

Tetsuo Sasao and André B. Fletcher

Introduction to VLBI Systems

Chapter 4

Lecture Notes for KVN Students

Partly based on Ajou University Lecture Notes

Version 1.

Issued on November 14, 2006.

Revised on June 18, 2010.

Revised on February 10, 2011.

(to be further edited)

NOT COMPLETE YET.

Very Long Baseline Interferometry

This chapter will focus on technological bases, actual implementation, and data processing of modern VLBI systems.

Contents

1	Technologies Which Made VLBI Possible	6
1.1	Basics of Digital Data Processing	7
1.1.1	Analog Processing Versus Digital Processing	7
1.1.2	Sampling and Clipping	8
1.1.3	Discrete-Time Random Process	9
1.1.4	Stationary Discrete-Time Random Process	10
1.1.5	Sampling	11
1.1.6	Comb Function	13
1.1.7	Fourier Transform of Comb Function Is Comb Function	15
1.1.8	Spectra of Discrete-Time Processes	16
1.1.9	Spectra of Sampled Data	19
1.1.10	Inverse Relations for Spectra of Sampled Data	20
1.1.11	Sampling Theorem	21
1.1.12	Optimum Sampling Interval	25
1.1.13	Sampling Function	26
1.1.14	Correlations of Nyquist Sampled Data with Rectangu- lar Spectra	30
1.1.15	S/N Ratio of Correlator Output of Sampled Data . . .	34
1.1.16	Nyquist Theorem and Nyquist Interval	37
1.1.17	Higher-Order Sampling in VLBI Receiving Systems . .	38

1.1.18	Clipping (or Quantization) of Analog Data	40
1.1.19	Probability Distribution of Clipped Data	43
1.1.20	Cross-Correlation of 1-Bit Quantized Data: van Vleck Relationship	47
1.1.21	van Vleck Relationship in Autocorrelation	53
1.1.22	Spectra of 1-Bit Quantized Data	53
1.1.23	Price's Theorem	57
1.1.24	Cross-Correlation of 2-Bit Quantized Data	61
1.1.25	Cross-Correlation Coefficient of 2-Bit Quantized Data	63
1.1.26	Power and Cross-Power Spectra of 2-Bit Quantized Data	65
1.1.27	Dispersion of Digital Correlator Output	68
1.1.28	S/N Ratio of Digital Correlator Output	70
1.1.29	Optimum Parameters v_0/σ and n in 2-Bit Quantization	72
1.1.30	Effect of "Oversampling" in S/N Ratio of Clipped Data	73
1.1.31	Coherence Factor and Sensitivity with Given Bit Rate	76
1.2	Frequency Standard	79
1.2.1	VLBI Requires "Absolute" Frequency Stability	79
1.2.2	How to Describe Frequency Stability?	81
1.2.3	Types of Phase and Frequency Noises	84
1.2.4	Time Domain Measurements	84
1.2.5	"True Variance" and "Allan Variance" of Fractional Frequency Deviation	86
1.2.6	True Variance and Allan Variance through Power Spec- trum of Fractional Frequency Deviation	88
1.2.7	Time-Interval Dependence of Allan Variance	90
1.2.8	True Variance and Allan Variance through Autocorre- lation of Phase Noise	91
1.2.9	Coherence Function	92
1.2.10	Approximate Estimation of Coherence Time	95
1.2.11	Estimation of Time-Averaged Phase Noise	96
1.3	Time Synchronization	98
1.4	Recording System	99
2	Overview of VLBI Systems	103
2.1	MK-3 As Prototype of Modern VLBI Systems	103
2.1.1	Dual-Frequency Reception	104
2.1.2	First Frequency Conversion	104
2.1.3	Transmission of Observed Data to Control Building . .	105
2.1.4	Intermediate Frequency Distributer	105
2.1.5	Baseband Conversion	106
2.1.6	Formatter	107

2.1.7	Data Recorder	108
2.1.8	Phase and Delay Calibration	108
2.1.9	Hydrogen Maser Frequency Standard	110
2.1.10	Automated Operation by Control Computer	110
2.2	Modern VLBI Systems	111
2.2.1	New Recording and Fiber-Link Systems	111
2.2.2	High-Speed Samplers and Digital Baseband Converters	114
2.2.3	e-VLBI	118
2.2.4	VLBI Standard Interface (VSI)	121
2.2.5	Future Geodetic VLBI System Proposed in “VLBI2010”	122
3	Effects of Clock Offset and Atmospheric Delay	123
3.1	Clock Offset.	123
3.1.1	Measuring Time with Clocks	123
3.1.2	Clock Time Deviation Due to Fluctuations in Frequency Standard	124
3.1.3	LO Phase Deviation Due to Fluctuations of Frequency Standard	125
3.1.4	Clock and LO-Phase Offsets between Two VLBI Stations	126
3.1.5	How Do Fluctuations in Frequency Standards Affect VLBI Data?	127
3.2	Atmospheric Propagation Delay	129
3.2.1	Three Major Components of Atmosphere That Cause Propagation Delay	130
3.2.2	Cross-Power Spectrum of IF Signals with Atmospheric Delay	132
3.3	Expected Correlator Output with Clock and Atmospheric Effects	133
3.3.1	Expected Correlator Output, General Expression . . .	133
3.3.2	Correlator Output for Continuum Spectrum Source . .	134
3.3.3	Simplest Case of Rectangular Filters	135
3.4	Limitations Imposed by Atmosphere and Clock Effects	135
3.4.1	Example of VLBI Fringe Phase	136
3.4.2	Difficulty in Direct Use of VLBI Phase	137
3.4.3	Limitation to Coherence Time	138
4	Correlation Processing in VLBI	139
4.1	Features of Correlation Processing in VLBI	139
4.2	First Step—Hardware Processing	142
4.2.1	Accumulation Period	142
4.2.2	“Classical” XF-Type Correlator	144
4.2.3	Data Synchronization	145

4.2.4	Detection of Phase Calibration Tones	146
4.2.5	Ring Buffer Memory	147
4.2.6	Delay Tracking by Programmable Delay Circuit	148
4.2.7	Address Shift Timing in Programmable Delay Circuit	148
4.2.8	“Fringe Rotator”: Fringe Stopping and Complex Correlation	149
4.2.9	Multi-Lag Correlation Unit	151
4.2.10	Algorithm for 1-Bit Correlation	153
4.2.11	Autocorrelation Unit	154
4.2.12	Coherence Loss in Digital Correlation Processing	154
4.2.13	Output of Multi-Lag Complex Correlator	155
4.2.14	Expected Output of Multiplier & Integrator	157
4.2.15	Cross-Correlation Coefficient as Expected Output	159
4.2.16	Expected Complex Cross-Correlation	161
4.2.17	Expected Complex Cross-Correlation for White Spectrum Source	164
4.2.18	Rectangular Videoband Case	166
4.2.19	Behavior of Correlator Outputs	166
4.2.20	PC-Based Correlators	170
4.2.21	New 16-Station FX-Type Correlator in Seoul	171
4.3	Second-Step, Software Processing (Fringe Fitting)	173
4.3.1	Some Words on Terminology	173
4.3.2	Two-Dimensional Search in Software Processing	174
4.3.3	How to Apply Corrections to Correlator Outputs?	176
4.3.4	Spectrum of Complex Cross-Correlation	177
4.3.5	Amplitude and Phase Spectra of Complex Cross-Correlation	180
4.3.6	Two-Dimensional Taylor Expansion of Residual Phase	181
4.3.7	Residual Group Delay and Residual Fringe Frequency .	182
4.3.8	Trial Corrections to Residual Phase	183
4.3.9	Finding Correlation Peak	185
4.3.10	Fringe Amplitude and Residual Fringe Phase in VLBI .	187
4.3.11	Coherence Factor Due to Imperfect Delay Model	190
4.3.12	Theoretical Form of Uncalibrated Fringe Amplitude	191
4.3.13	Search Window	192
4.3.14	Actual Software Processing	193
4.3.15	Correlation Peak of Continuum Source	197
4.3.16	Estimation of S/N of Detected Fringe	199
4.3.17	Improvement of S/N through Integration	201
4.3.18	Examples of Detected Fringes for Continuum Sources	203
4.3.19	Software Processing for Spectralline Sources	206
4.3.20	If We Do Not Find Any Peak	208

4.3.21	Threshold of Fringe Detection	209
5	Observables of VLBI	211
5.1	Four Basic Observables	211
5.1.1	Direct Products of Correlation Processing	211
5.1.2	Calibration of Fringe Amplitude	212
5.1.3	Estimates and Expected Forms for Observables	213
5.2	Accuracies of VLBI Observables	214
5.2.1	Statistics of Amplitude and Phase	214
5.2.2	Expectations of Amplitude and Phase	217
5.2.3	In Limiting Case When S/N is Large	218
5.2.4	Accuracy of Fringe Amplitude Observable	221
5.2.5	Accuracy of Fringe Phase Observable	222
5.2.6	Group Delay and Fringe Frequency Observables Free from Cycle Ambiguity	223
5.2.7	Accuracy Estimation for Group Delay and Fringe Fre- quency	224
5.2.8	Equivalence of Peak Search Method and Least Squares Estimation	225
5.2.9	Dispersions of Fringe Frequency and Group Delay Ob- servables	232
5.2.10	Accuracy of Fringe Frequency Observable	234
5.2.11	Atmospheric Limitation to Usefulness of Fringe Fre- quency	236
5.2.12	Accuracy of Group Delay Observable	238
5.2.13	Atmospheric Effect in Group Delay Observable	239
5.2.14	Phase Delay and Group Delay	241
5.3	Bandwidth Synthesis Technique	242
5.3.1	For Higher Accuracy of Group–Delay Estimation	242
5.3.2	Model Delay Resolution Function	246
5.3.3	Examples of Model Delay Resolution Functions	246
5.3.4	Ambiguity in Bandwidth Synthesis	249
5.3.5	Actual Bandwidth Synthesis	252
5.3.6	Group–Delay Accuracy in Bandwidth Synthesis	257
6	Atmosphere–Related Difficulties and Their Solutions in VLBI	258
6.1	Difficulties Associated with mm–Wave VLBI	259

1 Technologies Which Made VLBI Possible

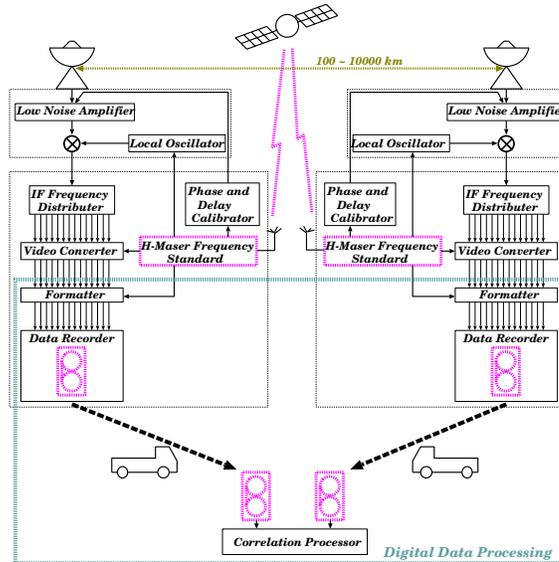


Figure 1: Schematic view of three major technologies and digital data processing which realized VLBI.

VLBI (Very Long Baseline Interferometry) is a kind of radio interferometer with very long baselines often exceeding thousands of kilometers, equipped with independent frequency standards (clocks) at each station. VLBI achieved the highest angular resolution in drawing images of astronomical radio sources (astrophysical VLBI), and the highest accuracy in measuring positions of radio telescope antennas and astronomical radio sources (geodetic VLBI).

There were three major technologies which enabled VLBI. They are:

1. high-stability frequency standard,
2. high-accuracy time synchronization, and
3. high-speed high-density recording, which is now being replaced by super-wideband data transmission via high-speed communication networks.

In addition, digital signal processing technology plays an essential role in VLBI. In fact, key elements of VLBI, such as time marking, recording, delay tracking, fringe stopping, and correlation processing, are all implemented digitally in modern VLBI systems (Figure 1). In this sense, it was the rapid progress in digital technology in the last decades that enabled to form fundamentals of VLBI.

In the present section, 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 three major technologies mentioned above.

1.1 Basics of Digital Data Processing

1.1.1 Analog Processing Versus Digital Processing

It is evident from following comparison that digital processing is suited to VLBI much better than analog processing,

Analog delay circuit	Digital delay circuit
Practically, a coaxial cable can be used as an instrumental delay (delay cable) only when delay is smaller than $\sim 1 \mu\text{sec}$ ($\sim 300 \text{ m}$). 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 characteristics 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. 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.

1.1.2 Sampling and Clipping

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

VLBI data are sampled and clipped (or digitized) usually with 1-bit or 2-bit quantization (Figure 2).

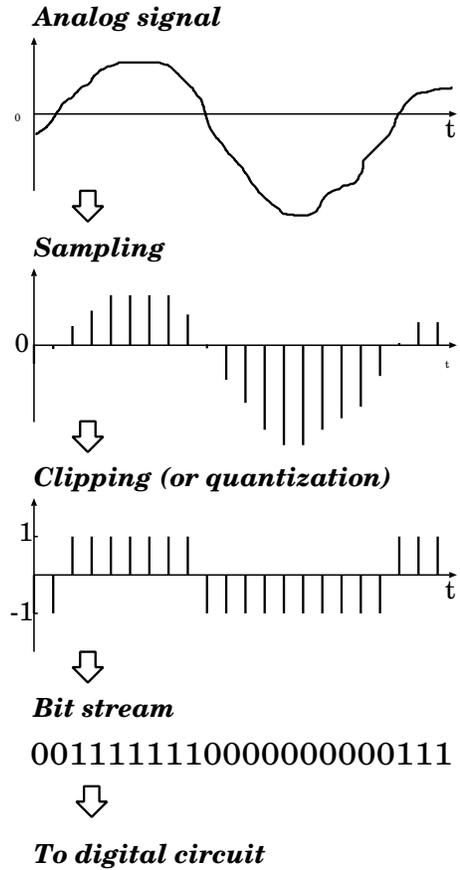


Figure 2: Analog-to-digital (A/D) conversion through sampling, clipping, and bit representation. This figure shows a case of 1-bit quantization.

We will discuss hereafter how we can restore information contained in analog data from sampled and clipped data. We will also consider loss of information due to the digitization.

First we start with statistical concepts useful for treating digital data.

1.1.3 Discrete-Time Random Process

Discrete sequence of variables $x[1], x[2], x[3], \dots, x[i], \dots$ is called “**random sequence**”, or “**discrete-time random process**”, if $x[i]$ at any argument (or sequence number) i is a random variable, i.e., may vary from trial to trial (Figure 3). This is a “discrete version” of the random process continuously varying in time shown in Chapter 3, which we call henceforth “continuous-time random process”.

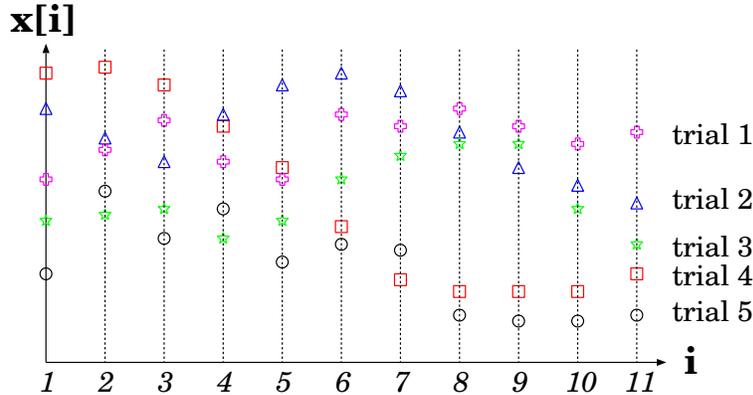


Figure 3: A discrete-time random process is a sequence $x[1], x[2], x[3], \dots, x[i], \dots$, whose value at any argument i is a random variable.

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

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

$$\eta[i] = \langle x[i] \rangle, \quad (1)$$

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

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

$$R[m, n] = \langle x[m] x^*[n] \rangle, \quad (2)$$

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

A discrete-time random process $x[i]$ is called “**white noise**” if its autocorrelation satisfies

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

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

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

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

$$\begin{aligned} &x[1], x[2], x[3], \dots \\ &y[1], y[2], y[3], \dots \end{aligned}$$

is defined by

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

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 transferred 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

$$\eta[i] = \langle x[i] \rangle = \eta, \quad (6)$$

is a constant independent of argument i , and if the autocorrelation

$$R[n + m, n] = \langle x[n + m] x^*[n] \rangle = R[m], \quad (7)$$

depends on difference m of arguments only.

In particular, the stationary random discrete-time process is called the “white noise”, if we have

$$R[m] = R[0] \delta_{m0}. \quad (8)$$

- **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

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

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 random process $x[i]$:

$$r[m] = \frac{R[m]}{R[0]}, \quad (10)$$

and

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

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

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 “**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 “**time-sample**” of the original continuous-time process.

So, 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 , then

$$x[n] = x(nT). \quad (12)$$

Note here that the argument n of the time sample $x[n]$ is really a sequence number of “discrete times” consecutively separated by the interval T .

Statistical properties of $x[n]$ are determined by statistical properties (i.e., by probability distribution) of $x(t)$, as we see below.

- **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

$$\eta[n] = \eta(nT), \quad (13)$$

and

$$R[m, n] = R(mT, nT), \quad (14)$$

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

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

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:

$$\eta[n] = \eta, \quad (16)$$

and autocorrelation:

$$R[n + m, n] = R[m] = R(mT). \quad (17)$$

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:

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

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 4, is called “**comb function**”. Thus a comb function $\sqcup(t; T)$ with period T is given in terms of delta functions $\delta(t)$ by an equation:

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

where k is an integer.

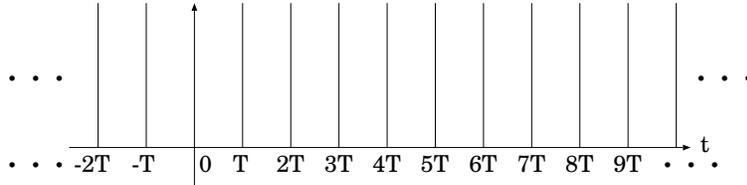


Figure 4: A comb function.

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

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

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

where n is an integer and a_n is a Fourier coefficient.

Following 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 $\frac{T}{2} < t \leq \frac{T}{2}$.

2. Then, we have

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

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

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

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 $a_n = \frac{1}{T}$, for any n , and hence

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

4. Although we derived this equality in a limited range: $-\frac{T}{2} < t \leq \frac{T}{2}$, it actually holds for wider range of t . In fact, functions in 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. $-\infty < t \leq \infty$. Thus, we have a general relation

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

which holds for any t , and, therefore,

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

is the alternative expression of the comb function.

1.1.7 Fourier Transform of Comb Function Is Comb Function

Fourier transform $\tilde{\square}(\omega; T)$ of a comb function $\square(t; T)$ of argument t with period T is a comb function of argument ω with period $\frac{2\pi}{T}$ (Figure 5).

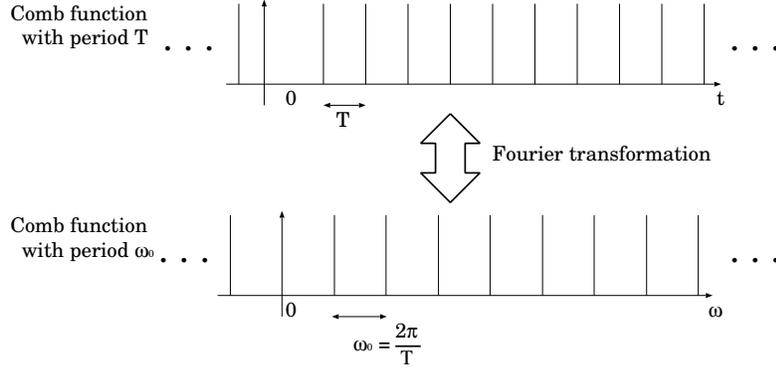


Figure 5: Fourier transform of a comb function of t with period T is a comb function of ω with period $\omega_0 = \frac{2\pi}{T}$.

Proof:

According to the general formula of Fourier transformation, we have

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

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

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

Thus,

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

where a symbol \Leftrightarrow means a Fourier transform pair.

1.1.8 Spectra of Discrete–Time Processes

We introduce following definitions.

- **Power spectrum.**

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

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

- **Cross–power spectrum.**

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

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

The spectra $S_D(\omega)$ and $S_{Dxy}(\omega)$ as given above are periodic functions of ω with a period $2\pi/\mathcal{T}$. Therefore, **equations (26) and (27) represent Fourier series expansions** of periodic functions $S_D(\omega)$ and $S_{Dxy}(\omega)$ with the period $2\pi/\mathcal{T}$ having $R[m]$ and $R_{xy}[m]$, respectively, as Fourier coefficients.

The spectra $S_D(\omega)$ and $S_{Dxy}(\omega)$ are dependent on the arbitrary parameter \mathcal{T} which determines their period $2\pi/\mathcal{T}$. Later, for particular cases of sampled discrete–time random 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.

An autocorrelation $R[m]$ of a stationary discrete-time random process $x[n]$ is derived from its power spectrum $S_D(\omega)$, while a cross-correlation $R_{Dxy}[m]$ of jointly stationary discrete-time random processes $x[n]$ and $y[n]$ is derived from their cross-power spectrum $S_{Dxy}(\omega)$, by inverse relations:

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

and

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

Naturally, equations (28) and (29) are the formulae for Fourier coefficients in the Fourier series expansion given in equations (26) and (27).

Mutual consistency of equations (26), (27) and (28), (29) is well known. For instance, we can directly confirm it as follows. We consider here the power spectrum case only, since the cross-power spectrum case can be confirmed exactly in the same way.

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

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

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

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

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

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

Now, $\sum_{n=-\infty}^{\infty} e^{in(\omega'-\omega)T}$ in the RHS is a comb function as given in

equation (23), since, introducing a notation $\omega_0 = \frac{2\pi}{T}$, we have

$$\begin{aligned} \sum_{n=-\infty}^{\infty} e^{in(\omega'-\omega)T} &= \sum_{n=-\infty}^{\infty} e^{i\frac{2\pi n}{\omega_0}(\omega'-\omega)} = \omega_0 \square(\omega' - \omega; \omega_0) \\ &= \omega_0 \sum_{k=-\infty}^{\infty} \delta(\omega' - \omega - k\omega_0) \\ &= \frac{2\pi}{T} \sum_{k=-\infty}^{\infty} \delta\left(\omega' - \omega - k\frac{2\pi}{T}\right), \end{aligned} \quad (30)$$

where we used equation (22). Therefore, we obtain

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

since the delta function in the integrand takes a non-zero value at

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

only when $k = 0$, as long as ω is confined within the interval $-\frac{\pi}{T} < \omega \leq \frac{\pi}{T}$.

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

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

for any range of ω .

Thus we confirmed the inverse relations.

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).

Now, if we choose the arbitrary parameter \mathcal{T} in equations (26) and (27) to be equal to the sampling interval T , i.e.,

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

then the power spectrum $S_D(\omega)$ and the cross-power spectrum $S_{Dxy}(\omega)$ of the stationary discrete-time random processes $x[n]$ and $y[n]$ given in equations (26) and (27) are related to power spectrum $S(\omega)$ and cross-power spectrum $S_{xy}(\omega)$, respectively, 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 $\omega_0 = \frac{2\pi}{T}$.

Proof:

We prove equation (32) only, since a proof of equation (33) is given just in the same way.

In view of equations (17), (26), and (31), the power spectrum of the time-sample $x[n] = x(nT)$ is now given by

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

since $\mathcal{T} = T$, 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), $\sum_{n=-\infty}^{\infty} e^{in(\omega'-\omega)T}$ is a comb function:

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

Therefore,

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

where $\omega_0 = \frac{2\pi}{T}$. Thus we proved equation (32). We can also prove equation (33) in the same way.

1.1.10 Inverse Relations for Spectra of Sampled Data

When we have a power spectrum $S_D(\omega)$ of a stationary discrete-time random process $x[n]$, we can always derive an autocorrelation $R[m]$ of the same process by the inverse relation shown in equation (28):

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

If the process $x[n]$ is a time-sample $x[n] = x(nT)$ with a sampling time T of a stationary continuous-time random process $x(t)$ which has an autocorrelation $R(\tau)$, the autocorrelation $R[m]$ derived from the inverse relation must satisfy the relation $R[m] = R(mT)$. We can directly confirm this as follows.

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

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

Similarly, we can confirm that the inverse relation in equation (29):

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

gives a cross-correlation $R_{xy}[m]$ of time-samples $x[n]$ and $y[n]$ which satisfies $R_{xy}[m] = R_{xy}(mT)$.

1.1.11 Sampling Theorem

Shannon (1949) formulated the sampling theorem 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.”

He gave a mathematical proof of the theorem based on the very principle of Fourier series expansion, noting that “this is not only approximately, but exactly, true”. His discussion is so concise and elegant that it would be better here to directly cite it rather than to attempt to summarize it.

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

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega) e^{i\omega t} d\omega = \frac{1}{2\pi} \int_{-2\pi B}^{2\pi B} F(\omega) e^{i\omega t} d\omega,$$

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

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

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 $-2\pi B$ to $+2\pi B$ 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 $F(\omega)$, 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.”

(We adjusted mathematical symbols to ours.)

Shannon (1949) mentioned that Nyquist had pointed out the fundamental importance of the time interval $1/2B$ in connection with telegraphy, and proposed to call it “**Nyquist interval**” corresponding to bandwidth B . Also, the sampling frequency $2B$ is often called “Nyquist rate”.

Shannon (1949) considered the case when a function has a spectrum of frequency ν confined within a range $|\nu| < B$. Nowadays, we formulate the sampling theorem for a slightly wider class of spectrum (see Figure 6).

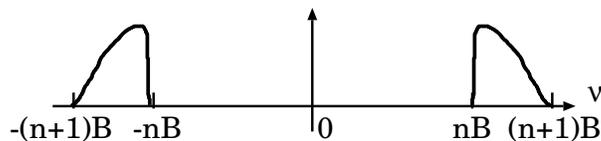


Figure 6: A passband spectrum conforming to the sampling theorem.

Sampling Theorem:

All 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, is preserved, provided that the signal is sampled with the Nyquist interval $1/(2B)$.

A proof of the above theorem is already given in equations (32) and (33) as illustrated in Figure 7.

In fact,

- if 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 7). Therefore, all 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 within a range $|\nu| < B$ 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 7).

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 7, or 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 7,

then focts 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.

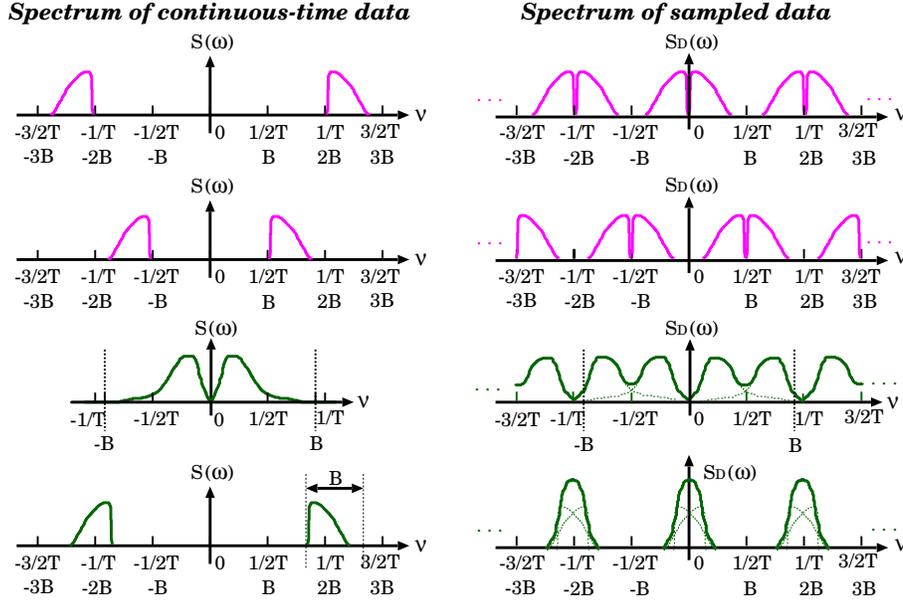


Figure 7: Four cases of the relation between spectrum of analog continuous-time data and spectrum of their time samples as given by equations (32) and (33):

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

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 spectrum 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 with 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.

