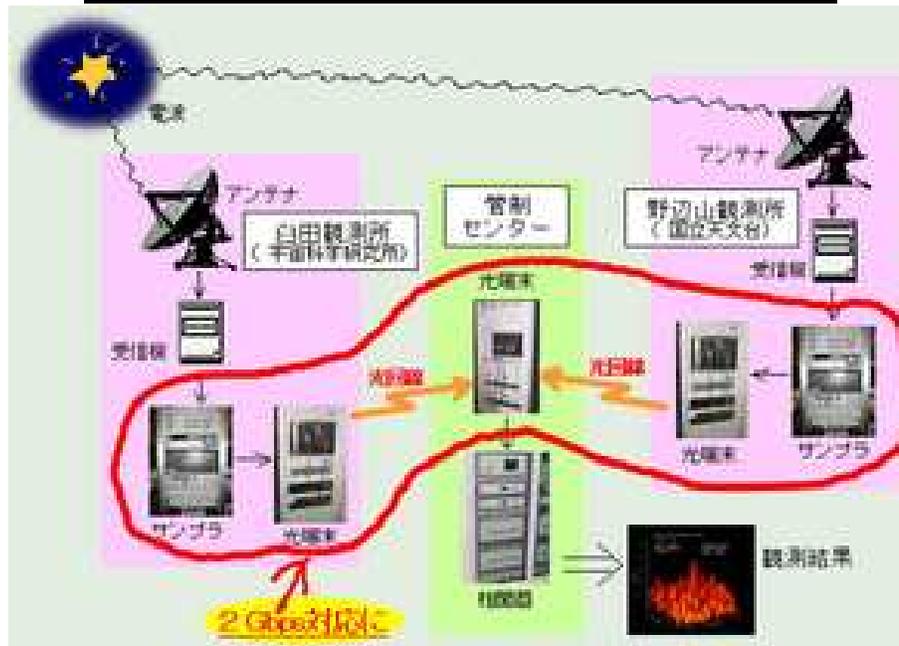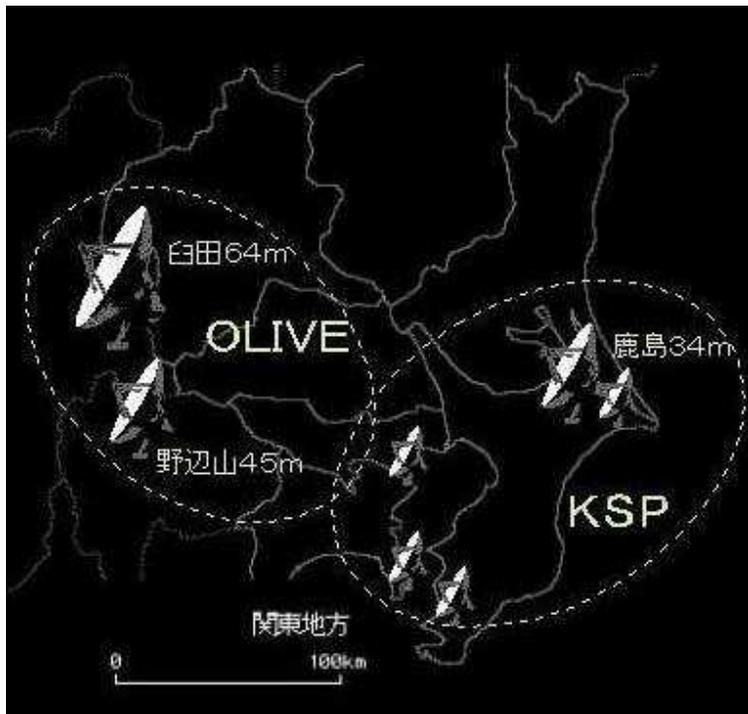
Figure 57: 2.5 Gbps fiber–linked real–time VLBI experiment "OLIVE–GALAXY", Japan (1998, December).
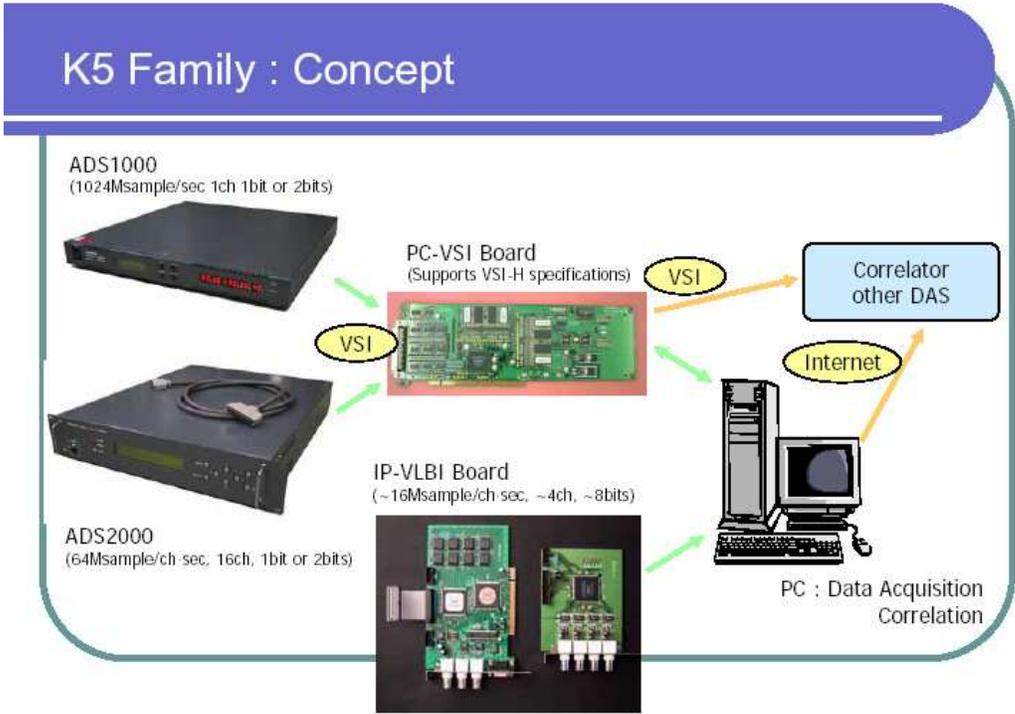
Figure 58: IP–based e–VLBI developed for the K–5 system.
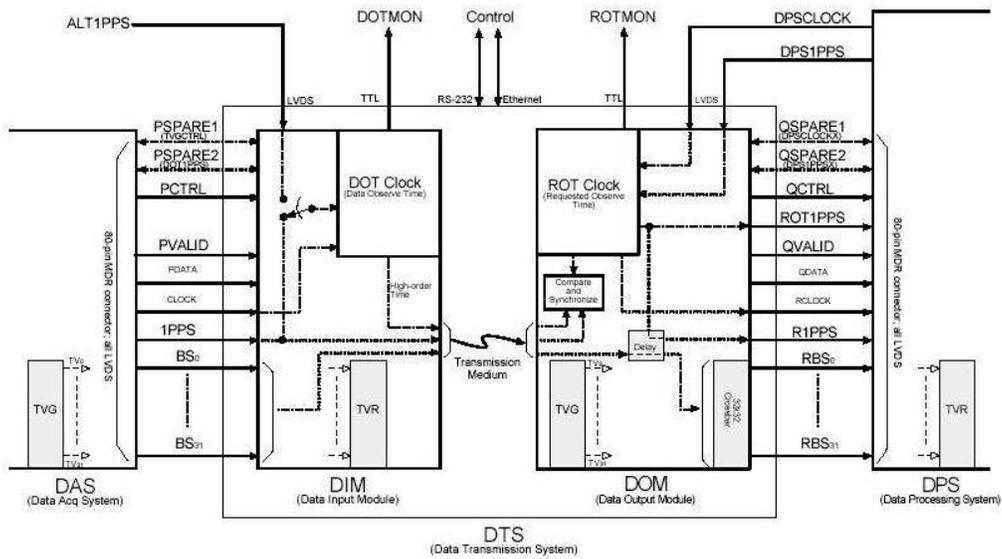
Figure 59: VSI-H functional diagram (http://web.haystack.edu/vsi).