1.1.12 Optimum Sampling Interval

The Nyquist interval is generally recognized as the optimum interval for the sampling. To be convinced, let us consider an analog continuous-time

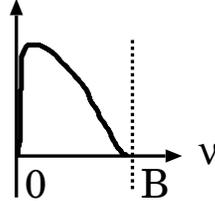


Figure 8: Positive frequency part of a baseband spectrum confined within a frequency range $|\nu| < B$.

spectrum which is confined within a “baseband” $|\nu| < B$. Here, the “baseband”, or otherwise called the “videoband”, implies a frequency band containing 0 Hz (or “DC”, which means “direct current”) as the lowest frequency, such as shown in Figure 8. This is a particular case of the passband spectrum within $nB \leq |\nu| < (n+1)B$ when $n = 0$. 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 $|\nu| < B$, but zero in the outside.

Then, we can conceive three cases which are shown in Figure 9.

1. Oversampling

If we sample an analog continuous-time signal with sampling interval smaller than the Nyquist interval $T < 1/(2B)$, as shown in top panel of Figure 9, 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 case shown in the middle panel of Figure 9. Such a sampling 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 9. This case is called the “undersampling”.

3. Nyquist sampling

Therefore, the Nyquist sampling, i.e. sampling with the Nyquist inter-

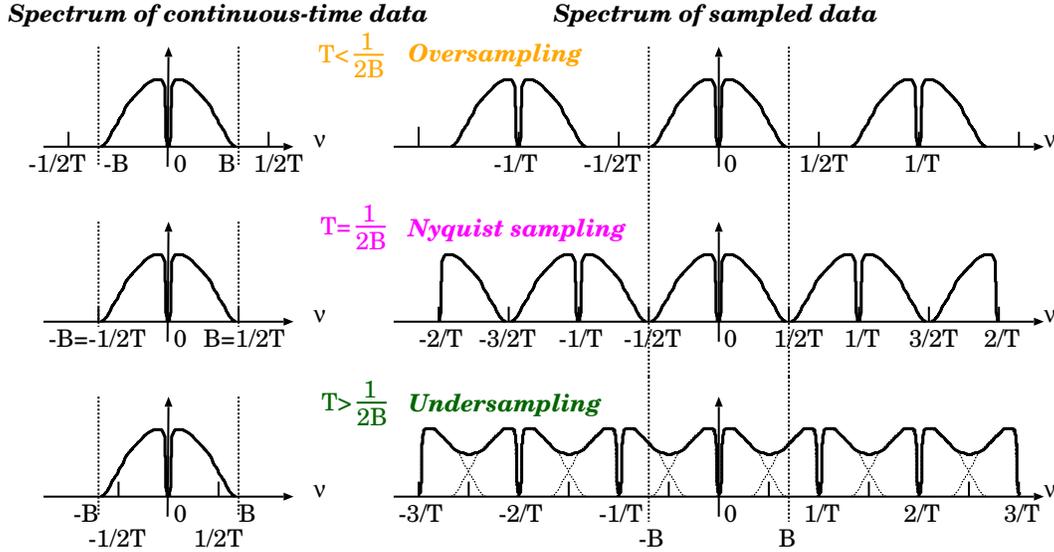


Figure 9: Spectra of sampled data in cases of oversampling $T < 1/(2B)$ (**top**), Nyquist sampling $T = 1/(2B)$ (**middle**), and undersampling $T > 1/(2B)$ (**bottom**).

val $T = 1/(2B)$, is the optimum sampling for an analog continuous-time signal with a band-limited baseband spectrum (middle panel of Figure 9).

1.1.13 Sampling Function

The sampling theorem says that autocorrelation $R(\tau)$ and cross-correlation $R_{xy}(\tau)$ of jointly stationary continuous-time random processes $x(t)$ and $y(t)$ are completely restored from autocorrelation $R[n] = R(nT)$ and cross-correlation $R_{xy}[n] = R_{xy}(nT)$, respectively, of time samples of the same processes $x[n] = x(nT)$ and $y[n] = y(nT)$, provided that their spectra are confined within a passband $mB \leq |\nu| < (m+1)B$ and the sampling interval T is equal to the Nyquist interval, $T = 1/(2B)$, where m, n are integers, ν is frequency, and B is bandwidth. But how can we actually express $R(\tau)$ and $R_{xy}(\tau)$ via $R[n]$ and $R_{xy}[n]$?

In a case of the baseband spectrum with $m = 0$, the answer is given in the so-called “**second part of the sampling theorem**”, which states that the correlations satisfy equations:

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

and

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

The sinc function here:

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

is called the “**sampling function**”.

Proof:

We prove equation (34) only, since equation (35) can be proven exactly in the same way.

According to equation (32), power spectrum $S(\omega)$ of a stationary continuous-time random 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 a relation:

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

Therefore, if the sampling interval T is equal to the Nyquist interval, $T = 1/(2B)$, we obtain the full continuous-time baseband spectrum $S(\omega)$ by multiplying to the spectrum of the time sample $S_D(\omega)$ a rectangular window function $P(\omega)$ which is defined by

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

as illustrated in Figure 10. In fact, we have

$$S(\omega) = T P(\omega) S_D(\omega), \quad (38)$$

in view of equation (32).

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

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

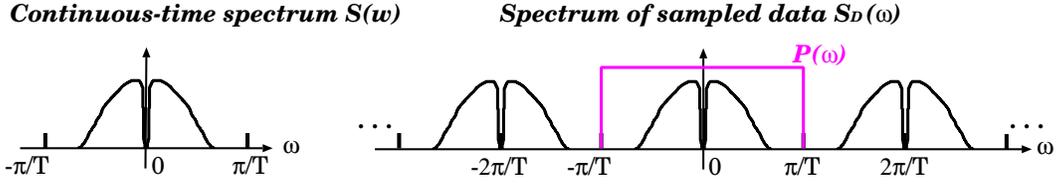


Figure 10: Power spectrum $S(\omega)$ of a continuous-time random process (left panel) is fully restored from power spectrum $S_D(\omega)$ of its time sample (right panel) by multiplying a rectangular window function $P(\omega)$ given in equation (37).

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:

$$p(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} P(\omega) e^{i\omega\tau} d\omega = \frac{1}{2\pi} \int_{-\pi/T}^{\pi/T} e^{i\omega\tau} d\omega = \frac{\sin\left(\frac{\pi\tau}{T}\right)}{\pi\tau}. \quad (40)$$

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

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

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

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

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

We can easily extend the second part of the sampling theorem to the more general case of the passband spectra confined within $mB \leq |\nu| < (m+1)B$. Indeed, all above discussions are valid also in this case if we only replace the rectangular window function $P(\omega)$ in equation (37) and its inverse Fourier transform $p(\tau)$ in equation (40) by new ones for the passband case. Specifically, $P(\omega)$ is now given by

$$P(\omega) = \begin{cases} 1 & \frac{m\pi}{T} \leq |\omega| < \frac{(m+1)\pi}{T}, \\ 0 & \text{otherwise,} \end{cases}$$

and therefore $p(\tau)$ is reduced to

$$p(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} P(\omega) e^{i\omega\tau} d\omega = \frac{1}{\pi} \Re \int_{\frac{m\pi}{T}}^{\frac{(m+1)\pi}{T}} e^{i\omega\tau} d\omega = \frac{\sin\left(\frac{\pi\tau}{2}\right)}{\frac{\pi\tau}{2}} \cos\left[\left(m + \frac{1}{2}\right) \frac{\pi\tau}{T}\right].$$

Note that we calculated a similar integral when we discussed the white fringe in Chapter 3.

Consequently, equations (34) and (35) are generalized to

$$R(\tau) = \sum_{n=-\infty}^{\infty} R[n] \frac{\sin\left[\frac{\pi}{2T}(\tau - nT)\right]}{\frac{\pi}{2T}(\tau - nT)} \cos\left[\left(m + \frac{1}{2}\right) \frac{\pi}{T}(\tau - nT)\right],$$

and

$$R_{xy}(\tau) = \sum_{n=-\infty}^{\infty} R_{xy}[n] \frac{\sin\left[\frac{\pi}{2T}(\tau - nT)\right]}{\frac{\pi}{2T}(\tau - nT)} \cos\left[\left(m + \frac{1}{2}\right) \frac{\pi}{T}(\tau - nT)\right],$$

in the passband case.

Now we are in position to answer to an interesting question: “Why accuracy of delay determination in VLBI can be much superior (i.e. shorter) than a sampling interval of digitized voltage signals, from which the delay is determined?” For example, we will see later that typical delay accuracy of modern VLBI systems is better than 0.1 nanosecond (1×10^{-10} sec), even when we sample the received voltages with 125 nanosecond interval (which is the Nyquist interval for a 2 MHz bandwidth: $1/(2 \times 2 \text{ MHz})$).

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

original continuous-time data are, irrespective of actual length of the Nyquist interval, since they are equivalent to each other as the sampling theorem states.

The $2B$ optimal rate and the sampling function have been independently discovered by a number of researchers in different countries, in addition to 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 Spectra

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

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

where n is an integer, and $n = 0$ corresponds to the particular case of the baseband spectrum (Figure 11).

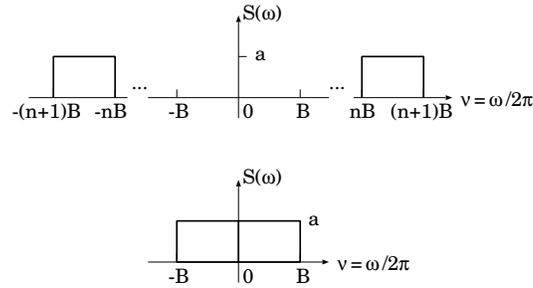


Figure 11: Rectangular passband (top) and baseband (bottom) spectra.

If we sample the data with Nyquist interval $T = 1/(2B)$, then how do the auto- and cross-correlations of the time samples look like?

Autocorrelation:

Autocorrelation $R(\tau)$ of an original continuous-time process $x(t)$ is obtained by the inverse Fourier transformation of the passband power spectrum $S(\omega)$ (top panel of Figure 11):

$$R(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S(\omega) e^{i\omega\tau} d\omega = \frac{a}{\pi} \Re \int_{2\pi nB}^{2\pi(n+1)B} e^{i\omega\tau} d\omega$$

$$= 2aB \frac{\sin(\pi B\tau)}{\pi B\tau} \cos \left[2\pi \left(n + \frac{1}{2} \right) B\tau \right], \quad (43)$$

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 11), we have $n = 0$, and the autocorrelation of the continuous-time process has a simple sinc function form:

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

For the correlation coefficient of the continuous-time process:

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

we have,

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

in the case of the passband spectrum, and

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

in the 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

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

for the passband spectrum, and

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

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

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

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 12.

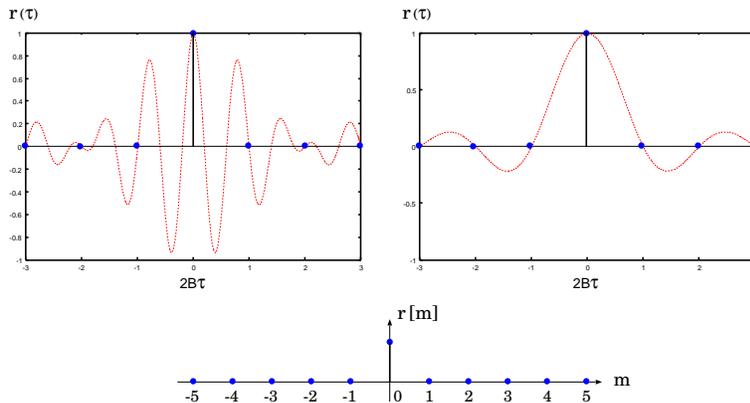


Figure 12: Correlation coefficient $r(\tau)$ of a continuous-time process with rectangular passband spectrum when $n = 2$ (top left) and baseband spectrum (top right), as given in equations (45) and (46). When this process is sampled with the Nyquist interval $T = 1/(2B)$, correlation coefficient $r[m]$ 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.

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 11, 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:

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

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 $R_{xy}(\tau)$:

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

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

$$\begin{aligned} r_{xy}(\tau) &= \frac{R_{xy}(\tau)}{\sqrt{R_{xx}(0) R_{yy}(0)}} \\ &= \begin{cases} \rho \frac{\sin[\pi B (\tau - \tau_d)]}{\pi B (\tau - \tau_d)} \cos \left[2\pi \left(n + \frac{1}{2} \right) B (\tau - \tau_d) \right] & \text{(passband),} \\ \rho \frac{\sin[2\pi B (\tau - \tau_d)]}{2\pi B (\tau - \tau_d)} & \text{(baseband),} \end{cases} \end{aligned} \quad (51)$$

where ρ is the maximum cross-correlation coefficient:

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

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”.

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

$$\begin{aligned} r_{xy}[m] &= r_{xy}(mT) \\ &= \begin{cases} \rho \frac{\sin \left[\frac{\pi}{2} \left(m - \frac{\tau_d}{T} \right) \right]}{\frac{\pi}{2} \left(m - \frac{\tau_d}{T} \right)} \cos \left[\pi \left(n + \frac{1}{2} \right) \left(m - \frac{\tau_d}{T} \right) \right] & \text{(passband),} \\ \rho \frac{\sin \left[\pi \left(m - \frac{\tau_d}{T} \right) \right]}{\pi \left(m - \frac{\tau_d}{T} \right)} & \text{(baseband).} \end{cases} \end{aligned} \quad (53)$$

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

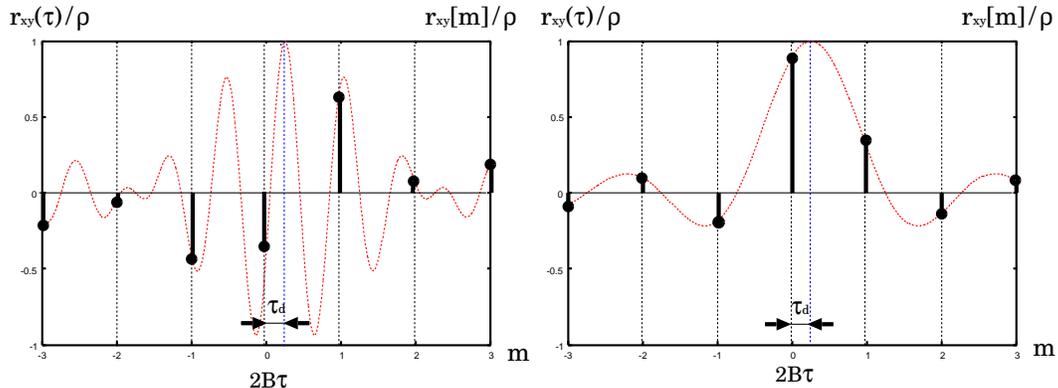


Figure 13: Cross-correlation coefficient $r_{xy}(\tau)$ of jointly stationary continuous-time processes $x(t)$ and $y(t)$ with rectangular passband spectrum when $n = 2$ (left: dotted line) and baseband spectrum with $n = 0$ (right: dotted line), as given by equation (51). τ_d is a time delay between 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.

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 12, 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 $x[i] = x(iT)$ and $y[i] = y(iT)$. 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 :

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

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

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

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

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

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 $x[i] = x(iT)$ and $y[j] = y(jT)$ obey the joint Gaussian probability distribution. Thus, we have

$$\begin{aligned} \langle \mathcal{R}_s^2 \rangle &= \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N \langle x[i] y[i] x[j] y[j] \rangle \\ &= \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N \{ \langle x[i] y[i] \rangle \langle x[j] y[j] \rangle \\ &\quad + \langle x[i] x[j] \rangle \langle y[i] y[j] \rangle + \langle x[i] y[j] \rangle \langle y[i] x[j] \rangle \} \end{aligned}$$

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

where $\rho = \frac{R_{xy}[0]}{\sqrt{R_{xx}[0] R_{yy}[0]}} = \frac{R_{xy}(0)}{\sqrt{R_{xx}(0) R_{yy}(0)}}$ 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:

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

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 S/N :

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

Here we denoted amplitudes of cross-correlation and cross-correlation coefficient as $|R_{xy}[0]|$ and $|\rho|$, respectively, regarding as if they are complex quantities derived from a complex correlator.

As we noticed in Chapter 3, the autocorrelations $R_{xx}(0)$ and $R_{yy}(0)$ in ρ are usually dominated by system noise contributions in a radio interferometer, while the cross-correlation $R_{xy}(0)$ contains contribution of the signal from a radio source only. Then, signal-to-noise ratio discussions in Chapter 3 showed that the maximum amplitude of the cross-correlation coefficient ρ is approximately given by

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

when we observe a continuum spectrum source, where T_{A_1} , T_{A_2} are antenna temperatures, 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

$$S/N = \frac{|\langle \mathcal{R}_s \rangle|}{\sigma_s} = |\rho| \sqrt{N} = \sqrt{\frac{T_{A_1} T_{A_2}}{T_{S_1} T_{S_2}}} \sqrt{N}. \quad (63)$$

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

$$N = \frac{\tau_a}{t_s} = 2 B \tau_a. \quad (64)$$

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

$$S/N = \sqrt{\frac{T_{A_1} T_{A_2}}{T_{S_1} T_{S_2}}} \sqrt{2 B \tau_a}, \quad (65)$$

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

$$W_\nu = k T, \quad (66)$$

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

$$E = B t k T. \quad (67)$$

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

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

1.1.17 Higher–Order Sampling in VLBI Receiving Systems

In digital data processings as applied to radio astronomy, sampling of received voltages has been traditionally done at the basebands (or the videobands), 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.

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) and Korean KVN (Korean VLBI Network). 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 frequency 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 14 shows a diagram of KVN 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.

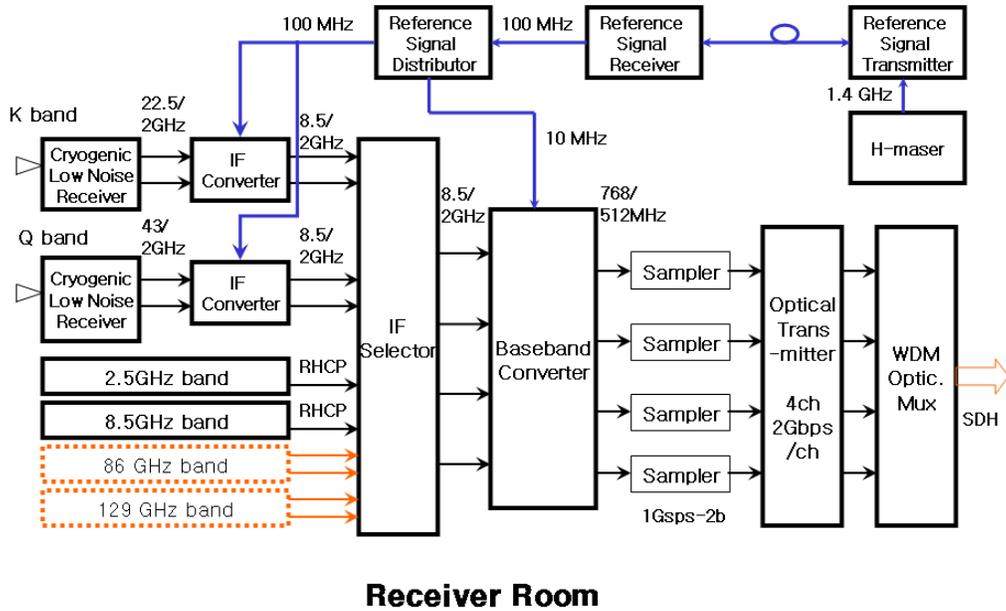


Figure 14: Diagram of receiving system in KVN (Korean VLBI Network) adopting the higher-order sampling technique (figure brought from KVN webpage <http://www.trao.re.kr/kvn/>).

Note here that, for the bandwidth $B = 512$ MHz, the passband $nB \leq \nu < (n + 1)B$ with the lower edge $nB = 512$ MHz means $n = 1$, i.e. an odd number of n , 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 7. 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 Kawaguchi clearly explains in the following way.

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, i.e. the time samples take continuously variable values 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 (“levels”) expressible by a certain number of bits. This step is called “**clipping**” or “**quantization**”.

Number of discrete values expressible by a given number of bits determines “number of levels of quantization”. Therefore, 1-bit, 2-bit, \dots , n -bit quantizations usually correspond to 2-level, 4-level, \dots , 2^n -level quantizations, respectively.

There have been exceptions to the n -bit & 2^n -level law. For example, 3-level quantization using 2 bits has been proposed for getting a high sensitivity without too much complicating logics of processing. In the present-day VLBI systems, however, such exceptional quantization schemes are rarely used.

We will denote henceforth a discrete-time process composed of quantized values as $\hat{x}[i]$. If number of quantization levels is m , with discrete values x_1, x_2, \dots, x_m , then $\hat{x}[i]$ must take one of these m values as illustrated in Figure

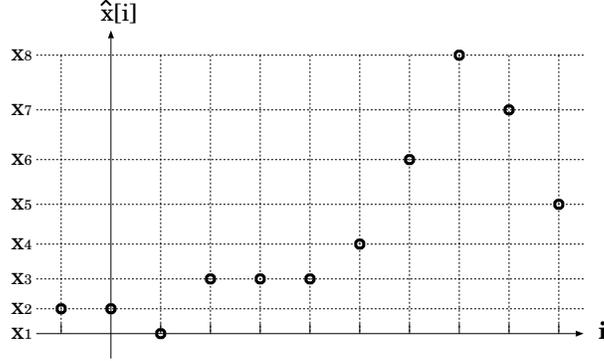


Figure 15: An image of a quantized discrete-time process.

15. A quantized process $\hat{x}[i]$ is supposed to be related in a prescribed way to an original unclipped discrete-time process $x[i]$ having analog values.

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 modern VLBI, 1-bit, 2-level and 2-bit, 4-level quantizations are mostly used.

Figure 16 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.

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 assignments 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

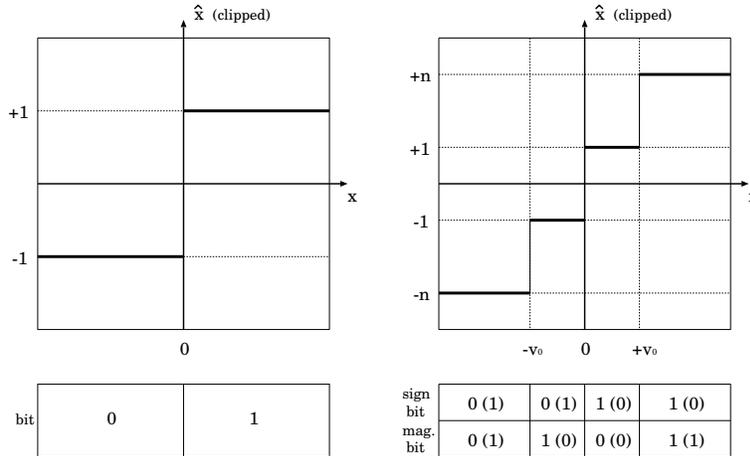


Figure 16: Relations between analog (unquantized) values x and clipped (quantized) values \hat{x} for 1-bit (left) and 2-bit (right) quantizations, respectively. Bit assignments are shown in bottom panels. In case of 2-bit quantization (left), a representative bit assignment for data recorder is shown along with that for a correlator chip given in parentheses.

values: $-n$, -1 , $+1$, and $+n$, as shown in left panel of Figure 16 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. On the other hand, a widely used correlator chip for 2-bit quantized data adopts 11, 10, 00, and 01 assignments, as shown in Figure 16 and Table 1. Therefore, recorded bits should be rearranged before correlation processing, when we use a correlator equipped 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. The simplest way is to use the 1-bit correlation mode, by just picking up sign bits only from the 2-bit quantized data, though we have to sacrifice a part of information contained in the 2-bit quantized data. Of course, direct cross-correlation of data with different quantization schemes, such as 1-bit and 2-bit, is also possible with no loss of information (see, for example, Hagen and Farley, 1973), provided that a special logical circuit for this purpose is

built in a correlator.

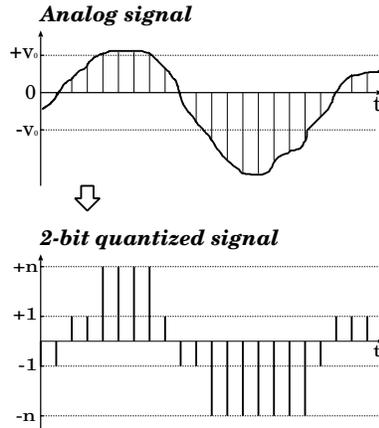


Figure 17: A schematic view of time variation of original analog data (top) and 2-bit quantized data (bottom).

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 2. In the 2-bit quantized data, some information on the magnitude (amplitude) of the analog data is left, as illustrated in Figure 17, 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, we consider that the data $x(t)$ and $y(t)$ from two antennas are jointly stationary continuous-time random processes, 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, the zero-mean second-order Gaussian probability density of jointly stationary continuous-time random processes $x(t + \tau)$ and $y(t)$ is given by

$$\begin{aligned} f(x, y; \tau) &\equiv f(x, y; t + \tau, t) \\ &= \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-r_{xy}^2(\tau)}} e^{-\frac{1}{2[1-r_{xy}^2(\tau)]} \left[\frac{x^2}{\sigma_x^2} - 2r_{xy}(\tau) \frac{xy}{\sigma_x\sigma_y} + \frac{y^2}{\sigma_y^2} \right]}, \end{aligned} \quad (68)$$

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

$$r_{xy}(\tau) \equiv \frac{R_{xy}(\tau)}{\sqrt{R_{xx}(0)R_{yy}(0)}}, \quad (69)$$

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 occur when $r_{xy}^2(\tau) = 1$.

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

$$R_{xy}(\tau) = \langle x(t + \tau) y(t) \rangle = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x y f(x, y; \tau) dx dy, \quad (70)$$

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 density 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 18). Here, the condition “for any t ” corresponds to our case of the stationary random processes.

Now let us 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

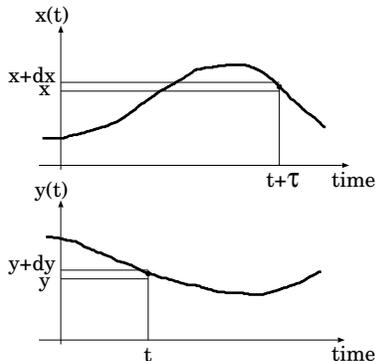


Figure 18: In case of analog continuous-time processes, joint probability density describes probability for the processes $x(t)$ and $y(t)$ to take values within infinitesimal ranges from x to $x + dx$ and from y to $y + dy$ at periods of time $t + \tau$ and t .

interval. Then, 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

$$R_{xy}[m] = R_{xy}(mT) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x y f(x, y; mT) dx dy. \quad (71)$$

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 , x_1, x_2, \dots, x_N and y_1, y_2, \dots, y_N , in case of N -level quantization.

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

$$R_{\hat{x}\hat{y}}[m] = \sum_{i=1}^N \sum_{j=1}^N x_i y_j P(x_i, y_j; m), \quad (72)$$

where $P(x_i, y_j; m)$ is a joint probability for clipped discrete-time processes $\hat{x}[l]$ and $\hat{y}[l]$ to take values:

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

at arguments $n + m$ and n for any n (see Figure 19).

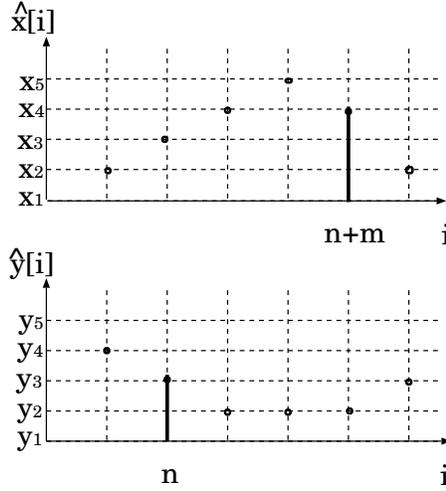


Figure 19: Joint probability $P(x_i, y_j; m)$ of clipped processes $\hat{x}[n+m]$ and $\hat{y}[n]$ describes probability for them to take discrete values x_i and y_j of i -th and j -th quantization levels at arguments $n+m$ and n .

Note that this cross-correlation is what we can expect to obtain as an output of a digital correlator.

In fact, a digital correlator is nothing but a machine which time-averages a large number of products of digital (i.e. sampled and clipped) data. Therefore, if we time-shift one data with respect to another by m , the correlator yields

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

for \mathcal{N} products of digital data. In view of the ergodicity, this should well approximate the cross-correlation $R_{\hat{x}\hat{y}}[m]$:

$$R_{\hat{x}\hat{y}}[m] \cong \frac{1}{\mathcal{N}} \sum_{n=1}^{\mathcal{N}} (\hat{x}[m+n] \hat{y}[n]), \quad (73)$$

if \mathcal{N} is sufficiently large.

Consequently, we can easily obtain the cross-correlation $R_{\hat{x}\hat{y}}[m]$ of the clipped data $\hat{x}[n]$ and $\hat{y}[n]$, using a digital correlator. But how can we relate this to the cross-correlation $R_{xy}[m]$ of the original analog data $x[n]$ and $y[n]$ which, in most cases, we actually wish to know?

At the first glance, equations (71) and (72) look quite different, and it appears difficult to relate $R_{\hat{x}\hat{y}}[m]$ to $R_{xy}[m]$. However, if the quantization

condition is clearly specified, we can calculate $P(x_i, y_j; m)$ rather easily from $f(x, y; mT)$ as far as the probability density is jointly Gaussian (van Vleck and Middleton, 1966).

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

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

$$\hat{y}[i] = \begin{cases} +1 & \text{for } y[i] = y(iT) \geq 0, \\ -1 & \text{for } y[i] = y(iT) < 0. \end{cases} \quad (74)$$

Therefore, the probability for $\hat{x}[i]$ to be +1, for example, is 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:

$$P(+1, +1; m) = \int_0^{+\infty} \left\{ \int_0^{+\infty} f(x, y; mT) dx \right\} dy,$$

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

$$P(+1, -1; m) = \int_{-\infty}^0 \left\{ \int_0^{+\infty} f(x, y; mT) dx \right\} dy,$$

$$P(-1, +1; m) = \int_0^{+\infty} \left\{ \int_{-\infty}^0 f(x, y; mT) dx \right\} dy. \quad (75)$$

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

$$N = 2, \quad \text{and} \quad \begin{cases} x_1 = -1, & x_2 = +1, \\ y_1 = -1, & y_2 = +1. \end{cases} \quad (76)$$

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

$$\begin{aligned}
R_{\hat{x}\hat{y}}[m] &= \sum_{i=1}^2 \sum_{j=1}^2 x_i y_j P(x_i, y_j; m) \\
&= (+1) \cdot (+1) \cdot P(+1, +1; m) + (-1) \cdot (-1) \cdot P(-1, -1; m) \\
&\quad + (+1) \cdot (-1) \cdot P(+1, -1; m) + (-1) \cdot (+1) \cdot P(-1, +1; m) \\
&= P(+1, +1; m) + P(-1, -1; m) \\
&\quad - P(+1, -1; m) - P(-1, +1; m). \tag{77}
\end{aligned}$$

Because of symmetric properties of the joint Gaussian probability density shown in equation (68), the joint probabilities given in equation (75) satisfy

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

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

$$\begin{aligned}
P(+1, +1; m) + P(-1, -1; m) + P(+1, -1; m) + P(-1, +1; m) \\
= 2P(+1, +1; m) + 2P(+1, -1; m) = 1. \tag{79}
\end{aligned}$$

Then, from equations (77), (78), and (79), we obtain

$$\begin{aligned}
R_{\hat{x}\hat{y}}[m] &= P(+1, +1; m) + P(-1, -1; m) \\
&\quad - P(+1, -1; m) - P(-1, +1; m) \\
&= 2P(+1, +1; m) - 2P(+1, -1; m) \\
&= 4P(+1, +1; m) - 1. \tag{80}
\end{aligned}$$

Substituting the explicit form of the joint Gaussian probability density in equation (68) into equation (75), we calculate $4P(+1, +1; m)$:

$$\begin{aligned}
4P(+1, +1; m) &= 4 \int_0^{+\infty} \left\{ \int_0^{+\infty} f(x, y; mT) dx \right\} dy \\
&= \int_0^{+\infty} \left\{ \int_0^{+\infty} \frac{2}{\pi \sigma_x \sigma_y \sqrt{1 - r_{xy}^2}} e^{-\frac{1}{2(1 - r_{xy}^2)} \left(\frac{x^2}{\sigma_x^2} - 2r_{xy} \frac{xy}{\sigma_x \sigma_y} + \frac{y^2}{\sigma_y^2} \right)} dx \right\} dy, \tag{81}
\end{aligned}$$

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

$$\begin{aligned} x &= \sigma_x \zeta \cos \phi, \\ y &= \sigma_y \zeta \sin \phi. \end{aligned} \quad (82)$$

Then the above integral is reduced to

$$\begin{aligned} 4P(+1, +1; m) &= \frac{2}{\pi \sqrt{1-r_{xy}^2}} \int_0^{\frac{\pi}{2}} d\phi \int_0^\infty e^{-\frac{\zeta^2(1-r_{xy} \sin 2\phi)}{2(1-r_{xy}^2)}} \zeta d\zeta \\ &= \frac{2}{\pi} \int_0^{\frac{\pi}{2}} \frac{\sqrt{1-r_{xy}^2}}{1-r_{xy} \sin 2\phi} d\phi. \end{aligned} \quad (83)$$

If we further introduce another new variable θ , which satisfies

$$\tan \theta = \frac{\tan \phi - r_{xy}}{\sqrt{1-r_{xy}^2}}, \quad \text{and, therefore,} \quad \theta = \arctan \left(\frac{\tan \phi - r_{xy}}{\sqrt{1-r_{xy}^2}} \right), \quad (84)$$

then we have

$$\frac{d\theta}{d\phi} = \frac{1}{1 + \left(\frac{\tan \phi - r_{xy}}{\sqrt{1-r_{xy}^2}} \right)^2} \frac{1}{\sqrt{1-r_{xy}^2}} \frac{1}{\cos^2 \phi} = \frac{\sqrt{1-r_{xy}^2}}{1-r_{xy} \sin 2\phi}. \quad (85)$$

Note that this has the same form as the integrand of equation (83). The limits of the integration $\phi = 0$ and $\phi = \frac{\pi}{2}$ now correspond to

$\theta = \theta_0 \equiv -\arctan \left(\frac{r_{xy}}{\sqrt{1-r_{xy}^2}} \right)$ and $\theta = \frac{\pi}{2}$, respectively. Therefore, equation (83) becomes

$$4P(+1, +1; m) = \frac{2}{\pi} \int_{\theta_0}^{\frac{\pi}{2}} d\theta = 1 - \frac{2}{\pi} \theta_0 = 1 + \frac{2}{\pi} \arctan \left(\frac{r_{xy}}{\sqrt{1-r_{xy}^2}} \right). \quad (86)$$

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

$$4P(+1, +1; m) = 1 + \frac{2}{\pi} \arctan(\tan \xi) = 1 + \frac{2}{\pi} \xi = 1 + \frac{2}{\pi} \arcsin(r_{xy}). \quad (87)$$

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:

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

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 (87) must be confined within a range:

$$-\frac{\pi}{2} \leq \arcsin(r_{xy}) \leq \frac{\pi}{2}. \quad (88)$$

Finally, from equation (80), we obtain

$$\begin{aligned} R_{\hat{x}\hat{y}}[m] &= 4P(+1, +1; m) - 1 = \frac{2}{\pi} \arcsin(r_{xy}[m]) \\ &= \frac{2}{\pi} \arcsin \left(\frac{R_{xy}(mT)}{\sqrt{R_{xx}(0) R_{yy}(0)}} \right). \end{aligned} \quad (89)$$

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:

$$r_{\hat{x}\hat{y}}[m] = \frac{R_{\hat{x}\hat{y}}[m]}{\sqrt{R_{\hat{x}\hat{x}}[0] R_{\hat{y}\hat{y}}[0]}} = R_{\hat{x}\hat{y}}[m]. \quad (90)$$

Therefore, equation (89) is described also as

$$r_{\hat{x}\hat{y}}[m] = \frac{2}{\pi} \arcsin(r_{xy}[m]). \quad (91)$$

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

$$R_{\hat{x}\hat{y}}[m] = \frac{2}{\pi} r_{xy}[m], \quad \text{or} \quad r_{\hat{x}\hat{y}}[m] = \frac{2}{\pi} r_{xy}[m]. \quad (92)$$

Although we derived equations (89) and (91) for $r_{xy}^2 \neq 1$ case, it is worth to confirm here that the resultant equations are valid in limiting cases of complete correlation ($r_{\hat{x}\hat{y}} = 1$ when $r_{xy} = 1$) and complete anti-correlation ($r_{\hat{x}\hat{y}} = -1$ when $r_{xy} = -1$), too.

Equation (89) or (91) 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:

$$r_{xy}[m] = \sin\left(\frac{\pi}{2} r_{\hat{x}\hat{y}}[m]\right). \quad (93)$$

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

$$r_{xy}[m] = \frac{\pi}{2} r_{\hat{x}\hat{y}}[m]. \quad (94)$$

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 (93) and (94) mean that, for Gaussian random processes **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 (see, as an example, Figure 20), though the amplitude is reduced by a factor of $\cong 2/\pi$.

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 $R_{xx}(0)$ 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 (75) 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”.

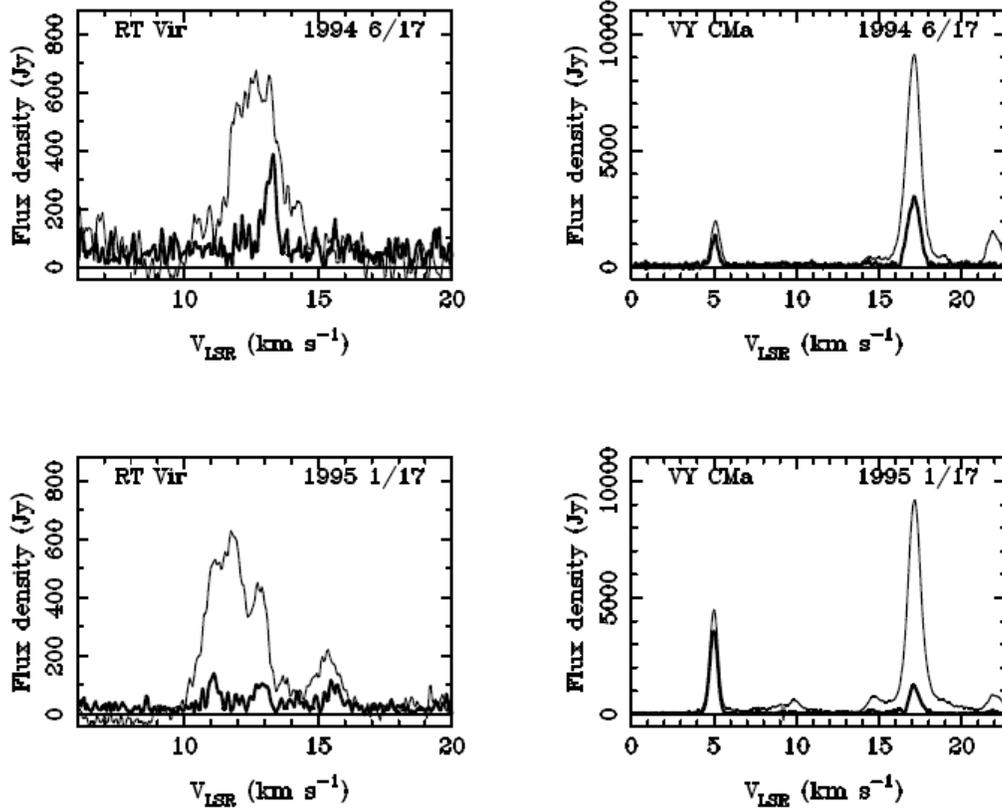


Figure 20: Spectra of water maser lines derived from 1-bit quantized data. Right and left panels show water maser lines of Asymptotic Giant Branch (AGB) stars RT Vir and VY CMa, respectively, obtained at two epochs (top and bottom panels). Thin lines show total-power (i.e. autocorrelation) spectra obtained 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 1300 km baseline. They exhibit large time variations within 7 months. (Figure courtesy of Imai et al., *Astron. Astrophys*, **317**, L67-L70, 1997.)

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 :

$$f(x_1, x_2; \tau) = \frac{1}{2\pi\sigma_x^2\sqrt{1-r_{xx}^2(\tau)}} e^{-\frac{1}{2(1-r_{xx}^2(\tau))} \left(\frac{x_1^2 - 2r_{xx}(\tau)x_1x_2 + x_2^2}{\sigma_x^2} \right)}, \quad (95)$$

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

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

$$r_{\hat{x}\hat{x}}[m] = R_{\hat{x}\hat{x}}[m] = \frac{2}{\pi} \arcsin(r_{xx}[m]) = \frac{2}{\pi} \arcsin\left(\frac{R_{xx}(mT)}{R_{xx}(0)}\right), \quad (96)$$

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 1-Bit Quantized Data

The derivation of the van Vleck relationship, as we saw above, is quite general and applicable to any data satisfying Gaussian probability density. In particular, the relationship is well valid for continuous-time data, although we have so far dealt with only sampled data in the context of our discussion on digitization of analog data.

Hence, we generally have

$$R_{\hat{x}\hat{y}}(\tau) = \frac{2}{\pi} \arcsin(r_{xy}(\tau)), \quad (97)$$

for the cross-correlation, and

$$R_{\hat{x}\hat{x}}(\tau) = \frac{2}{\pi} \arcsin(r_{xx}(\tau)), \quad \text{and} \quad R_{\hat{y}\hat{y}}(\tau) = \frac{2}{\pi} \arcsin(r_{yy}(\tau)), \quad (98)$$

for the autocorrelations, with continuous 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 transformation:

$$\begin{aligned}
S_{\hat{x}\hat{y}}(\omega) &= \int_{-\infty}^{\infty} R_{\hat{x}\hat{y}}(\tau) e^{-i\omega\tau} d\tau = \frac{2}{\pi} \int_{-\infty}^{\infty} \arcsin(r_{xy}(\tau)) e^{-i\omega\tau} d\tau, \\
S_{\hat{x}\hat{x}}(\omega) &= \int_{-\infty}^{\infty} R_{\hat{x}\hat{x}}(\tau) e^{-i\omega\tau} d\tau = \frac{2}{\pi} \int_{-\infty}^{\infty} \arcsin(r_{xx}(\tau)) e^{-i\omega\tau} d\tau, \\
S_{\hat{y}\hat{y}}(\omega) &= \int_{-\infty}^{\infty} R_{\hat{y}\hat{y}}(\tau) e^{-i\omega\tau} d\tau = \frac{2}{\pi} \int_{-\infty}^{\infty} \arcsin(r_{yy}(\tau)) e^{-i\omega\tau} d\tau. \quad (99)
\end{aligned}$$

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. Then, how spectra of the data will look like after clipping?

Figure 21 shows a power spectrum $S_{\hat{x}\hat{x}}(\omega)$ of 1-bit quantized data (solid line in bottom panel), which is obtained by clipping analog data with a rectangular power spectrum $S_{xx}(\omega)$ of bandwidth B (shown by broken line in bottom panel). The power spectrum of the analog data $S_{xx}(\omega)$ shown here is normalized by the dispersion $\sigma_x^2 = R_{xx}(0)$. In other words, the ‘‘power spectra’’ shown in this Figure are Fourier transforms of the correlation coefficients $r_{xx}(\tau)$ and $r_{\hat{x}\hat{x}}(\tau) = R_{\hat{x}\hat{x}}(\tau)$.

Also shown in Figure 21 are (1) correlation coefficient of the analog data $r_{xx}(\tau)$ (upper left panel) which shows a sinc function form similar to the white fringe, and (2) correlation coefficient of the clipped data $r_{\hat{x}\hat{x}}(\tau) = R_{\hat{x}\hat{x}}(\tau) = (2/\pi) \arcsin(r_{xx}(\tau))$ (upper right panel).

Although the bulk of the power spectrum after clipping is still contained within the original band $|\nu| \leq B$, 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. In a sense, the low-level skirt compensates reduction of the peak amplitude, since areas under the power spectra (i.e. powers) of the analog and clipped data should be equal to each other because of an obvious relation $r_{\hat{x}\hat{x}}(0) = r_{xx}(0) = 1$.

On the other hand, Figure 22 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 obtained by clipping analog data x and y having a real rectangular cross-power spectrum $S_{xy}(\omega)$ of bandwidth B (broken line in bottom panel). The analog

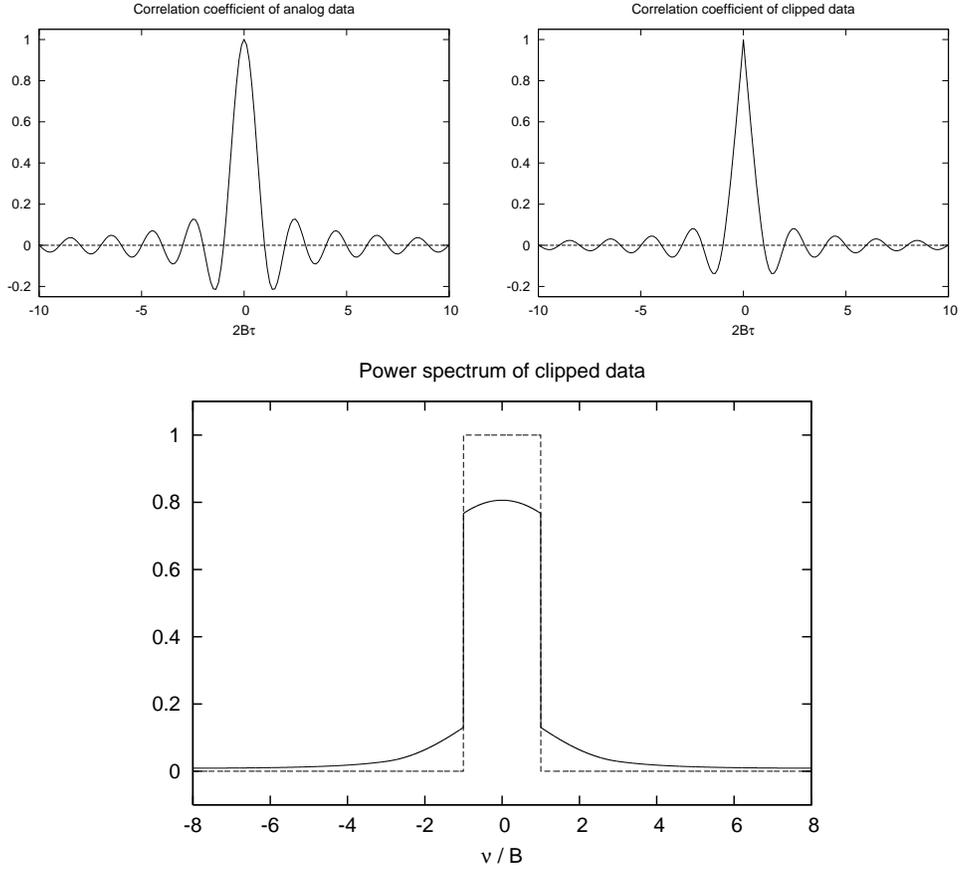


Figure 21: Original rectangular power spectrum $S_{xx}(\omega)$ with bandwidth B (broken line in bottom panel) of analog data x normalized by the dispersion σ_x^2 , and power spectrum $S_{\hat{x}\hat{x}}(\omega)$ (solid line in bottom panel) of clipped data \hat{x} which is derived from the analog data 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.

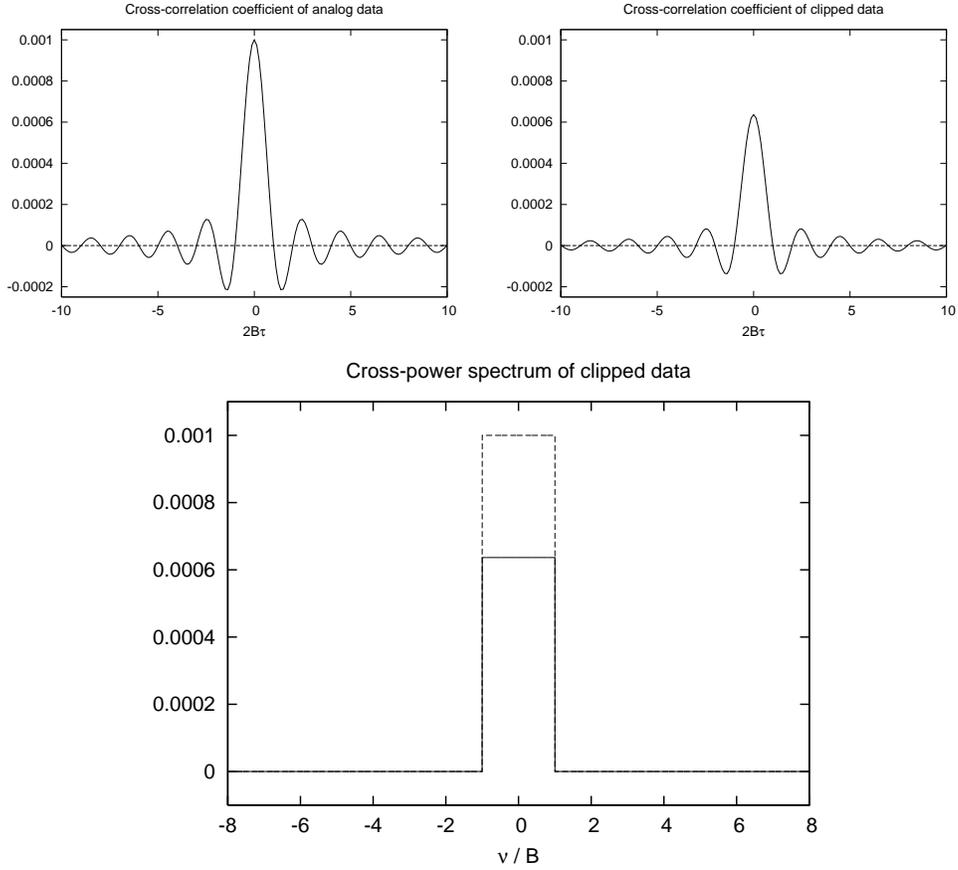


Figure 22: Original rectangular cross-power spectrum $S_{xy}(\omega)$ with bandwidth B (broken line in bottom panel) of analog data x and y normalized by a geometric mean of their dispersions $\sigma_x\sigma_y$, and cross-power spectrum $S_{\hat{x}\hat{y}}(\omega)$ (solid line in bottom panel) of clipped 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 as small as 0.001. Horizontal axes of the upper panels show delay τ normalized by the Nyquist interval ($1 / (2 B)$) of the original analog data.

cross-power spectrum $S_{xy}(\omega)$ shown in Figure 22 is normalized by a geometric mean of dispersions $\sqrt{R_{xx}(0)R_{yy}(0)}$. Therefore, the “cross-power spectra” in this Figure are Fourier transforms of the cross-correlation coefficients $r_{xy}(\tau)$ and $r_{\hat{x}\hat{y}}(\tau) = R_{\hat{x}\hat{y}}(\tau)$. We assume here that the maximum cross-correlation coefficient of the analog data $r_{xy}(\tau)$ is small. For definiteness, we set it to be 0.001 in Figure 22. 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 $r_{\hat{x}\hat{y}}(\tau) = R_{\hat{x}\hat{y}}(\tau) \cong (2/\pi)r_{xy}(\tau)$ (right).

Unlike in the power spectrum case, the cross-power spectrum after clipping remains rectangular, though peak amplitude is reduced by a factor of $2/\pi \cong 0.6366$. This is because $r_{\hat{x}\hat{y}}(\tau)$ is just proportional to $r_{xy}(\tau)$ with the proportionality coefficient $2/\pi$ when $|r_{xy}(\tau)| \ll 1$. Also, areas under the cross-power spectra of the analog and clipped data are not equal to each other, since $r_{\hat{x}\hat{y}}(0) \neq r_{xy}(0)$ as we see in the upper panels of Figure 22.

1.1.23 Price’s Theorem

Now we proceed to the 2-bit quantization case. In principle, we can derive cross-correlations of 2-bit quantized data by evaluating probabilities of quantization states, as we did in the 1-bit quantization case. However, use of Price’s theorem (Price, 1958) offers a simpler derivation.

The theorem states the following.

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

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

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

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

$$\langle g(x, y) \rangle = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f(x, y) dx dy, \quad (100)$$

we have

$$\frac{1}{\sigma_x^n \sigma_y^n} \frac{\partial^n \langle g(x, y) \rangle}{\partial r^n} = \left\langle \frac{\partial^{2n} g(x, y)}{\partial x^n \partial y^n} \right\rangle, \quad (101)$$

(Price, 1958).

Auxiliary Formula

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

$$\frac{1}{\sigma_x^n \sigma_y^n} \frac{\partial^n f(x, y)}{\partial r^n} = \frac{\partial^{2n} f(x, y)}{\partial x^n \partial y^n}, \quad \text{or} \quad \frac{\partial^n f(x', y')}{\partial r^n} = \frac{\partial^{2n} f(x', y')}{\partial x'^n \partial y'^n}, \quad (102)$$

where

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

$$f(x', y') = \frac{1}{2\pi \sigma_x \sigma_y \sqrt{1-r^2}} e^{-\frac{x'^2 - 2r x' y' + y'^2}{2(1-r^2)}}. \quad (103)$$

We verify this formula using the method of mathematical induction.

- For $n = 1$, simple differentiations show

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

and

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

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

- Now, if the formula is valid for $n = m$, i.e.

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

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

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

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

$$\begin{aligned} \frac{1}{\sigma_x \sigma_y} \frac{\partial \langle g(x, y) \rangle}{\partial r} &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g \frac{1}{\sigma_x \sigma_y} \frac{\partial f}{\partial r} dx dy = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g \frac{\partial^2 f}{\partial x \partial y} dx dy \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left[\frac{\partial}{\partial x} \left(g \frac{\partial f}{\partial y} \right) - \frac{\partial}{\partial y} \left(\frac{\partial g}{\partial x} f \right) + \frac{\partial^2 g}{\partial x \partial y} f \right] dx dy \\ &= \int_{-\infty}^{\infty} \left(g \frac{\partial f}{\partial y} \right) \Big|_{x=-\infty}^{x=+\infty} dy - \int_{-\infty}^{\infty} \left(\frac{\partial g}{\partial x} f \right) \Big|_{y=-\infty}^{y=+\infty} dx + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{\partial^2 g}{\partial x \partial y} f dx dy, \end{aligned}$$

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 $n = 1$.

- Now, if the theorem is valid for $n = m$, i.e.

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

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

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

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

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

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 1-Bit Quantaization Case

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

$$\langle g(x, y) \rangle = R_{\hat{x}\hat{y}}[m] = \langle \hat{x}[n+m] \hat{y}[n] \rangle,$$

with

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

In following derivations, we omit arguments m and n for simplicity. We again assume that x and y obey the joint Gaussian probability density:

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

where $r \equiv r_{xy}[m]$ is the analog cross-correlation coefficient.

Now, Price's theorem says

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

Since

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

in the 1-bit quantization case, the RHS of the above equation is equal to

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

Thus we have

$$\frac{\partial R_{\hat{x}\hat{y}}}{\partial r} = \frac{2}{\pi \sqrt{1-r^2}}, \quad \text{and hence} \quad R_{\hat{x}\hat{y}} = \frac{2}{\pi} \int_0^r \frac{dr'}{\sqrt{1-r'^2}} = \frac{2}{\pi} \arcsin(r),$$

where we chose the limits of the integration so that $R_{\hat{x}\hat{y}} = 0$ at $r = 0$. Therefore, we obtained

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

Thus, Price's theorem allows us to derive the van Vleck relationship as given in equation (91) in a really straightforward way.