input module (DIM)", and a "data output module (DOM)", which interface the common connectors, and the connectors of the handler hardware. The actual data formats in the individual data transmission media are arbitrary, provided that the DIM accepts the input data stream (specified by VSI-H), and the DOM yields the output data stream (also specified by VSI-H).

The VSI–S software specification defines protocols for handling VSI–H–compliant equipment.

A number of VSI–H–compliant VLBI systems have been developed in various countries, and successfully cross–correlated with each other.

3   Difficulties in Ground–Based VLBI

UNFINISHED.

4   Correlation Processing in VLBI

UNFINISHED.

5   Observables of VLBI

UNFINISHED.

References

Allan, D.W., 1966, Statistics of Atomic Frequency Standards, *Proc. IEEE*, **54**, 221–230.

Hagen, J.B., and Farley, D.T., 1973, Digital Correlation Techniques in Radio Science, *Radio Sci.*, **8**, 775–784.

Iguchi, S., and Kawaguchi, N., 2002, Higher–Order Sampling, Over Sampling and Digital Filtering Techniques for Radio Interferometry, *IEICE, Trans. Commun.*, **E85-B**, 1806–1816.

Meijering, E., 2002, A Chronology of Interpolation: From Ancient Astronomy to Modern Signal and Image Processing, *Proc, IEEE*, **90**, 319–342.

Papoulis, A., 1984, Probability, Random Variables, and Stochastic Processes, 2nd Edition, *McGraw-Hill* (New York).

Price, R., 1958, A Useful Theorem for Nonlinear Devices Having Gaussian Inputs, *IRE Trans. Inf. Theory*, **IT-4**, 69–72.

Rogers, A.E.E., and Moran, J.M., 1981, Coherence Limits for Very–Long–Baseline Interferometry, *IEEE Trans. Instrum. Meas.*, **IM–30**, 283–286.

Shannon, C.E., 1949, Communication in the Presence of Noise, *Proc. IEEE*, **37**, 10–21.

Thompson, A.R., Moran, J.M., and Swenson, G.W., Jr., 2001, Interferometry and Synthesis in Radio Astronomy, *John Wiley & Sons* (New York).

Takahashi, F., Kondo, T., Takahashi, Y., and Koyama, Y., 2000, Very Long Baseline Interferometer, *Ohmsha* (Tokyo).

Van Vleck, J.H., and Middelton, D., 1966, The Spectrum of Clipped Noise, *Proc. IEEE*, **54**, 2–19.