1.1.24 Cross-Correlation of 2-Bit Quantized Data

Now we consider cross-correlation $R_{\hat{x}\hat{y}}(\tau)$ of 2-bit quantized data $\hat{x}(t)$ and $\hat{y}(t)$. For generality, we regard that t and τ are continuous arguments, though we actually deal with discrete sampled data in our digital processings.

Let clipping criteria for the 2-bit quantization are those given in Figure 16 and Table 1. We assume a simple **equal dispersion** case where dispersions of the original analog data $x(t)$ and $y(t)$ are identical, i.e. $\sigma_x = \sigma_y = \sigma$, and hence the joint probability density is given by

$$f(x, y) \equiv f(x, y; \tau) = \frac{1}{2\pi\sigma^2\sqrt{1-r^2}} e^{-\frac{x^2 - 2rxy + y^2}{2\sigma^2(1-r^2)}}, \quad (104)$$

where $r \equiv r_{xy}(\tau)$ is the cross-correlation coefficient of the analog data. For simplicity, we will often omit the argument τ in following derivations.

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

$$\begin{aligned} \frac{\partial \hat{x}}{\partial x} &= (\mathbf{n} - 1) \delta(x + v_0) + 2 \delta(x) + (\mathbf{n} - 1) \delta(x - v_0), \\ \frac{\partial \hat{y}}{\partial y} &= (\mathbf{n} - 1) \delta(y + v_0) + 2 \delta(y) + (\mathbf{n} - 1) \delta(y - v_0), \end{aligned}$$

where $-v_0$ and $+v_0$ are threshold values between -1 and $-\mathbf{n}$ states and between $+1$ and $+\mathbf{n}$ states, respectively, in the clipping criteria of the 2-bit quantization given in Table 1.

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

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

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

$$\begin{aligned} R_{\hat{x}\hat{y}} &= \frac{1}{\pi} \int_0^r \frac{1}{\sqrt{1-r'^2}} \left\{ (\mathbf{n} - 1)^2 \left[e^{-\frac{v_0^2}{\sigma^2(1+r')}} + e^{-\frac{v_0^2}{\sigma^2(1-r')}} \right] \right. \\ &\quad \left. + 4(\mathbf{n} - 1) e^{-\frac{v_0^2}{2\sigma^2(1-r'^2)}} + 2 \right\} dr'. \end{aligned} \quad (105)$$

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

$$\begin{aligned} R_{\hat{x}\hat{y}}(\tau) &= \frac{1}{\pi} \left[2(\mathbf{n} - 1)^2 e^{-\frac{v_0^2}{\sigma^2}} + 4(\mathbf{n} - 1) e^{-\frac{v_0^2}{2\sigma^2}} + 2 \right] r \\ &= \frac{2}{\pi} [(\mathbf{n} - 1) E + 1]^2 r_{xy}(\tau), \end{aligned} \quad (106)$$

where we introduced a notation

$$E = e^{-\frac{v_0^2}{2\sigma^2}}. \quad (107)$$

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 2-Bit Quantized Data

Now we will 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 $\sigma_{\hat{y}\hat{y}}^2 = R_{\hat{y}\hat{y}}(0)$, in addition to 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:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}}, \quad \text{and} \quad f(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{y^2}{2\sigma^2}}. \quad (108)$$

If we denote probabilities for a 2-bit clipped data $\hat{x}(t)$ (or $\hat{y}(t)$) to take values $-\mathbf{n}$, -1 , $+1$, and $+\mathbf{n}$ of the 4 quantization states as $P(-\mathbf{n})$, $P(-1)$, $P(+1)$, and $P(+\mathbf{n})$, they satisfy

$$P(-\mathbf{n}) + P(-1) + P(+1) + P(+\mathbf{n}) = 1, \quad (109)$$

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

$$R_{\hat{x}\hat{x}}(0) = \mathbf{n}^2 P(-\mathbf{n}) + P(-1) + P(+1) + \mathbf{n}^2 P(+\mathbf{n}). \quad (110)$$

Combining equations (109) and (110), introducing a notation

$$\Phi = P(-1) + P(+1),$$

and repeating the same thing for $\hat{y}(t)$, we obtain

$$R_{\hat{x}\hat{x}}(0) = R_{\hat{y}\hat{y}}(0) = \Phi + \mathbf{n}^2 (1 - \Phi). \quad (111)$$

Using the explicit form of the probability density in equation (108), we can calculate Φ by an equation:

$$\Phi = P(-1) + P(+1) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-v_0}^{+v_0} e^{-\frac{\xi^2}{2\sigma^2}} d\xi = \frac{1}{\sqrt{2\pi}} \int_{-v_0/\sigma}^{+v_0/\sigma} e^{-\frac{\zeta^2}{2}} d\zeta, \quad (112)$$

where $-v_0$ and $+v_0$ are threshold values in the 2-bit quantization as before.

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

$$r_{\hat{x}\hat{y}}(\tau) = \frac{R_{\hat{x}\hat{y}}(\tau)}{\sqrt{R_{\hat{x}\hat{x}}(0) R_{\hat{y}\hat{y}}(0)}} = \frac{R_{\hat{x}\hat{y}}(\tau)}{\Phi + \mathbf{n}^2 (1 - \Phi)}$$

$$\begin{aligned}
&= \frac{1}{\pi [\Phi + \mathbf{n}^2 (1 - \Phi)]} \int_0^{r_{xy}(\tau)} \frac{1}{\sqrt{1 - r'^2}} \left\{ (\mathbf{n} - 1)^2 \left[e^{-\frac{v_0^2}{\sigma^2(1+r')}} + e^{-\frac{v_0^2}{\sigma^2(1-r')}} \right] \right. \\
&\quad \left. + 4(\mathbf{n} - 1) e^{-\frac{v_0^2}{2\sigma^2(1-r'^2)}} + 2 \right\} dr', \tag{113}
\end{aligned}$$

as a function of the cross-correlation coefficient of the original analog data $r = r_{xy}(\tau)$. This is the 2-bit-quantization analogue of the van Vleck relationship given in equation (91).

When $|r_{xy}(\tau)| \ll 1$, which is usually the case in radio interferometry, equations (106) and (111) yield

$$r_{\hat{x}\hat{y}}(\tau) = \frac{2 [(\mathbf{n} - 1) E + 1]^2}{\pi [\Phi + \mathbf{n}^2 (1 - \Phi)]} r_{xy}(\tau). \tag{114}$$

This is an analogue of equation (92) in the 2-bit quantization case.

In subsequent discussions, we will generally assume linear relationship between cross-correlation coefficients of clipped and analog data as given in equations (92) and (114) for the 1-bit and 2-bit quantization cases, respectively, in view of the smallness of cross-correlation coefficients in radio interferometry.

Exactly the same logic leads to the same functional forms as equations (113) and (114) 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

$$\begin{aligned}
r_{\hat{x}\hat{x}}(\tau) &= \frac{1}{\pi [\Phi + \mathbf{n}^2 (1 - \Phi)]} \int_0^{r_{xx}(\tau)} \frac{1}{\sqrt{1 - r'^2}} \left\{ (\mathbf{n} - 1)^2 \left[e^{-\frac{v_0^2}{\sigma^2(1+r')}} + e^{-\frac{v_0^2}{\sigma^2(1-r')}} \right] \right. \\
&\quad \left. + 4(\mathbf{n} - 1) e^{-\frac{v_0^2}{2\sigma^2(1-r'^2)}} + 2 \right\} dr', \tag{115}
\end{aligned}$$

for the general case, and

$$r_{\hat{x}\hat{x}}(\tau) = \frac{2 [(\mathbf{n} - 1) E + 1]^2}{\pi [\Phi + \mathbf{n}^2 (1 - \Phi)]} r_{xx}(\tau), \tag{116}$$

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

Figure 23 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 (91) or (96). The solid line shows the relationship for the 2-bit quantization case as given by equation (113) or (115) for a particular set of parameters $\mathbf{n} = 3$ and $v_0 = 0.996 \sigma$.

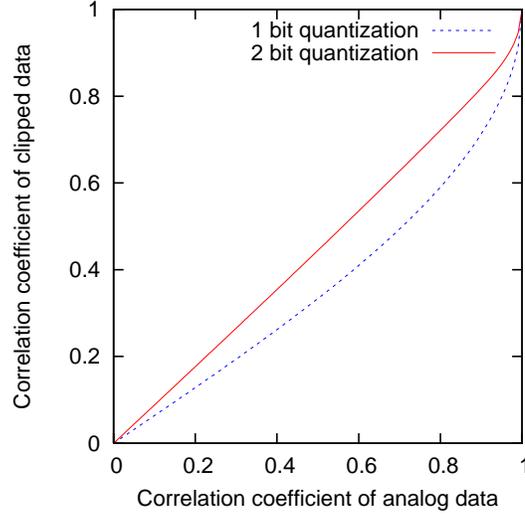


Figure 23: Correlation coefficients of 1-bit (dotted line) and 2-bit (solid line) quantized data with $\mathbf{n} = 3$ and $v_0 = 0.996 \sigma$ as functions of the analog correlation coefficient.

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. In fact, 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 (113), we obtain

$$\begin{aligned}
 S_{\hat{x}\hat{x}}(\omega) &= \int_{-\infty}^{\infty} r_{\hat{x}\hat{x}}(\tau) e^{-i\omega\tau} d\tau = \int_{-\infty}^{\infty} F(r_{xx}(\tau)) e^{-i\omega\tau} d\tau, \\
 S_{\hat{x}\hat{y}}(\omega) &= \int_{-\infty}^{\infty} r_{\hat{x}\hat{y}}(\tau) e^{-i\omega\tau} d\tau = \int_{-\infty}^{\infty} F(r_{xy}(\tau)) e^{-i\omega\tau} d\tau, \quad (117)
 \end{aligned}$$

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

$$F(r) = \frac{1}{\pi (\Phi + \mathbf{n}^2 (1 - \Phi))} \int_0^r \frac{1}{\sqrt{1 - r'^2}} \left\{ (\mathbf{n} - 1)^2 \left[e^{-\frac{v_0^2}{\sigma^2(1+r')}} + e^{-\frac{v_0^2}{\sigma^2(1-r')}} \right] + 4(\mathbf{n} - 1) e^{-\frac{v_0^2}{2\sigma^2(1-r'^2)}} + 2 \right\} dr'.$$

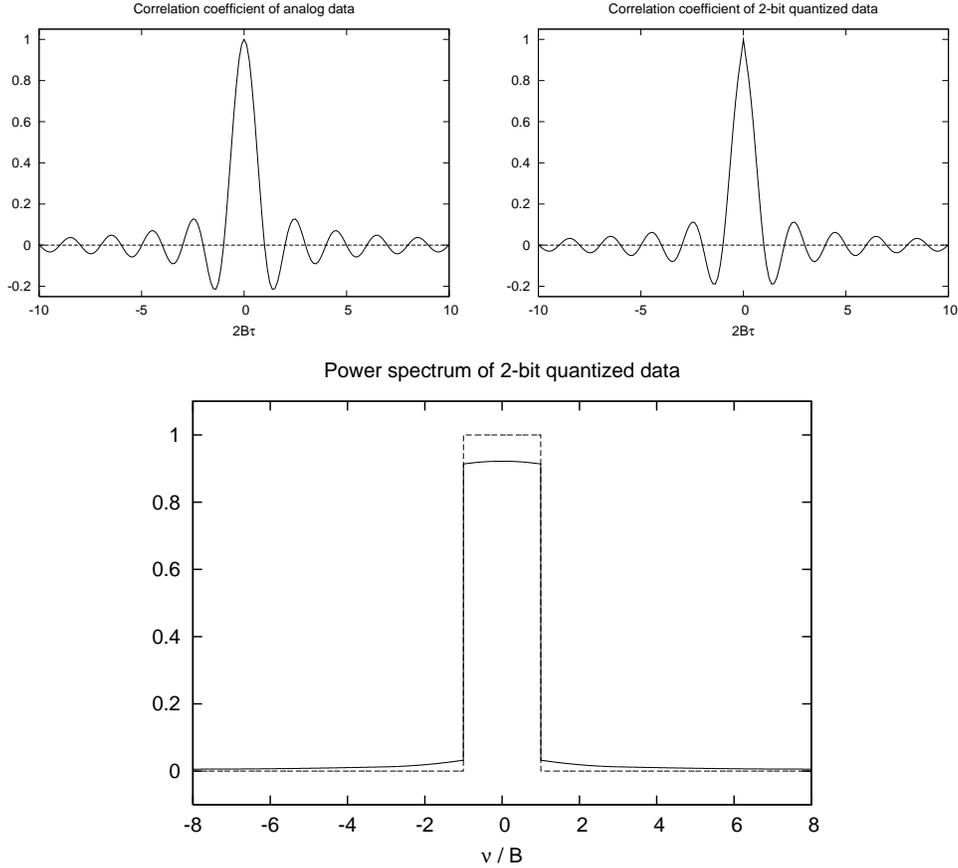


Figure 24: Original rectangular power spectrum $S_{xx}(\omega)$ with bandwidth B (broken line in bottom panel) of the analog data x normalized by the dispersion σ_x^2 , and power spectrum $S_{\hat{x}\hat{x}}(\omega)$ (solid line in bottom panel) of the 2-bit quantized data \hat{x} with $\mathbf{n} = 3$ and $v_0 = 0.996 \sigma$ normalized by the dispersion $\sigma_{\hat{x}}^2$. 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.

Let us again consider the case when analog data have rectangular spectra. Figure 24 shows a power spectrum $S_{\hat{x}\hat{x}}(\omega)$ of the 2-bit quantized data \hat{x} in a case with the set of parameters $\mathbf{n} = 3$ and $v_0 = 0.996 \sigma$ (solid line in bottom

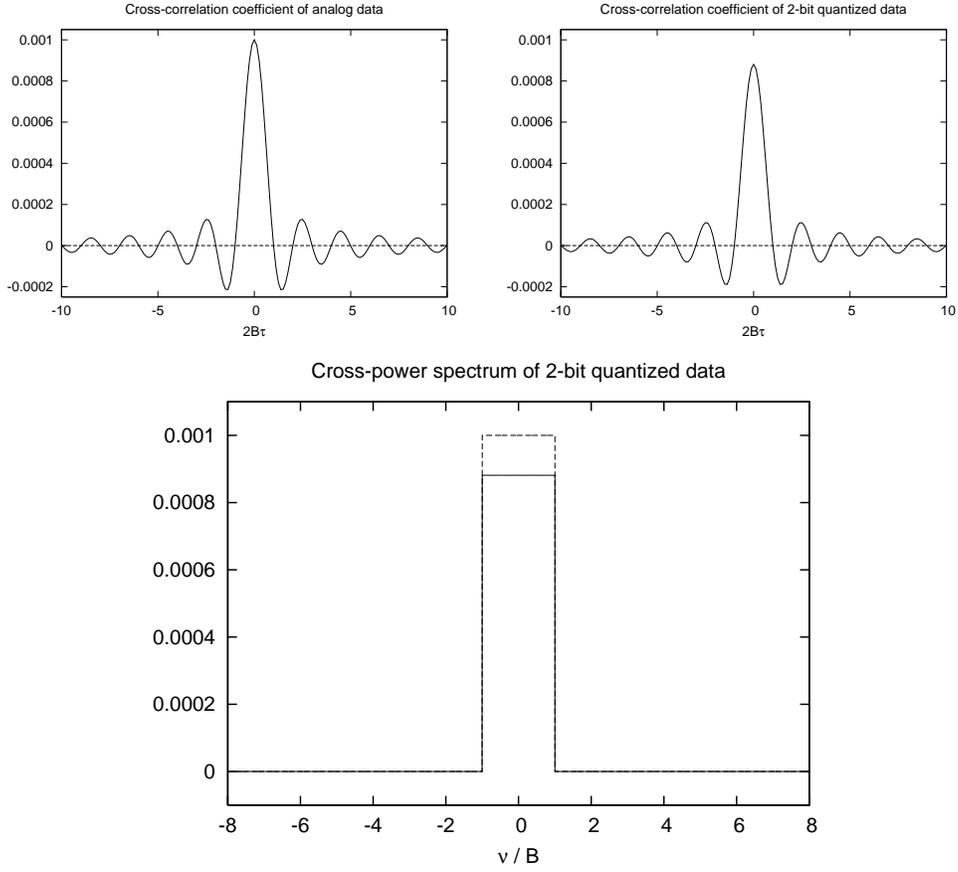


Figure 25: Original rectangular cross-power spectrum $S_{xy}(\omega)$ with bandwidth B (broken line in bottom panel) of the analog data x and y normalized by a geometric mean of their dispersions $\sigma_x\sigma_y$, and cross-power spectrum $S_{\hat{x}\hat{y}}(\omega)$ (solid line in bottom panel) of the 2-bit quantized data \hat{x} and \hat{y} with $\mathbf{n} = 3$ and $v_0 = 0.996\sigma$ normalized by a geometric mean of their dispersions $\sigma_{\hat{x}}\sigma_{\hat{y}}$. 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 quantized 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.

panel), together with a rectangular power spectrum $S_{xx}(\omega)$ with bandwidth B of the original analog data x normalized by the dispersion σ_x^2 (broken line in bottom panel). Similarly to the 1-bit case, areas under the power spectra (i.e. powers) of the analog and clipped data are equal to each other, since $r_{\hat{x}\hat{x}}(0) = r_{xx}(0) = 1$. 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 2-bit quantized data $r_{\hat{x}\hat{x}}(\tau)$ (upper right panel).

The power 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 25 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 $\mathbf{n} = 3$ and $v_0 = 0.996\sigma$ normalized by a geometric mean of their dispersions $\sigma_{\hat{x}}\sigma_{\hat{y}}$ (solid line in bottom panel) and a real rectangular cross-power spectrum $S_{xy}(\omega)$ of bandwidth B of the original analog data x and y normalized by a geometric mean of their dispersions $\sigma_x\sigma_y$ (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. Also, areas under the cross-power spectra of the analog and clipped data are not equal to each other, since $r_{\hat{x}\hat{y}}(0) \neq r_{xy}(0)$.

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 correlator, which we saw earlier? For this purpose, we first consider dispersion of the digital correlator output.

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

$$\mathcal{R} = \frac{1}{\mathcal{N}} \sum_{i=1}^{\mathcal{N}} \hat{x}[i] \hat{y}[i]. \quad (118)$$

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

$$\langle \mathcal{R} \rangle = \langle \hat{x}[i] \hat{y}[i] \rangle = R_{\hat{x}\hat{y}}[0], \quad (119)$$

$$\sigma_{\mathcal{R}}^2 = \langle \mathcal{R}^2 \rangle - \langle \mathcal{R} \rangle^2 = \frac{1}{\mathcal{N}^2} \sum_{i=1}^{\mathcal{N}} \sum_{j=1}^{\mathcal{N}} \langle \hat{x}[i] \hat{y}[i] \hat{x}[j] \hat{y}[j] \rangle - R_{\hat{x}\hat{y}}^2[0]. \quad (120)$$

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 a duration of 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 bandwidth 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 equal to zero, if $m \neq 0$. The same statement is valid for the 1-bit quantized and 2-bit quantized data, since correlation coefficients, and hence correlations, too, of these clipped data must also be equal to zero, as we easily see from equations (91), (96), and (113).

Thus, $R_{\hat{x}\hat{y}}[m]$, $R_{\hat{x}\hat{x}}[m]$, and $R_{\hat{y}\hat{y}}[m]$ are all equal to 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 (120):

$$\sum_{i=1}^{\mathcal{N}} \sum_{j=1}^{\mathcal{N}} \langle \hat{x}[i] \hat{y}[i] \hat{x}[j] \hat{y}[j] \rangle = \sum_{i=1}^{\mathcal{N}} \langle \hat{x}[i] \hat{y}[i] \hat{x}[i] \hat{y}[i] \rangle + \sum_{i=1}^{\mathcal{N}} \sum_{j \neq i}^{\mathcal{N}} \langle \hat{x}[i] \hat{y}[i] \hat{x}[j] \hat{y}[j] \rangle, \quad (121)$$

dominating terms will be

$$\begin{aligned} \sum_{i=1}^{\mathcal{N}} \sum_{j=1}^{\mathcal{N}} \langle \hat{x}[i] \hat{y}[i] \hat{x}[j] \hat{y}[j] \rangle &\cong \sum_{i=1}^{\mathcal{N}} \langle \hat{x}[i] \hat{x}[i] \rangle \langle \hat{y}[i] \hat{y}[i] \rangle + \sum_{i=1}^{\mathcal{N}} \sum_{j \neq i}^{\mathcal{N}} \langle \hat{x}[i] \hat{y}[i] \rangle \langle \hat{x}[j] \hat{y}[j] \rangle \\ &= \mathcal{N} R_{\hat{x}\hat{x}}[0] R_{\hat{y}\hat{y}}[0] + \mathcal{N}(\mathcal{N} - 1) R_{\hat{x}\hat{y}}^2[0] \cong \mathcal{N} R_{\hat{x}\hat{x}}[0] R_{\hat{y}\hat{y}}[0] + \mathcal{N}^2 R_{\hat{x}\hat{y}}^2[0], \end{aligned} \quad (122)$$

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

Now equations (120) and (122) yeild an approximate formula for the dispersion of the digital correlator output:

$$\sigma_{\mathcal{R}}^2 \cong \frac{1}{\mathcal{N}} R_{\hat{x}\hat{x}}[0] R_{\hat{y}\hat{y}}[0] + R_{\hat{x}\hat{y}}^2[0] - R_{\hat{x}\hat{y}}^2[0] = \frac{1}{\mathcal{N}} R_{\hat{x}\hat{x}}[0] R_{\hat{y}\hat{y}}[0]. \quad (123)$$

1.1.28 S/N Ratio of Digital Correlator Output

We are now ready to calculate a ratio S/N of the signal $|\langle \mathcal{R} \rangle|$ to the noise $\sigma_{\mathcal{R}}$ of the digital correlator output, using equations (119) and (123). Here we denote amplitude of $\langle \mathcal{R} \rangle$ as $|\langle \mathcal{R} \rangle|$ regarding as if the digital correlator output is a complex quantity derived from a complex correlator. We will use this convention for other relevant quantities as well. We obtain

$$S/N = \frac{|\langle \mathcal{R} \rangle|}{\sigma_{\mathcal{R}}} = \sqrt{\mathcal{N}} \frac{|R_{\hat{x}\hat{y}}[0]|}{\sqrt{R_{\hat{x}\hat{x}}[0] R_{\hat{y}\hat{y}}[0]}} = \sqrt{\mathcal{N}} |r_{\hat{x}\hat{y}}[0]| = \sqrt{\mathcal{N}} \left| \frac{r_{\hat{x}\hat{y}}[0]}{r_{xy}[0]} \right| |\rho|, \quad (124)$$

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 stopping. As we saw in equation (62), the amplitude of 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 throughout 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

$$\mathcal{N} = 2B\tau_a, \quad (125)$$

as we saw in equation (64). Therefore, equation (124) is reduced to

$$S/N = \left| \frac{r_{\hat{x}\hat{y}}[0]}{r_{xy}[0]} \right| \sqrt{\frac{T_{A_1} T_{A_2}}{T_{S_1} T_{S_2}}} \sqrt{2B\tau_a} = \left| \frac{r_{\hat{x}\hat{y}}[0]}{r_{xy}[0]} \right| (S/N)_{analog}, \quad (126)$$

where

$$(S/N)_{analog} \equiv \sqrt{\frac{T_{A_1} T_{A_2}}{T_{S_1} T_{S_2}}} \sqrt{2B\tau_a},$$

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 due to quantization’’ η_{cQ} , which is defined as a ratio of S/N after clipping to the analog one:

$$\eta_{cQ} \equiv \frac{S/N}{(S/N)_{analog}}, \quad (127)$$

is given by

$$\eta_{c_Q} = \left| \frac{r_{\hat{x}\hat{y}}[0]}{r_{xy}[0]} \right|, \quad (128)$$

in this case.

With the coherence factor due to quantization η_{c_Q} , equation (126) is reduced to

$$S/N = \eta_{c_Q} \sqrt{\frac{T_{A_1} T_{A_2}}{T_{S_1} T_{S_2}}} \sqrt{2 B \tau_a}, \quad (129)$$

which is an important formula for estimating sensitivity of a digital radio interferometer.

For the case of the small cross-correlation coefficient $|r_{xy}[m]| \ll 1$, which we generally assume as the usual case in radio interferometry, $r_{\hat{x}\hat{y}}[m]$ of the clipped data is proportional to $r_{xy}[m]$ of the original analog data. Therefore, the coherence factor due to quantization η_{c_Q} as given by equation (128) should be equal to the proportionality coefficient:

$$r_{\hat{x}\hat{y}}[m] = \eta_{c_Q} r_{xy}[m], \quad (130)$$

and, in view of equations (92) and (114), we have

$$\eta_{c_Q} = \frac{2}{\pi} \cong 0.64, \quad (131)$$

for the 1-bit quantization case, and

$$\eta_{c_Q} = \frac{2}{\pi} \frac{[(\mathbf{n} - 1) E + 1]^2}{[\Phi + \mathbf{n}^2 (1 - \Phi)]}, \quad (132)$$

where E and Φ are given in equations (107) and (112), for the 2-bit quantization case, when $|r_{xy}[0]| \ll 1$.

Note that equation (128), which takes into account the loss due to the effects of quantization only, is not a complete description of the coherence factor η_c which we introduced in Chapter 3 as

$$S/N = \eta_c \sqrt{\frac{T_{A_1} T_{A_2}}{T_{S_1} T_{S_2}}} \sqrt{2 B \tau_a}. \quad (133)$$

In addition to the quantization loss, we will take into account also losses associated with digital operations in the correlation processing, fringe-phase fluctuations due to instability of frequency standards and/or turbulent atmosphere, and imperfection of theoretically predicted delay model. In the result, the coherence factor is further reduced due to the losses.

1.1.29 Optimum Parameters v_0/σ and \mathbf{n} in 2–Bit Quantization

We mentioned earlier that the two parameters in the 2–bit quantization, the threshold value v_0 and the higher–level value \mathbf{n} , are chosen so that the signal–to–noise ratio of the digital correlator output is maximized. This condition is fulfilled when the coherence factor due to quantization η_{c_Q} takes the maximum. In the limiting case of the small cross–correlation coefficient $|r_{xy}[0]| \ll 1$, which is the usual case in radio interferometry, we can calculate η_{c_Q} using equation (132) with explicit forms for the functions E and Φ given in equations (107) and (112).

Figure 26 shows the coherence factor due to quantization as a function of

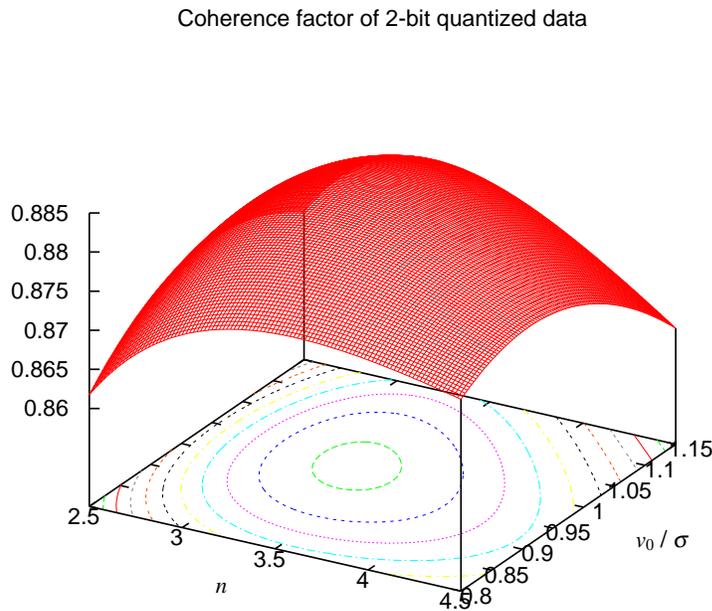


Figure 26: Coherence factor due to quantization η_{c_Q} of the 2–bit quantized data as a function of normalized threshold v_0/σ and higher–level value \mathbf{n} in the limiting case of the small cross–correlation coefficient.

the two parameters, v_0 normalized by the analog dispersion σ , and \mathbf{n} , in the small cross–correlation case.

As we see from the Figure, the maximum value of the coherence factor due to quantization $\eta_{c_Q} = 0.883$ is obtained when $\mathbf{n} = 3.34$ 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 \mathbf{n} is an integer.

Therefore, $\mathbf{n} = 3$, $v_0 = 0.996 \sigma$ with $\eta_{c_Q} = 0.881$, and $\mathbf{n} = 4$, $v_0 = 0.942 \sigma$ with $\eta_{c_Q} = 0.880$, are often used in existing VLBI 2-bit quantization systems. This is why we adopted $\mathbf{n} = 3$, $v_0 = 0.996 \sigma$ in Figures 23, 24, and 25.

Accordingly, in the limiting case of the small cross-correlation coefficient $|r_{xy}[0]| \ll 1$, the coherence factor due to quantization is given by

$$\eta_{c_Q} \cong 0.64, \quad (134)$$

for the 1-bit quantization case, and usually by

$$\eta_{c_Q} \cong 0.88, \quad (135)$$

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 rectangular spectra of the original analog data (Figures 21 and 24). Therefore, the Nyquist sampling for the original analog data with $2B$ rate is no longer optimum for the clipped data. This means that we can improve the signal-to-noise ratio of the clipped data by increasing sampling rate beyond $2B$.

Such a sampling with a rate higher than $2B$ is a little incorrectly called the ‘‘oversampling’’. This is not altogether correct since the higher rate results in an oversampling for the analog data, but does not for the clipped data. Anyway, we will see how an ‘‘oversampling’’ improves the signal-to-noise ratio of the clipped data.

In this case, the sampling interval is not equal to the optimum Nyquist one for the analog data. Therefore, different sample points are no longer independent to each other. In other words, autocorrelations between different sample points, i.e. $R_{xx}[m]$ and $R_{yy}[m]$ with $m \neq 0$ are not equal to zero any more. Consequently, we have to take into account their contributions 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 $2B$ of the original band-limited analog data. Then, in the calculation of the dispersion of the digital correlator output \mathcal{R} shown in equations (120) – (123), we must leave autocorrelation terms with non-zero arguments. Specifically, the second term in the RHS of equation (121) now must be

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

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

$$\begin{aligned}\sigma_{\mathcal{R}}^2 &= \frac{1}{\mathcal{N}^2} \left[\sum_{i=1}^{\mathcal{N}} \langle \hat{x}[i] \hat{y}[i] \hat{x}[i] \hat{y}[i] \rangle + \sum_{i=1}^{\mathcal{N}} \sum_{j \neq i}^{\mathcal{N}} \langle \hat{x}[i] \hat{y}[i] \hat{x}[j] \hat{y}[j] \rangle \right] - R_{\hat{x}\hat{y}}^2[0] \\ &\cong \frac{1}{\mathcal{N}} R_{\hat{x}\hat{x}}[0] R_{\hat{y}\hat{y}}[0] + \frac{1}{\mathcal{N}^2} \sum_{i=1}^{\mathcal{N}} \sum_{j \neq i}^{\mathcal{N}} R_{\hat{x}\hat{x}}[i-j] R_{\hat{y}\hat{y}}[i-j],\end{aligned}\quad (136)$$

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 difference $k = i - j$ is $\mathcal{N} - |k|$, and autocorrelations are likely to be large enough for small $|k| \ll \mathcal{N}$ only, we have

$$\begin{aligned}\sigma_{\mathcal{R}}^2 &\cong \frac{1}{\mathcal{N}} R_{\hat{x}\hat{x}}[0] R_{\hat{y}\hat{y}}[0] + \frac{2}{\mathcal{N}^2} \sum_{k=1}^{\mathcal{N}-1} (\mathcal{N} - k) R_{\hat{x}\hat{x}}[k] R_{\hat{y}\hat{y}}[k] \\ &\cong \frac{1}{\mathcal{N}} R_{\hat{x}\hat{x}}[0] R_{\hat{y}\hat{y}}[0] \left(1 + 2 \sum_{k=1}^{\infty} r_{\hat{x}\hat{x}}[k] r_{\hat{y}\hat{y}}[k] \right),\end{aligned}\quad (137)$$

where we introduced correlation coefficients of the clipped data $r_{\hat{x}\hat{x}}[k] = R_{\hat{x}\hat{x}}[k]/R_{\hat{x}\hat{x}}[0]$ and $r_{\hat{y}\hat{y}}[k] = R_{\hat{y}\hat{y}}[k]/R_{\hat{y}\hat{y}}[0]$, and we approximated the upper limit $\mathcal{N} - 1$ of summation by the infinity.

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

$$S/N = \frac{|\langle \mathcal{R} \rangle|}{\sigma_{\mathcal{R}}} = \left| \frac{r_{\hat{x}\hat{y}}[0]}{r_{xy}[0]} \right| \sqrt{\frac{T_{A_1} T_{A_2}}{T_{S_1} T_{S_2}}} \frac{\sqrt{2\beta B\tau_a}}{\sqrt{1 + 2 \sum_{k=1}^{\infty} r_{\hat{x}\hat{x}}[k] r_{\hat{y}\hat{y}}[k]}}. \quad (138)$$

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

$$\eta_{cQ} = \frac{S/N}{(S/N)_{analog}} = \left| \frac{r_{\hat{x}\hat{y}}[0]}{r_{xy}[0]} \right| \frac{\sqrt{\beta}}{\sqrt{1 + 2 \sum_{k=1}^{\infty} r_{\hat{x}\hat{x}}[k] r_{\hat{y}\hat{y}}[k]}}. \quad (139)$$

We can estimate this factor, using equations (96) and (115) 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 function $f_{cl}(r)$ is $f_{cl}(r) = r$ for the analog (i.e. non-clipped) data, $f_{cl}(r) = (\pi/2) \arcsin(r)$ for the 1-bit quantized data (equation (96)), and the function given in equation (115) for the 2-bit quantized data, correspondingly.

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 $T_s = 1/(2B\beta)$ are given by

$$r_{xx}[k] = r_{yy}[k] = r_{xx}(kT_s) = r_{yy}(kT_s) = \frac{\sin\left(\frac{\pi k}{\beta}\right)}{\frac{\pi k}{\beta}}. \quad (140)$$

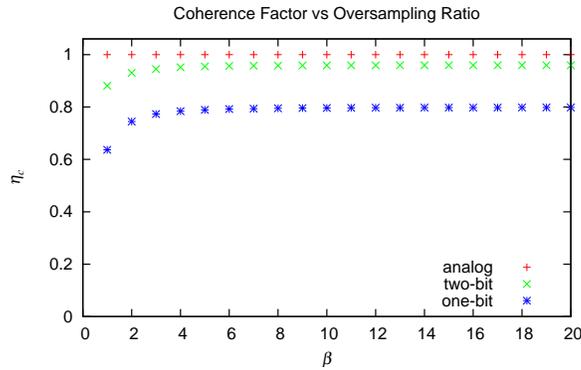


Figure 27: Coherence factor due to quantization η_{cQ} for the analog (+), 2-bit quantized (\times), and 1-bit quantized ($*$) data as functions of the oversampling ratio β .

Thus, the signal-to-noise ratio and the coherence factor due to quantization of the oversampled clipped data are given by

$$S/N = \eta_{cQ} \sqrt{\frac{T_{A_1} T_{A_2}}{T_{S_1} T_{S_2}}} \sqrt{2B\tau_a}, \quad (141)$$

$$\eta_{c_Q} = \frac{\left| \frac{r_{\hat{x}\hat{y}}[0]}{r_{xy}[0]} \right| \sqrt{\beta}}{\sqrt{1 + 2 \sum_{k=1}^{\infty} f_{cl}^2 \left(\frac{\sin\left(\frac{\pi k}{\beta}\right)}{\frac{\pi k}{\beta}} \right)}}. \quad (142)$$

Figure 27 shows the coherence factor due to quantization η_{c_Q} for the analog, 2-bit quantized, and 1-bit quantized data as functions of the oversampling ratio β , calculated by means of equation (142). 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 $\eta_{c_Q} = 0.744$ for the 1-bit quantized data, and $\eta_{c_Q} = 0.930$ for the 2-bit quantized data.

1.1.31 Coherence Factor and Sensitivity with Given Bit Rate

Coherence factor due to quantization η_{c_Q} with different clipping and oversampling is summarized in Table 2. It is obvious that the coherence factor

Number of bits N_b	Number of quantization levels	coherence factor η_{c_Q}	
		$\beta = 1$	$\beta = 2$
1	2	0.64	0.74
2	4	0.88	0.93

Table 2: Relationship of coherence factor due to quantization η_{c_Q} with number of bits N_b and oversampling ratio β .

increases as we increase 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 samplers, recorders, or correlators. If the maximum bit-rate ν_b is fixed, the maximum permissible sampling rate ν_S , which is β times as large as the Nyquist rate $2B$ with the analog bandwidth B , is $\nu_S = \nu_b/N_b = 2\beta B$. Therefore, the maximum allowable bandwidth is limited by the bit-rate as $B = \nu_b/(2\beta N_b)$. Hence, in view of equation (141), the maximum signal-to-noise ratio, i.e. the sensitivity of the interferometric observation, for a continuum spectrum source, is proportional to

$$S/N = \eta_{c_Q} \frac{T_{A_1} T_{A_2}}{T_{S_1} T_{S_2}} \sqrt{2B\tau_a} \propto \eta_{c_Q} \sqrt{B} \propto \frac{\eta_{c_Q}}{\sqrt{\beta N_b}}. \quad (143)$$

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

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

Table 3 shows the factor $\eta_{c_Q}/\sqrt{\beta N_b}$, 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.

On the other hand, when we observe a spectralline source, such as an astronomical maser source, signal from the source is confined in a limited range of frequency where we see narrow spectral lines. In such a case, it is more meaningful to know signal-to-noise ratio of a spectral line in a cross-power spectrum, that is a ratio of signal and noise contained within a narrow line spectrum of linewidth b , rather than the ordinary signal-to-noise ratio for the whole filter bandwidth B of our receiving system. We can estimate such a signal-to-noise ratio of a spectral line by replacing the filter bandwidth B in equation (141) by the linewidth b . Here we define T_{A_1} and T_{A_2} as peak antenna temperatures averaged over the linewidth b . The linewidth and peak antenna temperatures are intrinsic to the radio source and do not depend on digitization parameters, such as the number of bits N_b , or the oversampling ratio β . Therefore, the larger the coherence factor due to quantization η_{c_Q} , the higher is the sensitivity in this case. Consequently, modern VLBI systems tend to adopt the 2-bit quantization scheme for better performance in spectralline-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 we see in Table 3.

Figure 28 shows an example of the maser source spectra derived from 2-bit quantized data.

This concludes our brief discussion on basic elements of digital data processing as a fundament of modern VLBI systems. Now we proceed to examine the three specific technologies which enabled realization of VLBI.

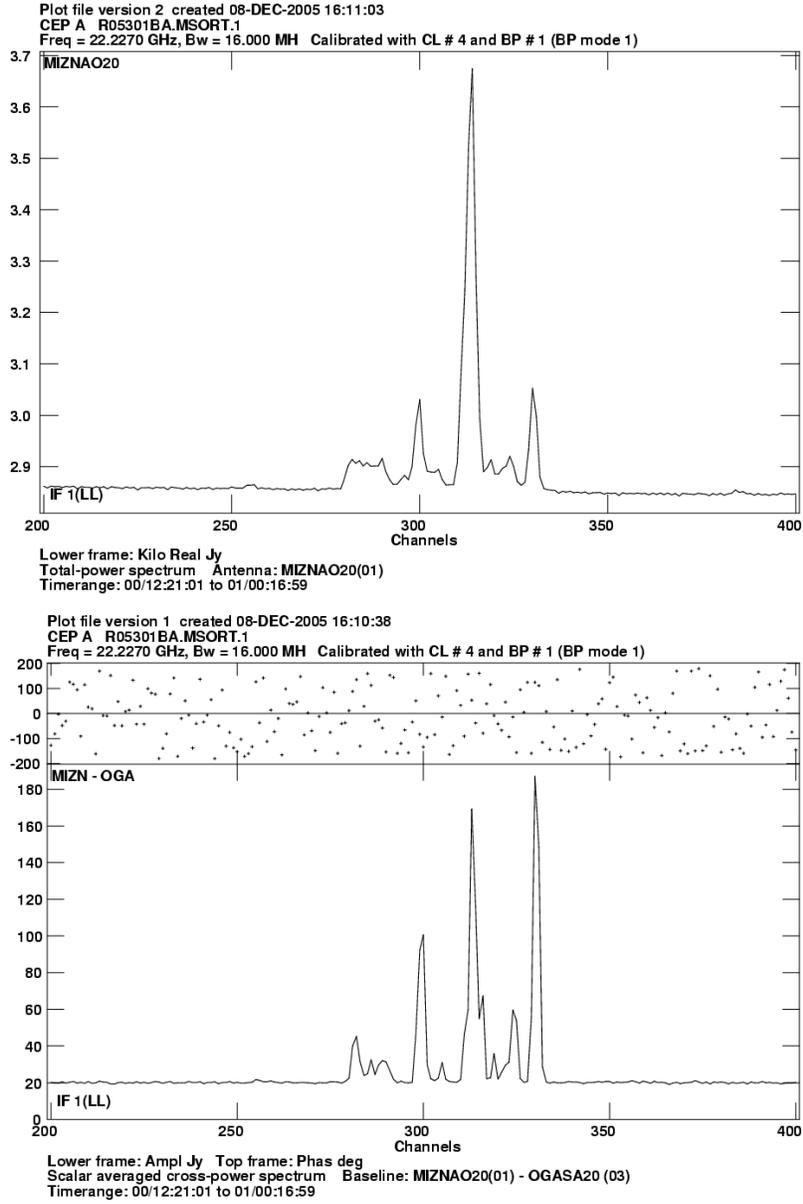


Figure 28: Spectra of water maser source Cep A in an active region of star formation derived from 2-bit quantized data. Upper panel shows total-power spectrum with Mizusawa 20m antenna, and lower panel shows phase and amplitude of cross-power spectrum with Mizusawa–Ogasawara baseline of the VERA. Profiles 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.)

1.2 Frequency Standard

1.2.1 VLBI Requires “Absolute” Frequency Stability

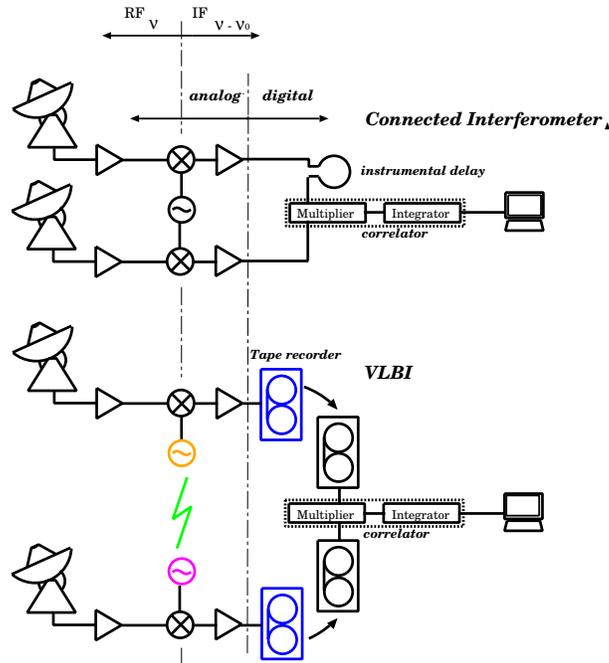


Figure 29: Connected-Element Interferometer vs. VLBI system.

Development of highly stable frequency standards was indispensable for realization of VLBI.

In the case of a connected-element interferometer array, a reference signal transmitted via cables from a common frequency standard is used to generate local oscillator (LO) signals for frequency conversion in all antennas of the array (see upper panel of Figure 29). Therefore, any phase noise, due to instability in the frequency standard, is commonly added to signals from all antennas. Such common phase noise from any 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

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

then, another signal from the same source received by antenna B must be

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

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 (time-averaging), we have

$$\overline{v_A(t) v_B(t)} \propto \frac{1}{2} \cos(\omega\tau_g), \quad (144)$$

since rapidly oscillating second term with frequency 2ω is averaged out. Thus, no effect of the phase noise due to the frequency standard remains in the correlator output!

Consequently, **correlation results in connected-element interferometers are unaffected by the instability of the common frequency standard.** Hence, we readily obtain the almost pure fringe pattern $\propto \cos(\omega\tau_g)$ in the correlation results as shown in equation (144). This is why very high stability of frequency standards is not necessarily required in connected-element interferometers.

How about VLBI, then?

Element antennas in a VLBI array are so far from each other that it is impossible to supply a reference signal to them through cables from a common frequency standard.

Therefore, each antenna in VLBI has to use its own independent frequency standard to generate the LO signal (see lower panel of Figure 29).

Thus, if a signal from a point source received by antenna A is

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

then, another signal from the same source 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 (time-averaging),

$$\overline{v_A(t) v_B(t)} \propto \frac{1}{2} \cos[\omega\tau_g - \phi_A(t) + \phi_B(t)]. \quad (145)$$

Thus, every correlation result in VLBI always contains a difference of phase noises $\phi_A(t) - \phi_B(t)$ due to the frequency standards.

Therefore, we cannot expect any automatic compensation of the phase noises. 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 (145)) gives rise to two big difficulties in VLBI.

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

Second, the phase noise, which is usually time variable, limits the sensitivity of VLBI. Indeed, it becomes impossible for us to completely stop the oscillation of the correlator output if the phase in the sinusoidal fringe pattern 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 system noise contribution in the correlator output must be surely suppressed by the integration, but amplitude of the averaged signal itself could be reduced even more rapidly due to the phase-noise-induced oscillation. Depending on the ratio between the integration time and 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 “**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 badly need “absolute” stability of frequency standards in VLBI, in order to ease these difficulties.

Of course, not only the noise due to the frequency standards, but any other phase noise, due in particular to irregular propagation delay induced by the **turbulent atmosphere**, causes the same sort of difficulties, as we will see later.

Now we will 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:

$$v_{ideal}(t) = v_0 \cos(\omega_0 t), \quad (146)$$

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

$$v(t) = v_0 \cos(\omega_0 t + \phi(t)). \quad (147)$$

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

$$\omega_a(t) = \omega_0 + \frac{d\phi(t)}{dt}. \quad (148)$$

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

$$y(t) = \frac{\delta\omega(t)}{\omega_0} = \frac{\omega_a(t) - \omega_0}{\omega_0} = \frac{1}{\omega_0} \frac{d\phi(t)}{dt}, \quad (149)$$

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 :

$$R_{\phi\phi}(\tau) = \langle \phi(t + \tau) \phi(t) \rangle, \quad \text{and} \quad R_{yy}(\tau) = \langle y(t + \tau) y(t) \rangle, \quad (150)$$

and their power spectra are given by Fourier transform relations:

$$\begin{aligned} S_{\phi\phi}(\omega) &= \int_{-\infty}^{\infty} R_{\phi\phi}(\tau) e^{-i\omega\tau} d\tau, & R_{\phi\phi}(\tau) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} S_{\phi\phi}(\omega) e^{i\omega\tau} d\omega, \\ S_{yy}(\omega) &= \int_{-\infty}^{\infty} R_{yy}(\tau) e^{-i\omega\tau} d\tau, & R_{yy}(\tau) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} S_{yy}(\omega) e^{i\omega\tau} d\omega. \end{aligned} \quad (151)$$

In view of equation (149), the autocorrelations of $y(t)$ and $\phi(t)$ are mutually related by

$$R_{yy}(t, t') = \langle y(t) y(t') \rangle = \frac{1}{\omega_0^2} \left\langle \frac{d\phi(t)}{dt} \frac{d\phi(t')}{dt'} \right\rangle = \frac{1}{\omega_0^2} \frac{\partial^2}{\partial t \partial t'} R_{\phi\phi}(t, t'). \quad (152)$$

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

$$R_{yy}(\tau) = -\frac{1}{\omega_0^2} \frac{d^2}{d\tau^2} R_{\phi\phi}(\tau), \quad (153)$$

where $\tau = t - t'$.

From equations (151) and (153), 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:

$$S_{yy}(\omega) = \frac{\omega^2}{\omega_0^2} S_{\phi\phi}(\omega). \quad (154)$$

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” $S_{\phi\phi}(\nu)$ and $S_{yy}(\nu)$:

$$\begin{aligned} S_{\phi\phi}(\nu) &= 4 \int_0^{\infty} R_{\phi\phi}(\tau) \cos(2\pi\nu\tau) d\tau, & R_{\phi\phi}(\tau) &= \int_0^{\infty} S_{\phi\phi}(\nu) \cos(2\pi\nu\tau) d\nu, \\ S_{yy}(\nu) &= 4 \int_0^{\infty} R_{yy}(\tau) \cos(2\pi\nu\tau) d\tau, & R_{yy}(\tau) &= \int_0^{\infty} S_{yy}(\nu) \cos(2\pi\nu\tau) d\nu. \end{aligned} \quad (155)$$

These single-sided spectra $S_{\phi\phi}(\nu)$ and $S_{yy}(\nu)$ have been widely used in the frequency stability discussions. They are related to our double-sided power spectra $S_{\phi\phi}(\omega)$ and $S_{yy}(\omega)$ by equations:

$$S_{\phi\phi}(\nu) = 2 S_{\phi\phi}(2\pi\nu), \quad \text{and} \quad S_{yy}(\nu) = 2 S_{yy}(2\pi\nu), \quad (156)$$

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

$$S_{yy}(\nu) = \frac{\nu^2}{\nu_0^2} S_{\phi\phi}(\nu), \quad (157)$$

instead of equation (154), where $\nu_0 = \frac{\omega_0}{2\pi}$.

However, we will continue to use our double-sided power spectrum forms in equations (151), following our previous discussions.

1.2.3 Types of Phase and Frequency Noises

A power spectrum $S_{yy}(\omega)$ of FFD due to instability of a frequency standard is often well approximated by a power-law spectrum $H_\alpha \omega^\alpha$ with a constant index α and a constant amplitude H_α within a certain range of frequency ω . The power-law index α is determined by a physical process which mainly generates the FFD (or frequency noise) in the frequency range.

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: Types of frequency and phase noise spectra with power-law indices α of the frequency noise (or FFD: fractional frequency deviation) spectra.

Therefore, noises in the frequency stability are conventionally classified into several types according to the index α of a model power-law spectrum which closely approximates the real FFD spectrum within a frequency range. 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 schematically shown in Figure 30.

1.2.4 Time Domain Measurements

Traditionally, it has been easier to make measurements in the time domain, than in the frequency domain. Hence, frequency stability characteristics are mostly given in terms of the time domain measurements.

Suppose that we have N phase values: $\phi(t_1), \phi(t_2), \dots, \phi(t_k), \dots, \phi(t_N)$, measured at equally spaced time points: $t_1, t_2, \dots, t_k, \dots, t_N$, 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**” $\bar{y}[k]$

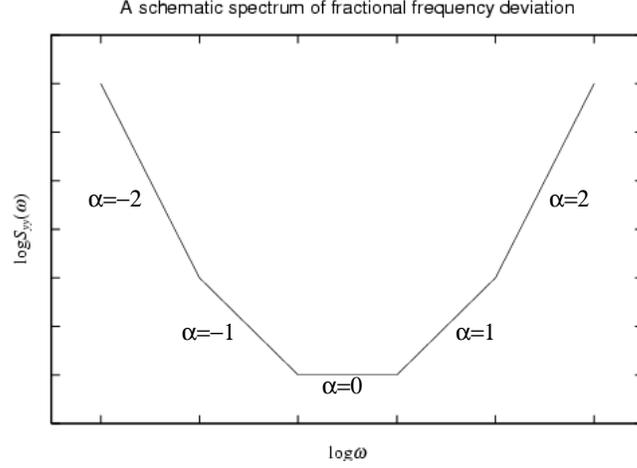


Figure 30: A schematic spectrum of the FFD, $y(t)$, of an oscillator showing all power-law components in Table 4 in various frequency ranges. Horizontal and vertical axes show the angular frequency ω and the power spectrum $S_{yy}(\omega)$, respectively, in log-scales.

which is defined by

$$\bar{y}[k] = \frac{\phi(t_{k+1}) - \phi(t_k)}{\omega_0 \mathcal{T}} = \frac{1}{\omega_0 \mathcal{T}} \int_{t_k}^{t_{k+1}} \frac{d\phi(t')}{dt'} dt' = \frac{1}{\mathcal{T}} \int_{t_k}^{t_{k+1}} y(t') dt', \quad (158)$$

and is described also as

$$\bar{y}[k] = \frac{1}{\mathcal{T}} \int_{t - \frac{\mathcal{T}}{2}}^{t + \frac{\mathcal{T}}{2}} y(t') dt', \quad (159)$$

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

$$\bar{y}[k] = y(t) * a(t) = \int_{-\infty}^{\infty} y(t - t') a(t') dt', \quad (160)$$

with an impulse response $a(t)$:

$$a(t) = \begin{cases} \frac{1}{\mathcal{T}} & \text{if } -\frac{\mathcal{T}}{2} < t \leq \frac{\mathcal{T}}{2}, \\ 0 & \text{otherwise,} \end{cases} \quad (161)$$

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

$$I^2(\mathcal{T}) = \langle \bar{y}^2[k] \rangle, \quad (162)$$

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 $\bar{y}^2[k]$:

$$\Sigma_t = \frac{1}{N - 1} \sum_{k=1}^{N-1} \bar{y}^2[k]. \quad (163)$$

At the first glance, 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 (163), 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:

$$\Delta \bar{y}[k] = \bar{y}[k + 1] - \bar{y}[k], \quad (164)$$

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

$$\Sigma_a = \frac{1}{2(N - 2)} \sum_{k=1}^{N-2} (\Delta \bar{y}[k])^2, \quad (165)$$

(Allan, 1966). In view of equation (158), the two sample difference $\Delta \bar{y}[k]$ of the mean FFD is easily derived from the measured phase noise $\phi[k]$ by

$$\Delta \bar{y}[k] = \frac{\phi(t_{k+2}) - 2\phi(t_{k+1}) + \phi(t_k)}{\omega_0 \mathcal{T}}. \quad (166)$$

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

$$\sigma_y^2(\mathcal{T}) = \frac{\langle (\Delta \bar{y}[k])^2 \rangle}{2}, \quad (167)$$

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

Estimation of the AVAR by means of Σ_a has a better chance to converge than that of the TVAR by means of Σ_t , as schematically illustrated in Figure 31. Therefore, the estimated AVAR is widely accepted as a good measure of the frequency stability.

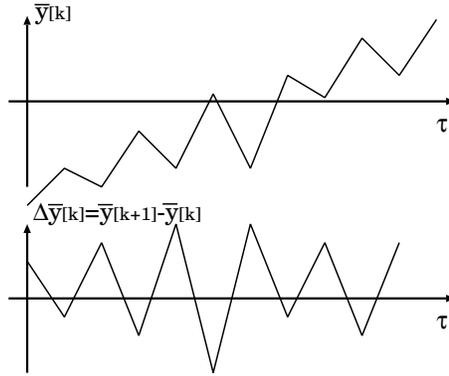


Figure 31: Mean square sum of the two-sample difference of the mean FFD (lower panel) has a better chance to converge than that of the mean FFD itself (upper panel), even when the mean FFD shows a diverging long-time-scale behavior.

In view of equation (158), we can describe the two-sample difference of the mean FFD $\Delta \bar{y}[k]$ through time integrations of the FFD $y(t)$:

$$\begin{aligned} \Delta \bar{y}[k] = \bar{y}[k+1] - \bar{y}[k] &= \frac{1}{\mathcal{T}} \left\{ \int_{t_{k+1}}^{t_{k+2}} y(t') dt' - \int_{t_k}^{t_{k+1}} y(t') dt' \right\} \\ &= \frac{1}{\mathcal{T}} \left\{ \int_t^{t+\mathcal{T}} y(t') dt' - \int_{t-\mathcal{T}}^t y(t') dt' \right\}, \end{aligned} \quad (168)$$

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

$$\Delta \bar{y}[k] = y(t) * b(t) = \int_{-\infty}^{\infty} y(t-t') b(t') dt', \quad (169)$$

with an impulse response:

$$b(t) = \begin{cases} \frac{1}{\mathcal{T}} & \text{if } -\mathcal{T} < t \leq 0, \\ -\frac{1}{\mathcal{T}} & \text{if } 0 < t \leq \mathcal{T}, \\ 0 & \text{otherwise,} \end{cases} \quad (170)$$

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

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

Equations (160) and (169) describe the discrete-time series $\bar{y}[k]$ (the mean 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 $\Delta\bar{y}(t)$:

$$\begin{aligned} \bar{y}(t) &= y(t) * a(t) = \int_{-\infty}^{\infty} y(t-t') a(t') dt', \\ \Delta\bar{y}(t) &= y(t) * b(t) = \int_{-\infty}^{\infty} y(t-t') b(t') dt', \end{aligned} \quad (171)$$

at arbitrary time t , and introduce their autocorrelations

$$\begin{aligned} R_{\bar{y}\bar{y}}(\tau) &= \langle \bar{y}(t+\tau) \bar{y}(t) \rangle, \\ R_{\Delta\bar{y}\Delta\bar{y}}(\tau) &= \langle \Delta\bar{y}(t+\tau) \Delta\bar{y}(t) \rangle, \end{aligned} \quad (172)$$

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.

$$\begin{aligned} \bar{y}(t) &= \bar{y}[k] \quad \text{at } t = t_k + \frac{\mathcal{T}}{2}, \\ \Delta\bar{y}(t) &= \Delta\bar{y}[k] \quad \text{at } t = t_{k+1}. \end{aligned} \quad (173)$$

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

$$I^2(\mathcal{T}) = \langle (\bar{y}[k])^2 \rangle = R_{\bar{y}\bar{y}}(0) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_{\bar{y}\bar{y}}(\omega) d\omega. \quad (174)$$

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

$$\sigma_y^2(\mathcal{T}) = \frac{\langle (\Delta \bar{y}[k])^2 \rangle}{2} = \frac{R_{\Delta \bar{y} \Delta \bar{y}}(0)}{2} = \frac{1}{4\pi} \int_{-\infty}^{\infty} S_{\Delta \bar{y} \Delta \bar{y}}(\omega) d\omega. \quad (175)$$

Since the functions $\bar{y}(t)$ and $\Delta \bar{y}(t)$ are related to the FFD $y(t)$ through the linear systems in equation (171), we can describe the power spectra $S_{\bar{y}\bar{y}}(\omega)$ 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 (151). For this purpose, let us introduce system functions $A(\omega)$ and $B(\omega)$ of the impulse responses $a(t)$ and $b(t)$, respectively:

$$\begin{aligned} A(\omega) &= \int_{-\infty}^{\infty} a(t) e^{-i\omega t} dt = \frac{1}{\mathcal{T}} \int_{-\frac{\mathcal{T}}{2}}^{\frac{\mathcal{T}}{2}} e^{-i\omega t} dt = \frac{\sin(\frac{\omega \mathcal{T}}{2})}{(\frac{\omega \mathcal{T}}{2})}, \\ B(\omega) &= \int_{-\infty}^{\infty} b(t) e^{-i\omega t} dt = \frac{1}{\mathcal{T}} \left\{ \int_{-\mathcal{T}}^0 e^{-i\omega t} dt - \int_0^{\mathcal{T}} e^{-i\omega t} dt \right\} = 2i \frac{\sin^2(\frac{\omega \mathcal{T}}{2})}{(\frac{\omega \mathcal{T}}{2})}. \end{aligned} \quad (176)$$

From general properties of linear systems, we have

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

and

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

Therefore, equations (174) and (175) are reduced to

$$I^2(\mathcal{T}) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_{yy}(\omega) \frac{\sin^2(\frac{\omega \mathcal{T}}{2})}{(\frac{\omega \mathcal{T}}{2})^2} d\omega, \quad (177)$$

$$\sigma_y^2(\mathcal{T}) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_{yy}(\omega) 2 \frac{\sin^4(\frac{\omega \mathcal{T}}{2})}{(\frac{\omega \mathcal{T}}{2})^2} d\omega. \quad (178)$$

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

Figure 32 shows squared system functions $|A(\omega)|^2$ and $|B(\omega)|^2$ for the TVAR and the AVAR, respectively. It is evident from this Figure that the low frequency noise is effectively suppressed in the AVAR. Therefore, AVAR is really more suited than TVAR for describing noises diverging at low frequency.

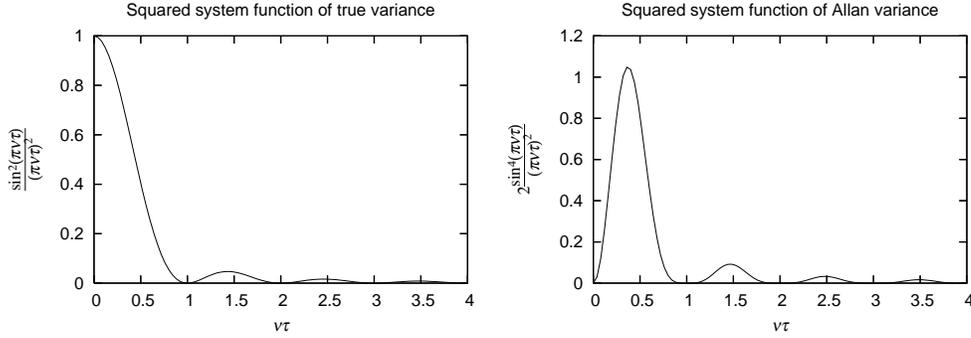


Figure 32: Squared system functions for the TVAR, $|A(\omega)|^2$, (left) and the AVAR, $|B(\omega)|^2$, (right). Horizontal axes show the product νT of the frequency and the time interval.

1.2.7 Time–Interval Dependence of Allan Variance

We can calculate dependence of the TVAR and the AVAR on the time interval T according to equations (177) and (178) 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	$S_{yy}(\omega)$	$S_{\phi\phi}(\omega)$	$\sigma_y^2(T)$	$I^2(T)$
White phase	$H_2 \omega^2$	$\omega_0^2 H_2$	$\frac{3H_2\omega_h}{\pi T^2}$	$\frac{2H_2\omega_h}{\pi T^2}$
Flicker phase	$H_1 \omega^1$	$\omega_0^2 H_1 \omega^{-1}$	$\frac{3H_1 \ln(\omega_h T)}{\pi T^2}$	—
White frequency	H_0	$\omega_0^2 H_0 \omega^{-2}$	$\frac{H_0}{T}$	$\frac{H_0}{T}$
Flicker frequency	$H_{-1} \omega^{-1}$	$\omega_0^2 H_{-1} \omega^{-3}$	$\frac{(2 \ln 2) H_{-1}}{\pi}$	—
Random walk of frequency	$H_{-2} \omega^{-2}$	$\omega_0^2 H_{-2} \omega^{-4}$	$\frac{T H_{-2}}{3}$	—

Table 5: Time–interval dependence of the AVAR, $\sigma_y^2(T)$, and the TVAR, $I^2(T)$, 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 33 schematically shows the time–interval dependence of the “**Allan standard deviation (ASD)**”, $\sigma_y(T)$, which is defined as the square root of the AVAR. It is clear from this Figure that, having measured ASD as a function of time interval between samples T , we can readily distinguish noise types, except for the case of the white phase and the flicker phase. This

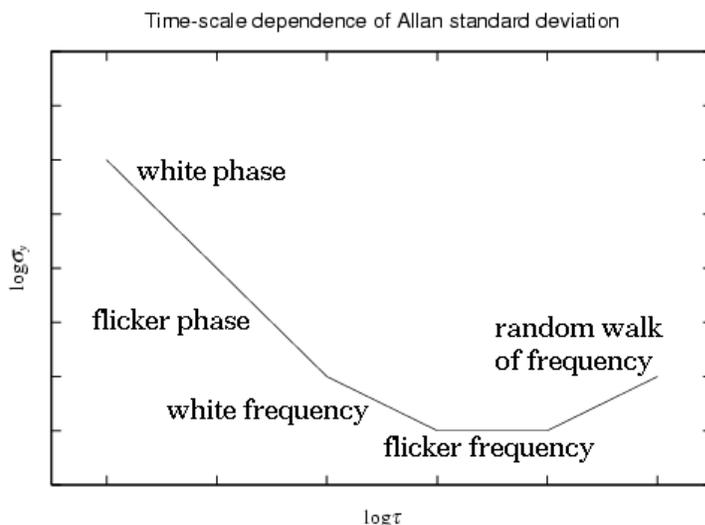


Figure 33: A schematic view of the time-interval \mathcal{T} dependence of the Allan standard deviation $\sigma_y(\mathcal{T})$ in log-scales. Each noise type shows its own gradient in this diagram, though lines of the white phase and the flicker phase are hardly distinguished by the difference of their gradients.

demonstrates real usefulness of the ASD (or the AVAR) as a measure of the frequency stability.

Figure 34 shows characteristic performance of existing frequency standards in terms of their ASD as functions of 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 $\text{ASD} \leq 10^{-15}$ in its flicker frequency regime, exhibits the highest frequency stability in the time-interval range of 1,000 to 10,000 seconds.

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 (158) and (162) yield

$$I^2(\mathcal{T}) = \langle \bar{y}^2[k] \rangle = \frac{\langle [\phi(t_k + \mathcal{T}) - \phi(t_k)]^2 \rangle}{(\omega_0 \mathcal{T})^2} = \frac{2[R_{\phi\phi}(0) - R_{\phi\phi}(\mathcal{T})]}{(\omega_0 \mathcal{T})^2}. \quad (179)$$

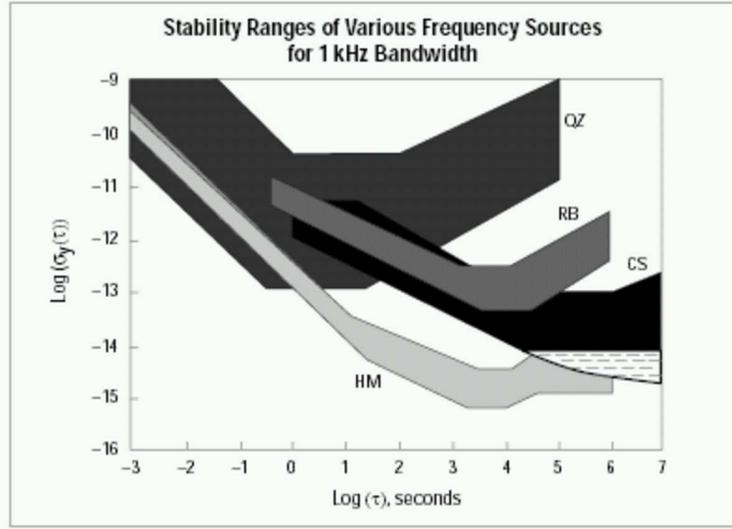


Figure 34: Performance of various frequency standards (QZ: Quartz, RB: Rubidium, CS: Cesium, HM: Active Hydrogen Maser) in terms of the ASD—time interval relation. (Figure courtesy of HEWLETT PACKARD, Application Note 1289).

Also, for the AVAR, equations (166) and (167) lead to

$$\begin{aligned}\sigma_y^2(\mathcal{T}) &= \frac{\langle (\Delta \bar{y}[k])^2 \rangle}{2} = \frac{\langle [\phi(t_k + 2\mathcal{T}) - 2\phi(t_k + \mathcal{T}) + \phi(t_k)]^2 \rangle}{2(\omega_0 \mathcal{T})^2} \\ &= \frac{3R_{\phi\phi}(0) - 4R_{\phi\phi}(\mathcal{T}) + R_{\phi\phi}(2\mathcal{T})}{(\omega_0 \mathcal{T})^2}.\end{aligned}\quad (180)$$

Now, from equation (179), we have,

$$\begin{aligned}I^2(\mathcal{T}) - I^2(2\mathcal{T}) &= \frac{2[R(0) - R(\mathcal{T})] - \frac{1}{2}[R(0) - R(2\mathcal{T})]}{(\omega_0 \mathcal{T})^2} \\ &= \frac{3R_{\phi\phi}(0) - 4R_{\phi\phi}(\mathcal{T}) + R_{\phi\phi}(2\mathcal{T})}{2(\omega_0 \mathcal{T})^2}.\end{aligned}\quad (181)$$

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

$$\sigma_y^2(\mathcal{T}) = 2[I^2(\mathcal{T}) - I^2(2\mathcal{T})].\quad (182)$$

1.2.9 Coherence Function

As we saw earlier, noisy fluctuations in the fringe phase due to unstable frequency standards and/or turbulent atmosphere causes a 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

$$C(T) = \left| \frac{1}{T} \int_0^T e^{i\phi(t)} dt \right|. \quad (183)$$

This is a monotonically decreasing function of integration time T within a range $0 \leq C(T) \leq 1$. Magnitude of the squared coherence function is represented by its dispersion:

$$\langle C^2(T) \rangle = \frac{1}{T^2} \int_0^T \int_0^T \langle e^{i[\phi(t) - \phi(t')]} \rangle dt dt'. \quad (184)$$

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

$$f(\Phi) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{\Phi^2}{2\sigma^2}}, \quad (185)$$

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 (184) to

$$\langle C^2(T) \rangle = \frac{1}{T^2} \int_0^T \int_0^T e^{-\frac{\langle [\phi(t) - \phi(t')]^2 \rangle}{2}} dt dt' = \frac{1}{T^2} \int_0^T \int_0^T e^{-\frac{D(t, t')}{2}} dt dt', \quad (186)$$

where $D(t, t')$ is the so-called “**temporal structure function**” defined by

$$D(t, t') = \langle [\phi(t) - \phi(t')]^2 \rangle. \quad (187)$$

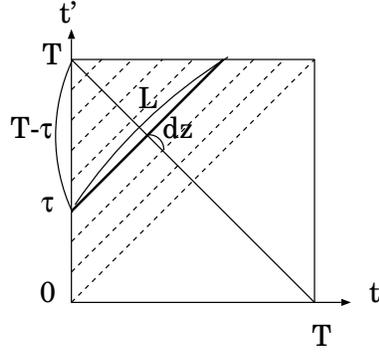


Figure 35: Geometry of integration in equation (186).

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 (186) to a single integral. In fact, noting that $dz = \frac{1}{\sqrt{2}} d\tau$ and $L = \sqrt{2}(T - \tau)$ in Figure 35, we obtain

$$\langle C^2(T) \rangle = \frac{2}{T^2} \int_0^T (T - \tau) e^{-\frac{D(\tau)}{2}} d\tau = \frac{2}{T} \int_0^T \left(1 - \frac{\tau}{T}\right) e^{-\frac{D(\tau)}{2}} d\tau. \quad (188)$$

According to equation (158), 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

$$D(\tau) = \omega_0^2 \tau^2 \langle \bar{y}^2 \rangle = \omega_0^2 \tau^2 I^2(\tau). \quad (189)$$

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

$$\begin{aligned} \langle C^2(T) \rangle &= \frac{2}{T} \int_0^T \left(1 - \frac{\tau}{T}\right) e^{-\frac{\omega_0^2 \tau^2 I^2(\tau)}{2}} d\tau \\ &= \frac{2}{T} \int_0^T \left(1 - \frac{\tau}{T}\right) e^{-\frac{\omega_0^2 \tau^2}{4\pi} \int_{-\infty}^{\infty} S_{yy}(\omega) \frac{\sin^2(\frac{\omega\tau}{2})}{(\frac{\omega\tau}{2})^2} d\omega} d\tau. \end{aligned} \quad (190)$$

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 $\langle C^2(\tau_c) \rangle \sim 0.85$.

The coherence loss due to fluctuations in the fringe phase reduces signal to noise ratio S/N of VLBI by a certain factor η_{cP} as

$$S/N \propto \eta_{cP} \sqrt{\frac{T_{A_1} T_{A_2}}{T_{S_1} T_{S_2}}} \sqrt{2 B \tau_a},$$

where $\eta_{cP} = \sqrt{\langle C^2(T) \rangle}$ is the “coherence factor due to fringe-phase fluctuations” which varies depending on type and magnitude of the phase noise and on integration time T . Other notations are the same with those used in Subsections 1.1.15 and 1.1.28. This factor η_{cP} should be multiplied to the coherence factor due to quantization η_{cQ} discussed in Subsection 1.1.28 when we simultaneously take into account losses due to quantization and fringe-phase fluctuations.

1.2.10 Approximate Estimation of Coherence Time

Precise estimation of the coherence time using equation (190) is sometimes impractical when the estimation Σ_t of the TVAR in equation (163) diverges.

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

Phase noise accumulated during a time T is approximately given by $\simeq \omega_0 \sigma_y(T) T$, where $\sigma_y(T)$ is “some standard deviation” of the FFD during the time interval T , 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 . In fact, this means averaging an oscillating signal for more than a cycle which could not give anything but nearly zero result. Thus the accumulated phase noise should be kept much smaller than 2π . It is the usual practice to require that the accumulated phase noise must not exceed one radian:

$$\omega_0 \sigma_y(T) T \leq 1 \text{ radian.} \quad (191)$$

(Note that σ_y is, in general, a function of the time interval T , i.e. $\sigma_y(T)$, as we saw in Figure 33.) Thus, the coherence time τ_c is estimated as a time interval which satisfies

$$\omega_0 \sigma_y(\tau_c) \tau_c \simeq 1. \quad (192)$$

For example, if the FFD behaves as a white phase noise, then $\sigma_y(T) \propto 1/T$, and hence $\sigma_y(T) T$ is independent of T . Therefore, in the white phase noise case, we can integrate the correlated results for arbitrarily long time interval, provided that $\sigma_y(T_*) < 1/(\omega_0 T_*)$ for any time interval T_* . Note that we get the same result by means of equation (190), too. In fact, $I^2(T) = \frac{2H_2\omega_h}{\pi T^2}$ in the white phase noise case as given in Table 5. Therefore, the equation yields $\langle C^2(T) \rangle = e^{-\frac{H_2\omega_0^2\omega_h}{\pi}}$, which is independent of T . Since $\sigma_y^2(T) = \frac{3}{2} I^2(T)$ in this case, the dispersion of the coherence function $\langle C^2(T) \rangle$ is kept close to 1 as long as $\sigma_y(T_*)$ is well smaller than $1/(\omega_0 T_*)$.

On the other hand, in the flicker–frequency region of frequency standards (often called the “flicker floor”), $\sigma_y(T)$ is almost independent of the time interval T as shown in Figure 33. In such a case, the longer the time interval, the larger is the accumulated phase noise, and the coherence time is inversely proportional to the frequency ω_0 .

If we observe at frequency $\nu = 8$ GHz, and require that the coherence time is as long as $\tau_c = 1000$ 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 34), fulfilled the requirement, and is widely used in the world VLBI observations.

Nowadays, the coherence time is most severely limited by the phase fluctuation due to the turbulent atmosphere which we will discuss later.

1.2.11 Estimation of Time–Averaged Phase Noise

In some applications of VLBI observation, 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:

$$\bar{\phi} = \frac{1}{T} \int_0^T \phi(t) dt, \quad (193)$$

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 (183) 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

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

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

$$\Delta\phi = \phi(t) - \bar{\phi} = \phi(t) - \frac{1}{T} \int_0^T \phi(t') dt', \quad (194)$$

and that of the time-averaged phase itself, given by equation (193).

The dispersions are

$$\begin{aligned} \sigma_{\Delta\phi}^2(T) &= \left\langle \left[\frac{1}{T} \int_0^T \left[\phi(t) - \frac{1}{T} \int_0^T \phi(t') dt' \right] dt \right]^2 \right\rangle \\ &= \frac{2}{T^2} \int_0^T (T - \tau) [R_{\phi\phi}(0) - R_{\phi\phi}(\tau)] d\tau = \frac{1}{T^2} \int_0^T (T - \tau) \omega_0^2 \tau^2 I^2(\tau) d\tau. \end{aligned} \quad (195)$$

$$\sigma_{\bar{\phi}}^2(T) = \left\langle \left[\frac{1}{T} \int_0^T \phi(t) dt \right]^2 \right\rangle = \frac{2}{T^2} \int_0^T (T - \tau) R_{\phi\phi}(\tau) d\tau. \quad (196)$$

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 (195) describes the accumulated phase noise around its time average. In particular, equation (190) can be reduced to

$$\begin{aligned} \langle C^2(T) \rangle &= \frac{2}{T} \int_0^T \left(1 - \frac{\tau}{T}\right) e^{-\frac{\omega_0^2 \tau^2 I^2(\tau)}{2}} d\tau \\ &\cong \frac{2}{T} \int_0^T \left(1 - \frac{\tau}{T}\right) \left[1 - \frac{\omega_0^2 \tau^2 I^2(\tau)}{2}\right] d\tau = 1 - \sigma_{\Delta\phi}^2(T), \end{aligned} \quad (197)$$

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

$$I^2(\tau) = \frac{2 H_2 \omega_h}{\pi \tau^2}, \quad \text{and} \quad \sigma_y^2(\tau) = \frac{3 H_2 \omega_h}{\pi \tau^2}, \quad (198)$$

for the TVAR and the AVAR, and

$$S_{\phi\phi}(\omega) = \omega_0^2 H_2, \quad \text{and} \quad R_{\phi\phi}(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_{\phi\phi}(\omega) e^{i\omega\tau} d\omega = \omega_0^2 H_2 \delta(\tau), \quad (199)$$

for the power spectrum and the autocorrelation of the phase noise. Then, equations (195) and (196) yield

$$\sigma_{\Delta\phi}^2(T) = \frac{1}{T^2} \int_0^T (T - \tau) \omega_0^2 \tau^2 I^2(\tau) d\tau = \frac{H_2 \omega_0^2 \omega_h}{\pi}, \quad (200)$$

for the dispersion of the residual phase noise, and

$$\sigma_{\phi}^2(T) = \frac{2}{T^2} \int_0^T (T - \tau) R_{\phi\phi}(\tau) d\tau = \frac{H_2 \omega_0^2}{T}, \quad (201)$$

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 $\sigma_{\Delta\phi}^2(T)$ does not depend on the averaging time T . This dispersion is readily estimated, if we have measured AVAR $\sigma_y^2(\tau)$, from which we can easily extract the coefficient $H_2 \omega_h$.

On the other hand, the dispersion of the time-averaged phase noise $\sigma_{\phi}^2(T)$ 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 $\sigma_{\phi}^2(T)$, 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, clocks of 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 a typical bandwidth of $B = 2$ MHz, we needed high accuracy of time synchronization:

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

Such accuracy was not easily available from time transfer systems at that time which mainly used surface waves.

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

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

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, commercially available time synchronization systems using signals from GPS (Global Positioning System) Satellites are now capable of providing time synchronization 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

VLBI required recording systems with very high data rate and very large data capacity from the beginning. Recording was indispensable at that time since it was impossible to transfer observed data through cables to a correlator which could be thousands of kilometers apart.

For example, if we wish to record digitized data, with bandwidth $B = 2$ MHz, sampled at Nyquist rate (which is equal to $2B$ 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. This was not an easy requirement in the 1960's. Now we use recording systems with 1 Gbit/sec rate and 4 Tbit capacity.

Recording was so important for VLBI that generations of "VLBI systems" have been marked by development of the digital data recording technology, as shown in Figure 36.

The Communications Research Laboratory (CRL, now NICT: National Institute of Information and Communications Technology) has developed the Mark II-compatible K1, the experimental real-time VLBI system K2, the Mark III-compatible K3, and cassette-based K4 systems, by its own efforts (Takahashi, et al., 2000).

Nowadays, we have powerful new VLBI systems which use hard disk recorders and even direct data transfer systems based on high-speed optical fiber networks. We will discuss these topics in the next section.

TABLE 9.5 Characteristics of Some VLBI Tape-Recording Systems

System	Period of Use	Basic Description	Tape Recorder	Sample Rate ^e (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. (1995), 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 36: VLBI systems (from Thompson, Moran, and Swenson, 2001).



Figure 37: VLBA recorder (developed by MIT Haystack Observatory/NRAO, USA).

Figures 37, 38, 39 and 40 show the VLBA recorder, S2 recorder, K4 recorder, and the tape handler for K4 recorder. Figure 41 shows the S2, K4, and VLBA (Mark IV) tapes.



Figure 38: S2 recorder (developed by Center for Research in Earth and Space Technology, Canada).



Figure 39: K4 recorder (developed by NICT, Japan).



Figure 40: Automated tape handler for K4 recorder (developed by NICT, Japan).



Figure 41: Magnetic tapes used in VLBI: S2 (top left), K4 (left), and VLBA/Mark IV (right).

2 Overview of VLBI Systems

2.1 MK-3 As Prototype of Modern VLBI Systems

Figure 42 shows a schematic view of the Mark III / K3 VLBI system, often called “MK-3” system, 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 5, VLBA, K4, 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.

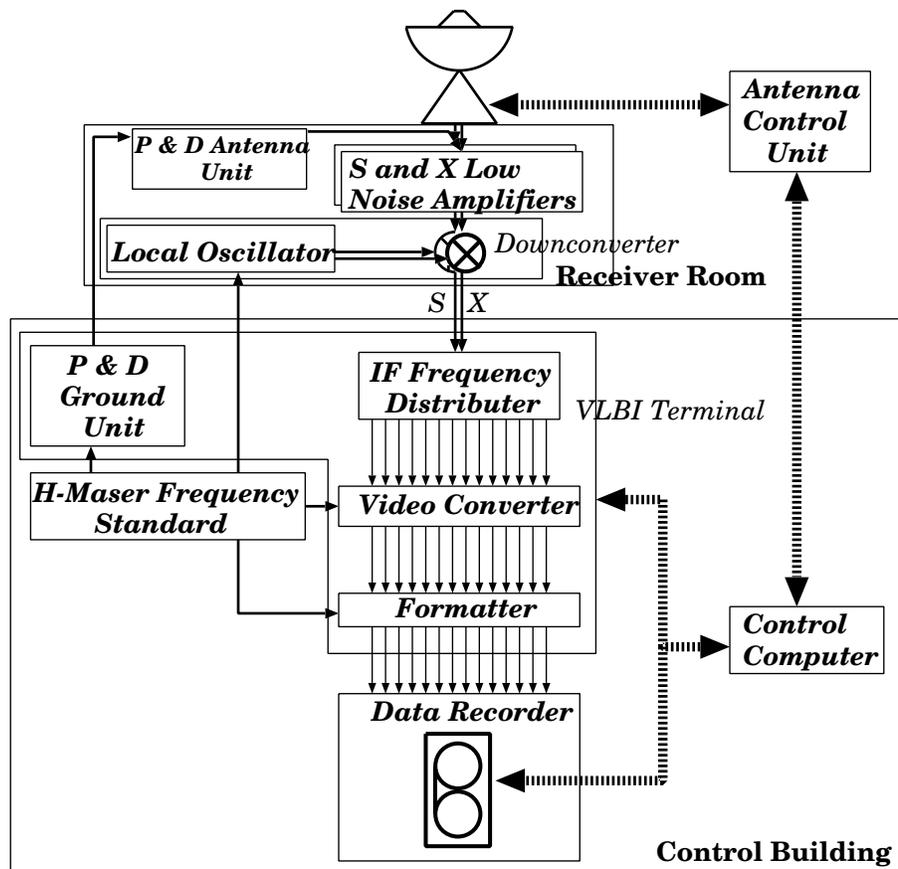


Figure 42: A schematic view of the Mark III / K3 VLBI system.

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 for realizing the so-called “bandwidth synthesis” technique, which we will discuss later, 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 simultaneously at two frequencies, in order to correct the effects of frequency-dependent propagation delay in the ionosphere. For 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 helical arrays (see, e.g., Figure 43), are also used.

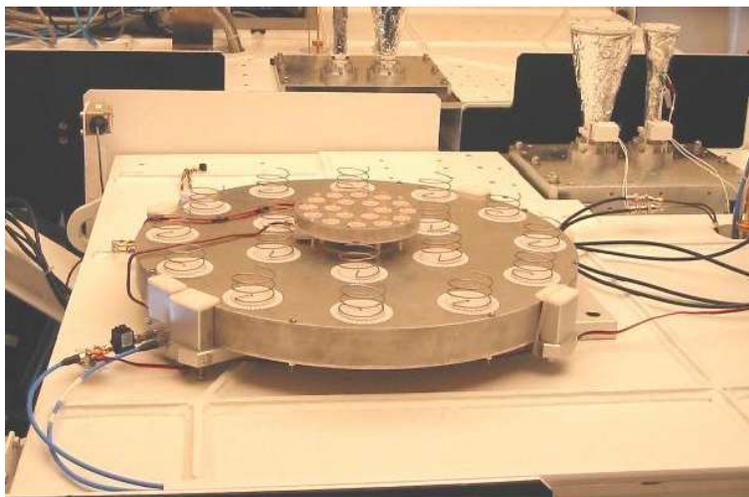


Figure 43: S/X helical-array feed system, developed by Hosei University, Japan, is used for dual-frequency reception in VERA.

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 of Observed Data to Control Building

Then, the IF signals are fed to a so-called MK-3 VLBI terminal rack usually located 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

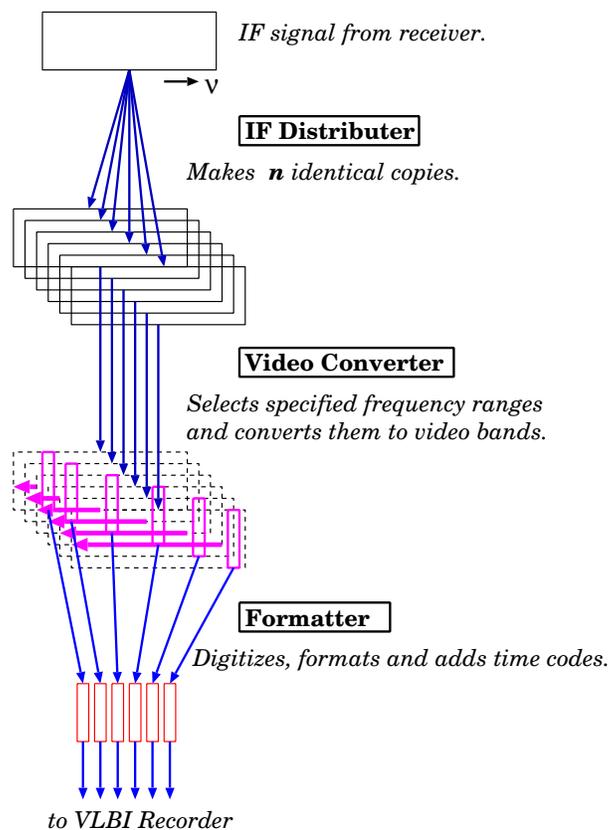


Figure 44: Data flow in the MK-3 VLBI terminal.

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

over to 6 (S-band) and 8 (X-band) identical signals, respectively, by the so-called “Intermediate Frequency Distributer” unit (Figure 44).

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, with 2 MHz bandwidth, say, and downconverts this range into a baseband, typically from 0 to 2 MHz. For this purpose, a builtin frequency synthesizer generates an LO frequency at the edge of the selected IF range, using a reference signal provided by the Hydrogen Maser Frequency Standard. The synthesizer allows flexible setting of the LO frequency with 10 kHz or finer step. 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 into separate basebands. Thus, in total up to 28 (14×2 SB’s) baseband frequency channels (“IF channels” or “BB 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 45 show an example of distribution of RF frequency ranges used in a recent geodetic VLBI observation with 16 VC’s. Note that the “LO frequency” in Table 6 is not an LO frequency of a Video Converter, but a “total” LO frequency which converts an RF band to a baseband. We call such distribution of RF frequency ranges as “frequency distribution of IF channels”.

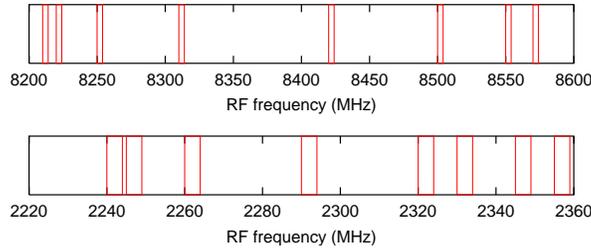


Figure 45: Frequency distributions of IF channels within observed RF frequency ranges, in a sample geodetic VLBI observation, as used in Table 6. Top: X-band, and bottom: S-band.

As we see in Figure 45, frequency distribution of IF channels here appears a little strange. RF frequency ranges converted to IF channels are somewhat

irregularly spread over fairly wide ranges of observed S (2 GHz) and X (8 GHz) bands. Such frequency distribution of IF channels is especially designed for better estimation of group delay observables via the bandwidth synthesis technique, which will be discussed later.

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 frequency distribution of IF channels. Note that this example is taken from a recent geodetic VLBI observation, which uses the K4 VLBI system. The K4 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, in a one-bit (i.e. two-level, at the time

of MK-3) quantization scheme, and with Nyquist sampling rate (i.e. $2B_{VC}$ samples/sec, where $B_{VC} \leq 2$ MHz is the channel bandwidth). At the same time, the Formatter periodically generates time mark codes, using reference signal and clock pulse provided by the Hydrogen Maser Frequency Standard, and overwrites them over a part of the digitized data. These time mark codes play basic and important roles later in the digital correlation processing.

2.1.7 Data Recorder

Up to 28 channels of the digitized and 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 some of its successors (see Figure 37). 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 videoband signals. It is necessary, for accurate estimation of the group delay, to calibrate the phase offsets among the videobands due mainly to the VC initial phases. For that purpose, a special device called the "Phase and Delay Calibrator" is used (Figure 46). 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 ("tones") 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 comb-tones, usually at integer multiples of 1 MHz in the frequency domain, and adds them to the observed data in the RF band. In actual observations, LO frequencies of Video Converters are selected in such a way that at least one comb-tone signal falls into each of the videobands. In the MK-3 system a comb-tone usually appeared at a frequency of 10 kHz in a baseband. This is why we had odd LO frequencies like 8209.99 MHz, 8219.99 MHz, ... in Table 6.

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 basebands are compared with each other, in order to estimate, and then remove from the observed data, the instrumental phase offsets due mainly to the VC initial phases.

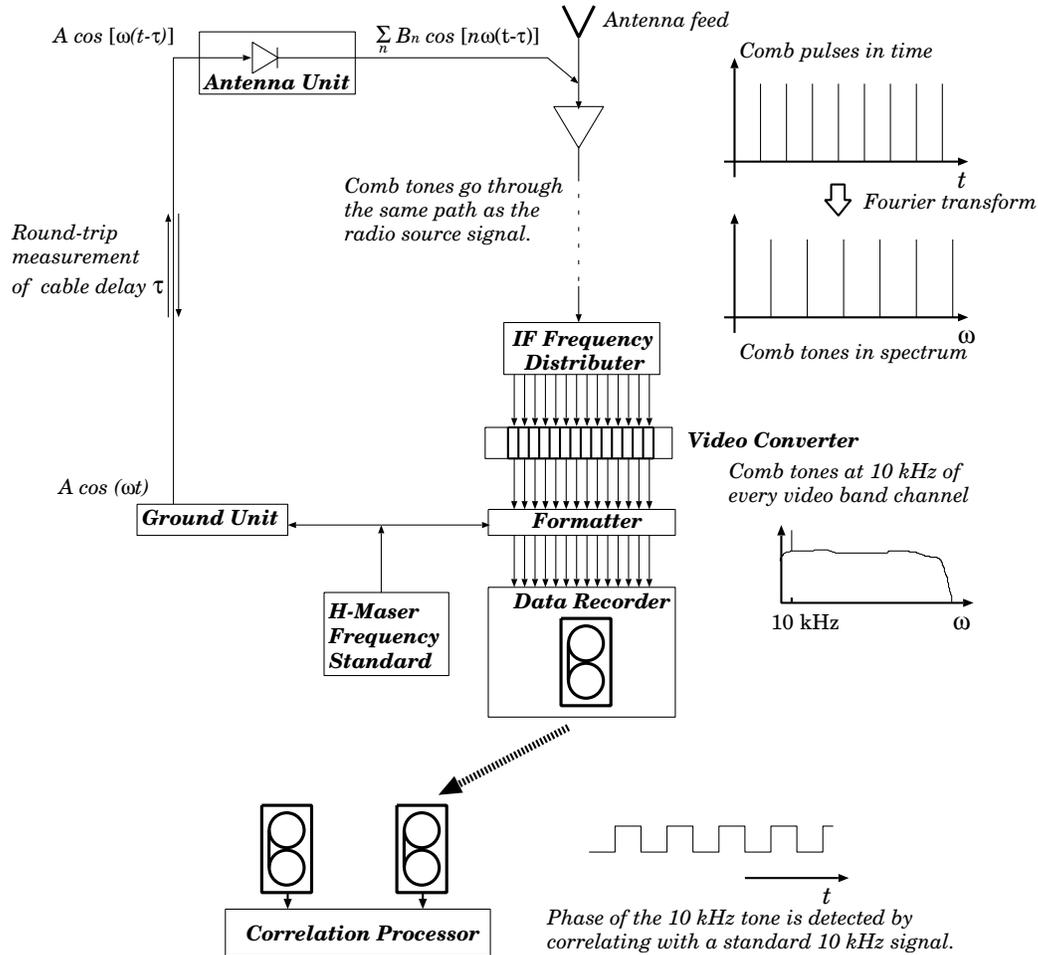


Figure 46: Phase and delay calibration system in the MK-3.

2.1.9 Hydrogen Maser Frequency Standard

The Hydrogen Maser Frequency Standard provides very high stability reference signals, with Allan Standard Deviation of 10^{-15} during 1000 sec, say, usually at 5 MHz or 10 MHz, to first downconverters in the receiver room, Video Converters, Formatters, and 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 be protected against external disturbances as much as possible. Figure 47 shows a Hydrogen Maser Frequency Standard installed in the VERA Mizusawa station.



Figure 47: 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 48). All commands for initial setting of the system such as LO frequencies, sampler modes, recording rates, etc., are remotely given by the control computer. Also, commands for antenna operations, recorder 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 observed radio sources, and station locations are listed. Also, detailed time sequence of observing events, such as repointing the antenna to a new

source, tape start, tape stop, and so on, are given in Universal Times (UT). 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

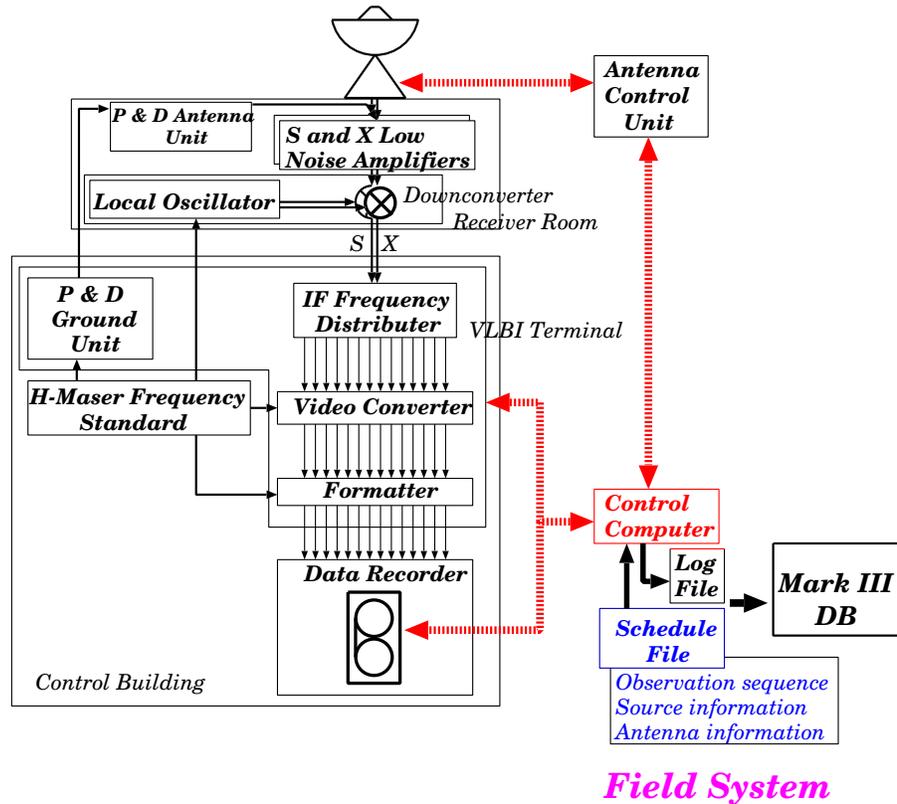


Figure 48: A typical automated VLBI operation system.

recording tapes once an hour or 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 36. They are the VLBA and Mark IV systems based on the advanced head controls of the

Honeywell M-96 recorder (Figure 37), the S2 system using 8 video cassette recorders in parallel (Figure 38), and the K4 system based on an ID1 digital video cassette recorder (Figure 39).

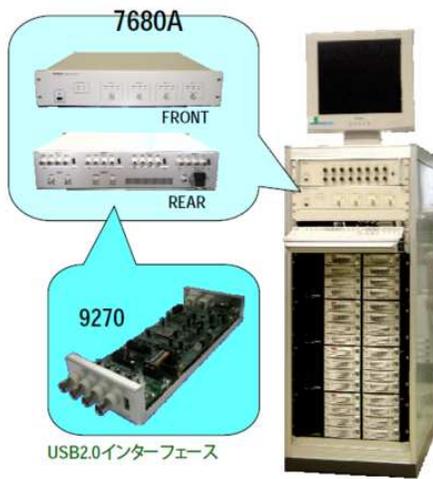
Recently, modern VLBI systems with 1 Gbps or higher data rate have been successfully developed (Table 7). They include the Mark 5 system using a hard disk array instead of the magnetic tapes (Figure 49), the K-5 system also based on the hard disk array (Figure 50), the GBR-1000 system based on an HDTV video cassette recorder (Figure 51), the VERA recording system based on a new ID1 digital video cassette recorder (Figure 52), and the direct data transfer technique based on optical fiber network which is called “e-VLBI”.

name	type	bitrate	main developer
Mark 5	hard disk array	1024 Mbps	MIT Haystack Obs.
K-5	hard disk array	512 & 2048 Mbps	NICT
GBR-1000	HDTV	1024 Mbps	NICT
VERA	ID1	1024 Mbps	NOAJ
e-VLBI	real-time fiber link	> 2048 Mbps	worldwide

Table 7: New generation recording and fiber-link systems.



Figure 49: Mark 5B recording system based on a new hard disk array design.



Specifications

Sampling rate (MHz)	0.04, 0.1, 0.2, 0.5, 1, 2, 4, 8, 16, 32, 64
Digital LP filter (MHz)	2, 4, 8, 16, through
Analog bandwidth	300 MHz
AD bits	1, 2, 4, 8
# of ch/unit	1, 4
Maximum data rate	64 Mbps/ch, 256 Mbps/unit, 1024 Mbps/4 unit
DC offset	adjustable from host
Reference signal	1 PPS, 10 MHz / 5 MHz
PC Interface	USB2.0

Figure 50: K-5 VSSP32 VLBI system based on PC architecture with hard disk array (left) and its specifications (right). K-5 VSSP32 is not a mere recorder but Sampler + Digital low-pass filter + Formatter + Recorder.



Figure 51: GBR-1000 1 Gbps VLBI system based on an HDTV video cassette recorder.



Figure 52: VERA recording system based on a new ID1 digital video cassette recorder.

2.2.2 High-Speed Samplers and Digital Baseband Converters

Another remarkable progress is being made in several countries for implementing digital BBC (baseband converter) systems based on high-speed sampling and digital filtering technologies. The idea is first to digitize a wideband

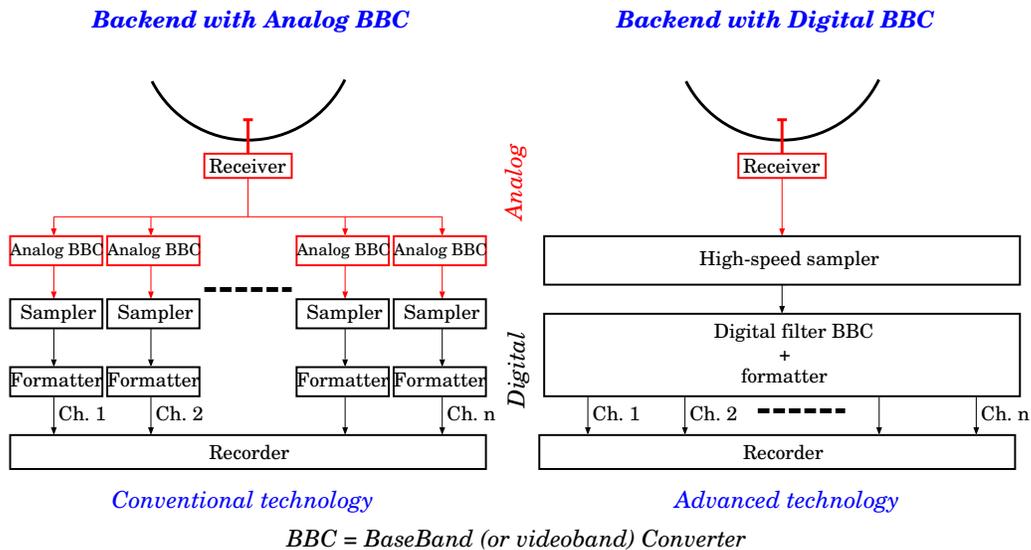


Figure 53: Data acquisition systems using analog BBC(left) and digital BBC(right).

IF data using a high-speed sampler (analog-to-digital converter), and second to “slice” the wideband into baseband channels by means of a high-speed digital filter. Figure 53 illustrates principles of the digital BBC system (left) in comparison with the conventional analog BBC system (right).

The digital BBC system has several advantages over the analog one. First, multichannel baseband data generated by a digital BBC system will no longer suffer from individual bandpass characteristics of analog BBC's. Second, they will no longer involve unequal initial phases added by LO's of the BBC's, which are not predictable beforehand. As a result, the digital BBC system can in principle remove the phase calibration systems discussed before, as far as the LO phase problem only is concerned. Third, the digital BBC system allows flexible organization of bandwidth of baseband channels, in order to meet various scientific requirements. For example, a single channel of wideband (256 MHz, say) data could be favorable for imaging a very weak continuum source, while 16 channels of 8 MHz-wide data could be suited to geodetic VLBI observations. A single digital BBC system with a suitable digital filter circuit can realize both. In case of an analog BBC system, however, number of baseband channels is limited by number of analog BBC units in the system and allowable bandwidth of a baseband channel is determined by analog filters built in a BBC. Finally, a digital BBC system can be made much less expensive than an equivalent analog BBC system.

Figure 54 and Table 8 show examples of existing high-speed samplers



Figure 54: High-speed samplers ADS1000, ADS3000 developed by NICT, and iBOB developed by Haystack Observatory.

	ADS1000	ADS3000	iBOB
Sampling rate	1024 MHz	2048 MHz	dual 1024 MHz
AD bits	1, 2	2, 4, 8	1 – 8
Number of ch.	1	FPGA programmable	FPGA programmable
Max. data rate	2048 Mbps	4096 Mbps	4096 Mbps

Table 8: Main performance parameters of existing high-speed samplers.

used in digital BBC systems.

Note that 2048 Msps (mega sample per second) system is capable of sampling 1024 MHz-wide frequency band at Nyquist rate. Such a wideband can cover almost all observing frequency bands currently used in VLBI. Also note that ADS3000 and iBOB are equipped with FPGA (Field Programmable Gate Array) which are capable of filtering and formatting digitized data. In this sense, they are not mere samplers, but digital BBC systems themselves with digital filtering functions.

Even higher sampling rate is challenged. Figure 55 shows a preliminary result of direct sampling of a K band signal. Such an “RF sampling” technology, if applied to radio interferometry, will enable us to realize a fine delay tracking at RF band, which will make a separate fringe stopping unnecessary, as we saw in Chapter 3.

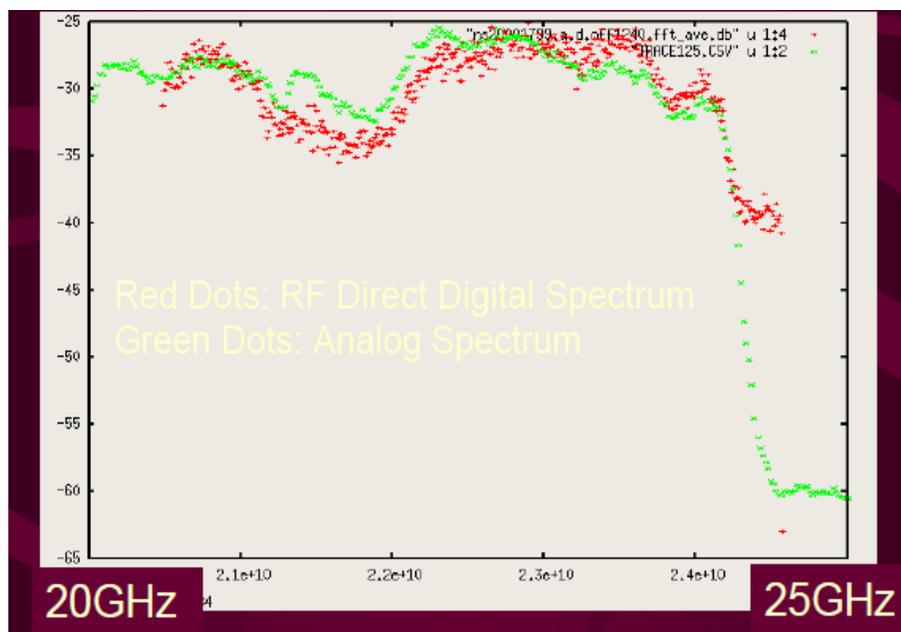


Figure 55: “RF sampling” result at K band obtained with a prototype InP HBT(Indium Phosphide Heterojunction Bipolar Transistor) ultra-high-speed sampler. A power spectrum calculated from the RF-sampled digital data (red) is compared with the one obtained with an analog spectrometer (green) (Kawaguchi et al., 2007, IVSTDC Symp.).

A digital filtering system for VLBI was developed and successfully utilized for the first time in 4 stations of the VERA array. This system uses ADS1000 as a sampler.

There are still some limitations in flexible selection of frequency distribution of IF channels under the current level of digital filtering technology. Nevertheless, digital BBC systems already support many observation modes.

For example, Figure 56 shows Data Acquisition System (DAS) of the KVN

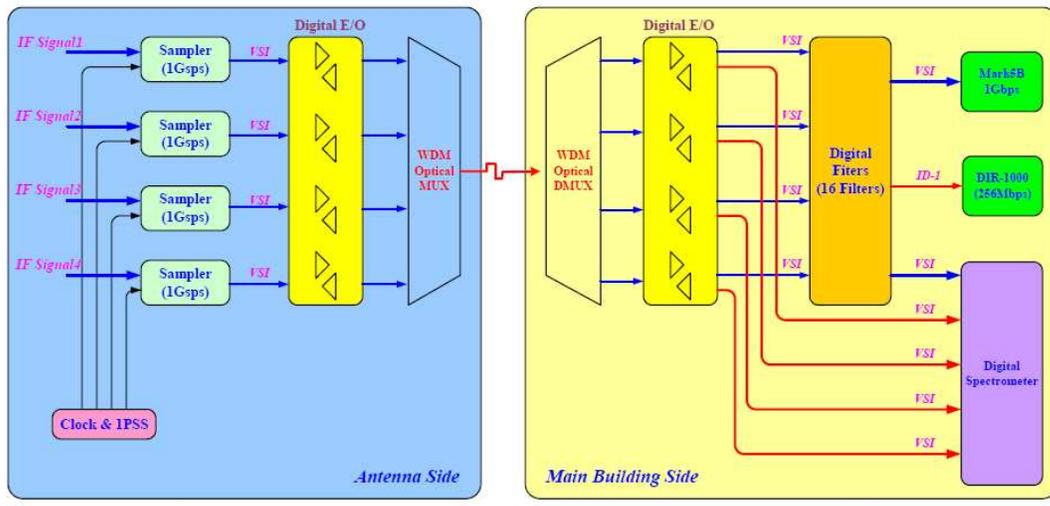


Figure 56: Data Acquisition System of the KVN adopting the digital filter system for the baseband conversion.

Mode	Bandwidth [MHz]	Output streams	Bits/Sample	Output data rate [Mbps]	Output clock speed [MHz]
1	256	1	2	1024	32
2	128	2	2	1024	32
3	64	4	2	1024	32
4	32	8	2	1024	32
5	16	16	2	1024	32
6	8	16	2	512	16
7	64/128	2/1	2	1024	32
8	32/64/128	2/1/1	2	1024	32
9	32/128	4/1	2	1024	32
10	16/32/128	2/3/1	2	1024	32

Table 9: Observing modes to be supported by the digital filter of the KVN.

(Korean VLBI Network) which uses a digital BBC system. The DAS system includes 4 high-speed samplers located in the receiver room (antenna

side) and digital filter units in the control room (main building side). 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 maximum 512 MHz bandwidth with 1024 Msps speed in the 2-bit quantization mode (therefore, maximum output bit rate is 2 Gbps), using the higher-order sampling technique discussed earlier. The digital data are transmitted to the main building where the digital filter units form 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 Gbps input, depending on scientific purposes. Table 9 shows observation modes to be supported in the KVN.

Now efforts are made to enhance digital BBC systems to support all existing frequency distributions of IF channels. For example, very flexible frequency setting with 0.5 kHz step is planned in an advanced version of ADS3000 based on a high performance FPGA.

2.2.3 e-VLBI

A number of groups in the world are developing VLBI systems based on high-speed data transmission techniques via optical fiber cables. The idea is to replace the data tapes by ultra-wide-band transmission cables. This technology is now called “e-VLBI”.

Of course, this technology still does not transform VLBI into a connected-element interferometer, since the frequency standards in different stations remain independent. It is still essentially difficult to transmit reference signals generated by a common frequency standard to stations thousands of kilometers apart without significant delays or phase fluctuations, although some efforts are being made in this direction using “round trip” technique.

Nevertheless, the e-VLBI 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.

For example, pioneering real-time e-VLBI observations with 256 Mbps data rate, which were conducted in the KSP (Key Stone Project, 1995–2001) of the CRL (now the NICT), regularly yielded final geodetic results a few minutes after each observation.

Moreover, the optical fiber cables offer even higher data transmission rates, and therefore higher sensitivity, than the magnetic tapes, or hard disk arrays. This is especially true when dedicated optical fiber cables are available for VLBI purposes. Real-time VLBI experiments with 2.5 Gbps transmission rate have been successfully conducted since 1998 in “VONUS” (“VLBI Optically linked Network Using Super-sinet”) project and its prede-

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.

Figure 58 shows an intercontinental geodetic e-VLBI network connecting Wettzell, Germany, Haystack Observatory, USA, and Tsukuba, Japan, via the broadband Internet. This network is regularly in use for transmitting 24 hour and 1 hour geodetic observation data with maximum transfer rates of 110 Mbps and 25 Mbps, respectively. Currently, more than 50 hours are needed for transferring 24 hour data, depending on the traffic of the Internet.

Figure 59 shows the first successful EVN (European VLBI Network) e-VLBI experiment (“Internet Telescope”) conducted by three observatories, in

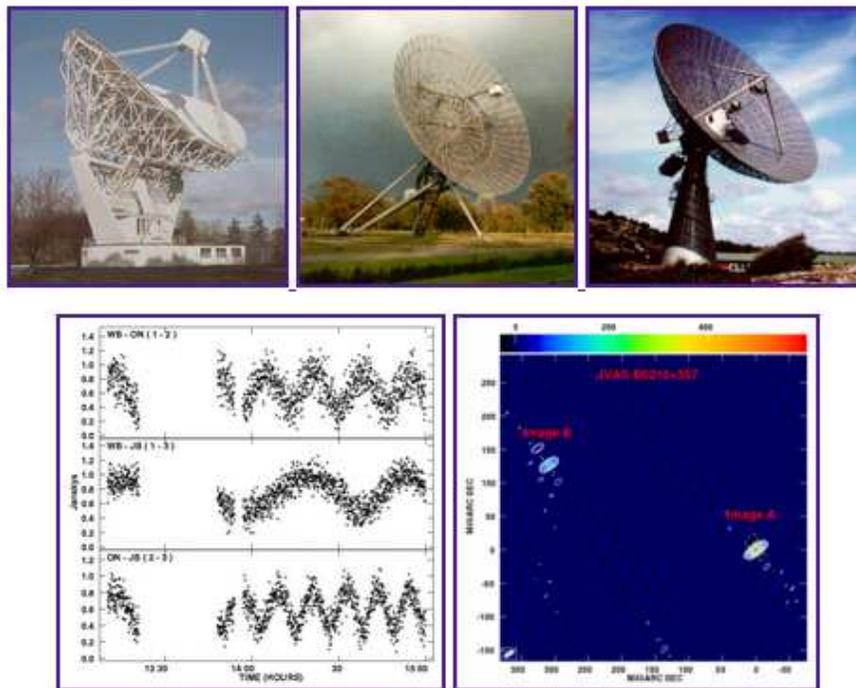


Figure 59: First EVN e-VLBI observation (2004 April) using radio telescopes at Jodrell Bank, UK; Westerbork, the Netherlands; and Onsala, Sweden. The lower panels show beats of the fringe amplitudes produced by two closely spaced sources (gravitational lens object JVAS B0218+357), and an image map of the object obtained during the observation (figure courtesy of KVN, <http://www.jive.nl>).

April 2004, and a beautiful image of the gravitational lens object B0218+357 obtained through this coordinated effort.

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.

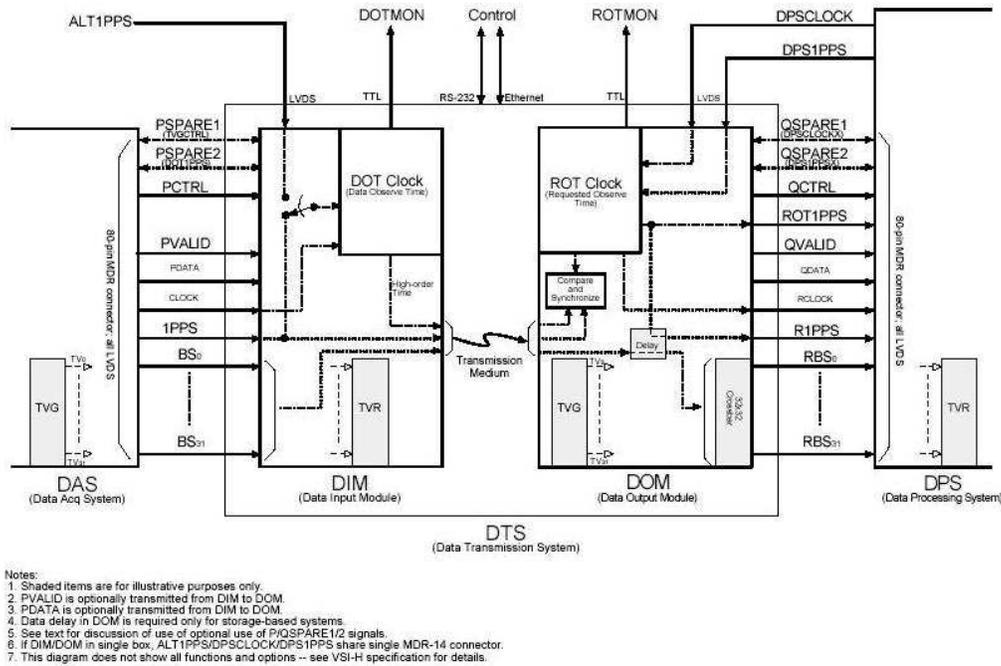


Figure 60: VSI-H functional diagram (<http://web.haystack.edu/vsi>).

The world VLBI community has long suffered from incompatibility of different VLBI systems, developed in different institutions. All of them have their own excellent points. However, in order to cross-correlate data recorded on different tapes by different VLBI systems A and B, say, one had to convert data of B’s format to data of 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.

A hardware specification called “VSI-H”, a software specification called “VSI-S”, and a VLBI data interface specification called “VDIF” are defined (<http://www.vlbi.org/vsi>).

The VSI-H hardware specification defines a standard interface between the “data-acquisition system (DAS)” (i.e. VLBI terminals) and the “data-transmission system (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 60 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 input module (DIM)”, and a “data output module (DOM)”, which interface the common connectors and 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 their data have been successfully cross-correlated with each other.

2.2.5 Future Geodetic VLBI System Proposed in “VLBI2010”

In the light of growing roles of precise global geodetic measurements in Earth and Space sciences together with recent advances in VLBI technology, the IVS (International VLBI Service for Geodesy & Astrometry) published in September, 2005, a working-group report called “VLBI2010”. In this document, the IVS proposed a concept of future geodetic VLBI system which shall replace a little obsolete system currently used in international observations. Goals for the new system were outlined as

- 1 mm measurement accuracy on global baselines,
- continuous measurements for time series of station positions and Earth orientation parameters,
- turnaround time to initial geodetic results of less than 24 hrs.

Main points of the system proposed in 2005 and further enhanced in subsequent discussions include the followings. These ambitious plans will eventually change appearance of the world VLBI.

Antenna	10~12 m dish; 60% efficiency; Az & El slew rates $> 5^\circ/\text{sec}$
Feed	Dual polarization; low cross-polarization leakage
Frontend	2–18 GHz continuous RF coverage, or 8/32 GHz; $T_{\text{sys}} \cong 45 \text{ K}$
Backend	Early digitization; flexible channelization
Data rate	2~4 Gbps initially, expanding to 8~16 Gbps, or 32 Gbps
Transport	Mixture of disk recording and high-speed network transfer
Correlation	Near real time, perhaps distributed among processors

3 Effects of Clock Offset and Atmospheric Delay

At the present time, VLBI is the most powerful technology for high-resolution radio source imaging and also for high-precision geodesy and astrometry. However, it has its own limitations and difficulties, not surprisingly. Besides obvious heavy loads incurred in data transportation, as well as in the operation and maintenance of distant stations, we saw earlier a fundamental limitation that VLBI is able to observe only very bright and compact radio sources, at least at the present level of sensitivity.

Furthermore, VLBI suffers from two more essential difficulties, compared with connected-element interferometers, as long as it is located on the surface of the Earth. One is an effect of unpredictable frequency fluctuations in the independent frequency standards, which causes time-variable offsets in station clocks and their rates, and the other is an effect of refraction in turbulent atmosphere, which causes highly irregular and independent propagation delays at stations separated by great distances. We will discuss these effects in the present section.

3.1 Clock Offset.

3.1.1 Measuring Time with Clocks

The clock time value t_C is measured “by counting number of cycles \mathcal{N} of an oscillator, and multiplying it to the nominal period P_0 of the oscillator”, i.e.

$$t_C = \mathcal{N} \times P_0. \quad (202)$$

Since $2\pi\mathcal{N}$ is the phase ϕ of the oscillator, the clock time is also given by

$$t_C = \frac{\phi \times P_0}{2\pi} = \frac{\phi}{\omega_0}, \quad (203)$$

i.e. “by measuring the phase ϕ of the oscillator, and dividing it by the nominal angular frequency ω_0 ”, which is equal to

$$\omega_0 = 2\pi / P_0, \quad (204)$$

as illustrated in Figure 61.

Of course, if the the true period, or the true frequency, of the oscillator deviates from the nominal period P_0 , or the nominal frequency ω_0 , we cannot obtain the correct time value. This is the reason why we have to consider the “clock offset” problem.



Figure 61: Usually, we look at phases of hands of a clock (rotation angles of the hands) and divide them by nominal angular frequencies of the hands (360° per 60 seconds for second hand, 360° per 60 minutes for minute hand, and 360° per 12 hours for hour hand) to know the time.

3.1.2 Clock Time Deviation Due to Fluctuations in Frequency Standard

In VLBI, the “oscillator” is the sinusoidal reference signal, provided from a frequency standard (usually the Hydrogen Maser). Such a frequency standard is often called the “clock”, though it does not have any hands. The actual instantaneous frequency of the reference signal $\omega_a(t)$ always deviates, though slightly, from its nominal value ω_0 , as

$$\omega_a(t) = \omega_0 + \delta\omega(t), \quad (205)$$

due to finite stability of the frequency standard.

The phase $\phi(t)$ of the reference signal and the actual frequency $\omega_a(t)$ are related to each other by an equation:

$$\frac{d\phi(t)}{dt} = \omega_a(t),$$

or equivalently

$$\phi(t) = \int_{t_i}^t \omega_a(t) dt = \omega_0(t - t_i) + \int_{t_i}^t \delta\omega(t) dt, \quad (206)$$

where t_i is an initial epoch of the phase count.

The measured clock time value t_C , i.e. “the phase divided by the nominal frequency”:

$$t_C = \frac{\phi(t)}{\omega_0} = t - t_i + \int_{t_i}^t \frac{\delta\omega(t)}{\omega_0} dt, \quad (207)$$

deviates from the true time $t - t_i$ by a “clock time deviation” Δt_C :

$$\Delta t_C \equiv t_C - (t - t_i) = \int_{t_i}^t \frac{\delta\omega(t)}{\omega_0} dt. \quad (208)$$

Note that the ratio $\delta\omega(t)/\omega_0$ here is the “fractional frequency deviation” which we introduced in Section 1.2.2.

3.1.3 LO Phase Deviation Due to Fluctuations of Frequency Standard

The frequency deviation also causes a slight deviation $\Delta\omega_{LO}(t)$ of the actual Local Oscillator (LO) frequency $\omega_{LOa}(t)$ from its nominal value ω_{LO} :

$$\omega_{LOa}(t) = \omega_{LO} + \Delta\omega_{LO}(t). \quad (209)$$

Since the LO signal is usually generated from the reference signal by frequency multiplication with a constant ratio ω_{LO}/ω_0 , the actual LO frequency is given by

$$\omega_{LOa}(t) = \omega_a(t) \frac{\omega_{LO}}{\omega_0} = [\omega_0 + \delta\omega(t)] \frac{\omega_{LO}}{\omega_0} = \omega_{LO} + \delta\omega(t) \frac{\omega_{LO}}{\omega_0}, \quad (210)$$

and hence the LO frequency deviation is equal to

$$\Delta\omega_{LO}(t) = \frac{\delta\omega(t)}{\omega_0} \omega_{LO}. \quad (211)$$

Now the actual phase ϕ_{LOa} of the LO signal is given by

$$\begin{aligned} \phi_{LOa} &= \int_{t_i}^t \omega_{LOa} dt + \theta' = \omega_{LO} \int_{t_i}^t \left[1 + \frac{\delta\omega(t)}{\omega_0} \right] dt + \theta' \\ &= \omega_{LO}(t - t_i) + \omega_{LO} \Delta t_C + \theta', \end{aligned} \quad (212)$$

where θ' is an initial LO phase at time $t = t_i$. Here we used the relation for the clock time deviation Δt_C given in equation (208).

Then, the actual LO phase ϕ_{LOa} deviates from its nominal value $\phi_{LO_0} = \omega_{LO}(t - t_i)$ by an “LO phase deviation”:

$$\Delta\phi_{LO} \equiv \phi_{LOa} - \phi_{LO_0} = \omega_{LO}\Delta t_C + \theta'. \quad (213)$$

We will assume this equation in further discussions, though this may not strictly hold in cases when the LO frequency is generated by a method other than the simple frequency multiplication of the reference signal.

3.1.4 Clock and LO–Phase Offsets between Two VLBI Stations

Now we consider two stations in a VLBI baseline. Let us assume that

- frequency deviations at the two stations are $\delta\omega_1(t)$, and $\delta\omega_2(t)$, respectively,
- clock time deviations at the two stations are $\Delta t_{C1}(t)$, and $\Delta t_{C2}(t)$, respectively, and
- a synchronization error between the station clocks at the initial epoch $t = t_i$ is τ_{Ci} .

Then, the **clock offset** τ_C between the two stations is given by

$$\tau_C(t) = \Delta t_{C1}(t) - \Delta t_{C2}(t) + \tau_{Ci} = \int_{t_i}^t \frac{\delta\omega_1(t) - \delta\omega_2(t)}{\omega_0} dt + \tau_{Ci}. \quad (214)$$

In actual VLBI observations, the clock offset $\tau_C(t)$ could be several tens of nanoseconds or larger, which far exceeds the VLBI measurement accuracy (~ 0.1 nsec). Therefore, we must estimate and correct the clock offset in an early stage of the correlation processing.

Also, the **LO phase offset** between the two stations due to LO phase deviations $\Delta\phi_{LO1}$ and $\Delta\phi_{LO2}$ at respective stations is given by:

$$\Delta\phi_{LO1} - \Delta\phi_{LO2} = \omega_{LO}[\Delta t_{C1}(t) - \Delta t_{C2}(t)] + \theta'_1 - \theta'_2 = \omega_{LO}\tau_C(t) + \theta_1 - \theta_2. \quad (215)$$

Here, we newly redefined the initial phase offset as

$$\theta_1 - \theta_2 = \theta'_1 - \theta'_2 - \omega_{LO}\tau_{Ci}, \quad (216)$$

which now includes initial values of the LO phases and the effect of the clock synchronization error at the initial epoch.

In actual VLBI observations, time variation of the clock offset due to the mismatch of reference frequencies $\dot{\tau}_C = (\delta\omega_1 - \delta\omega_2)/\omega_0$ could be as large as 10^{-12} sec/sec = 1 psec/sec, which could cause rather rapid variation $\omega_{LO} \dot{\tau}_C$ in the VLBI fringe phase (10 mHz, or 1 cycle per 100 sec, if LO frequency is 10 GHz). Therefore, this effect also must be estimated and corrected in the early stage of the correlation processing.

3.1.5 How Do Fluctuations in Frequency Standards Affect VLBI Data?

We discussed in Chapter 3 cross-power spectra of IF (videoband in our VLBI case) signals v_{F1} and v_{F2} of a single baseline radio interferometer, after the delay tracking (DT) and the fringe stopping (FS). In a positive frequency range $\omega \geq 0$, we derived equations for a Hermitian symmetric cross-power spectrum $S_{v_{F1}v_{F2}}$, which are

$$S_{v_{F1}v_{F2}}^{USB}(\omega) = \frac{1}{4} A_0 e^{-i[\omega_{LO} \Delta\tau_g + \omega \Delta\tau_g + \phi_{LO1} - \phi_{LO2}]} \mathcal{V}(\omega_{LO} + \omega) H_1(\omega) H_2^*(\omega), \quad (217)$$

for the upper sideband (USB) reception case, and

$$S_{v_{F1}v_{F2}}^{LSB}(\omega) = \frac{1}{4} A_0 e^{i[\omega_{LO} \Delta\tau_g - \omega \Delta\tau_g + \phi_{LO1} - \phi_{LO2}]} \mathcal{V}^*(\omega_{LO} - \omega) H_1^*(-\omega) H_2(-\omega), \quad (218)$$

for the lower sideband (LSB) reception case, respectively.

Notations used here are the followings: A_0 is the geometric mean of effective apertures of two antennas, $\Delta\tau_g = \tau_{g0} - \tau_i$ is the residual geometric delay, $\tau_{g0} = \mathbf{D} \cdot \mathbf{s}_0 / c$ is the geometric delay at a reference direction \mathbf{s}_0 in a radio source observed with an interferometer of baseline vector \mathbf{D} , τ_i is the theoretical prediction of the geometric delay used in the DT and FS (“the instrumental delay”), $\phi_{LO1} - \phi_{LO2}$ is the offset of the LO phases, \mathcal{V} is the complex visibility, $H_1(\omega) H_2^*(\omega)$ is the “bandpass characteristics of combined IF filters”, ω_{LO} is the nominal local oscillator frequency, and ω is the IF (videoband) frequency.

Then, how do the fluctuations in the frequency standards modify the cross-power spectrum?

There are two places where the fluctuations in the frequency standards affect the VLBI data (Figure 62).

1. The clock offset τ_C causes an offset in time mark codes of VLBI records, which is equivalent to inserting an additional delay in the videoband.
2. The LO phase offset $\Delta\phi_{LO1} - \Delta\phi_{LO2}$ is added to the phase term in the cross-power spectrum of the videoband signals.

Specifically,

1. we need to add to the $\omega\Delta\tau_g$ term in equations (217) and (218) a contribution of the “additional delay” $\omega\tau_C$ corresponding to the time code offset. Thus we get $\omega(\Delta\tau_g + \tau_C)$ instead of $\omega\Delta\tau_g$ in equations (217) and (218).
2. Also, since the LO phase offset at the two stations due to the phase deviations in the frequency standards is given by equation (215):

$$\Delta\phi_{LO1} - \Delta\phi_{LO2} = \omega_{LO}\tau_C(t) + \theta_1 - \theta_2,$$

we need to replace the LO phase offset term $\phi_{LO1} - \phi_{LO2}$ in equations (217) and (218) by this term.

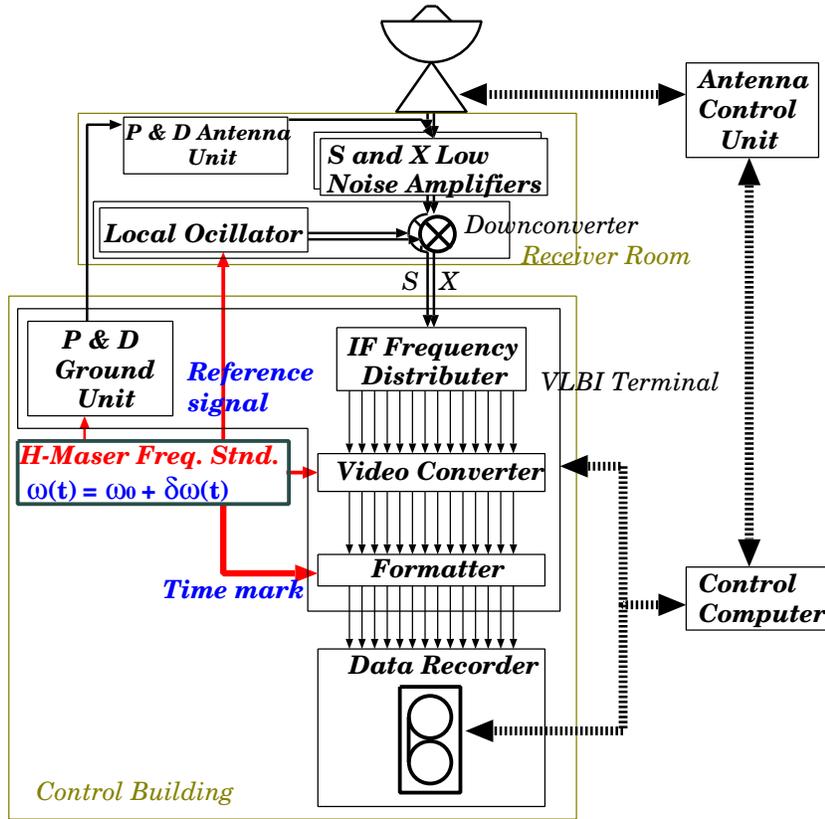


Figure 62: Frequency deviations in frequency standards affect VLBI data through the clock offset in the time mark settings and the LO phase offset in the frequency conversions.

Consequently, the cross-power spectra of the videoband signals in equations (217) and (218) are now modified to

$$S_{v_{F1}v_{F2}}^{USB}(\omega) = \frac{1}{4} A_0 e^{-i[(\omega_{LO}+\omega)(\Delta\tau_g+\tau_C)+\theta_1-\theta_2]} \mathcal{V}(\omega_{LO} + \omega) H_1(\omega) H_2^*(\omega), \quad (219)$$

and

$$S_{v_{F1}v_{F2}}^{LSB}(\omega) = \frac{1}{4} A_0 e^{i[(\omega_{LO}-\omega)(\Delta\tau_g+\tau_C)+\theta_1-\theta_2]} \mathcal{V}^*(\omega_{LO} - \omega) H_1^*(-\omega) H_2(-\omega), \quad (220)$$

for the USB and LSB reception cases, respectively, when the clock offset and the LO phase offset are taken into account.

Thus, effects of the clock offset and the LO phase offset taken together are practically equivalent to inserting an additional delay τ_C in the RF band with frequency $(\omega_{LO} \pm \omega)$!

3.2 Atmospheric Propagation Delay

The Earth's atmosphere causes two kinds of negative effects in VLBI observations, one through absorption and another through refraction.

The absorption of the electromagnetic wave in the atmosphere leads to attenuation of the signal amplitude and generation of the thermal noise radiation, which both deteriorate the signal-to-noise ratio. This effect becomes increasingly severe as we go to higher observing frequency.

Though severe, especially in bad weather conditions, the absorption effect is common in all radioastronomical observations, including single-dish ones. We discussed this effect to some extent in Chapter 2, related to the system-noise-temperature issue. Here we concentrate our attention to the VLBI-specific problems associated with the atmospheric refraction effect.

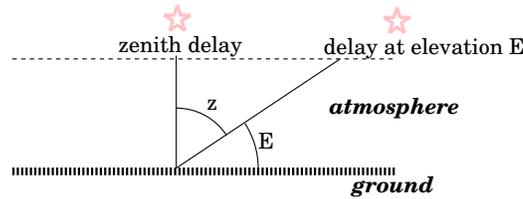


Figure 63: Elevation dependence of the propagation delay.

The refraction in the atmosphere makes the speed of the electromagnetic wave a little slower than in the vacuum. Therefore, the arrival time of a wave front at an antenna is delayed compared with the arrival time in the vacuum.

This additional delay is called the “**propagation delay**” τ_{prop} . The same effect is described also in terms of the “**excess path length (EPL)**” \mathcal{L} , defined by $\mathcal{L} = c \tau_{prop}$, where c is the light velocity. The propagation delay is the smallest at zenith direction, and varies with zenith distance z roughly in proportion to $\sec z$ as expected in a simple model of plane-parallel atmosphere (Figure 63).

Difference of the propagation delays at two stations of a VLBI baseline is called the “**atmospheric delay**” τ_A , and given by $\tau_A = \tau_{prop1} - \tau_{prop2}$, where τ_{prop1} and τ_{prop2} are the propagation delays at station 1 and station 2, respectively (Figure 64).

☆

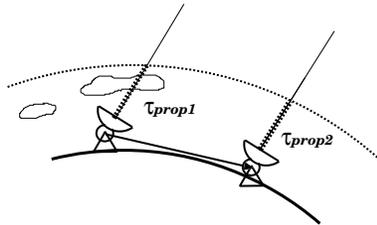


Figure 64: Atmospheric delay.

In many of cm-wave connected-element interferometer arrays, the atmosphere above element antennas are practically the same. Therefore, the propagation delays at individual antennas are also the same, and hence the atmospheric delay is almost zero everywhere. However, this is certainly not the case for the VLBI, unfortunately.

3.2.1 Three Major Components of Atmosphere That Cause Propagation Delay

Figure 65 shows an image of atmospheric layers. It is known that the atmospheric propagation delay in the radio wave is mainly caused by the **ionosphere**, and by the “**dry**” (non-water-vapor) and “**wet**” (water-vapor) components in the neutral atmosphere (troposphere and stratosphere).

- The ionospheric propagation delay is rather large in relatively low frequency. For example, in the daytime of the active period in the solar cycle, the ionospheric EPL at 2 GHz may exceed 10 m. However, the

ionospheric delay has a clear dependence on frequency, and, therefore, can be effectively estimated and removed if we observe a radio source at two different frequencies simultaneously, as we will see later.

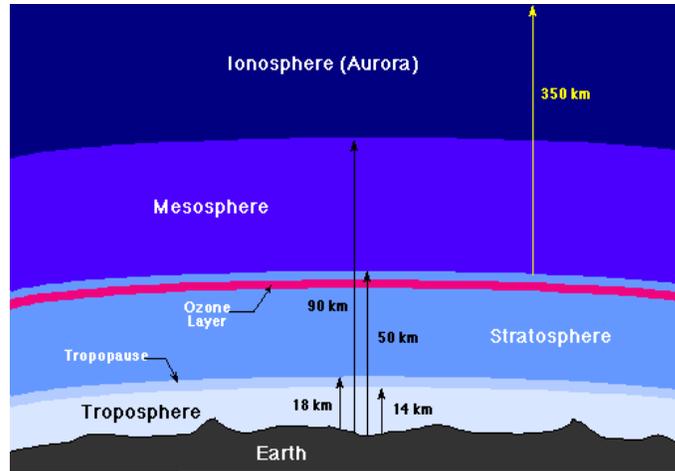


Figure 65: An image of atmospheric layers (Figure brought from a webpage <http://csep10.phys.utk.edu/astr161/lect/earth/atmosphere.html>).

- The propagation delay due to the dry component of the neutral atmosphere is also large. The EPL due to the dry component is usually larger than 2 m. However, the dry delay (or, more exactly, the “hydrostatic delay”) can be almost completely determined by the air pressure at the ground, as we will discuss later. Its time variation is rather slow and can be well traced by using the ground pressure data.
- On the other hand, the refractivity of the water vapor contained mainly in the lower troposphere is a big problem. The refractivity is about 20 times as large in the radio wave region as the one in the optical and near infrared regions, and resulting propagation EPL is as large as a few to several tens cm.

Its time variability is largely caused by wind-blown turbulent distribution of the water vapor in the troposphere, which is highly irregular and almost unpredictable.

Also, it is difficult at present to measure the water vapor distribution in the atmosphere by means of existing ground measurements.

Therefore, it is the water vapor component in the troposphere that causes the most serious propagation delay effect on the VLBI observation.

Typical Allan standard deviation of the wet delay is around 10^{-13} in a timescale around 100 s (Rogers and Moran, 1981), which is much worse than that of modern Hydrogen Maser Frequency Standards (Figure 66).

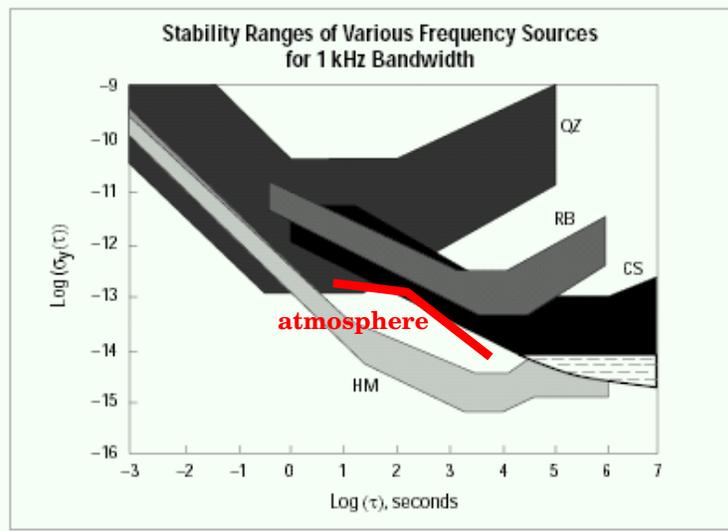


Figure 66: Allan standard deviation vs. time interval plots for various frequency standards as given in Figure 34, and for the atmospheric phase fluctuation.

Currently, the water vapor in the troposphere is the largest factor which causes fluctuation of the VLBI fringe phase.

3.2.2 Cross-Power Spectrum of IF Signals with Atmospheric Delay

Atmospheric delay τ_A occurs before the radio wave arrives at antennas. Therefore, effects of the τ_A must appear in the RF band, just as the geometric delay. Consequently, the VLBI cross-power spectrum of IF (videoband in our VLBI case) signals in the positive frequency range ($\omega \geq 0$) after the DT and FS, with the clock offset and atmospheric delay taken into account, is now given by adding $(\omega_{LO} \pm \omega) \tau_A$ to the arguments of the exponential functions of equations (219) and (220), where + and - correspond to USB and LSB cases, respectively. Thus we have, for the USB reception case,

$$S_{v_{F1}v_{F2}}^{USB}(\omega) = \frac{1}{4} A_0 e^{-i[(\omega_{LO} + \omega)(\Delta\tau_g + \tau_C + \tau_A) + \theta_1 - \theta_2]} \times \mathcal{V}(\omega_{LO} + \omega) H_1(\omega) H_2^*(\omega), \quad (221)$$

and for the LSB reception case,

$$S_{v_{F1}v_{F2}}^{LSB}(\omega) = \frac{1}{4} A_0 e^{i[(\omega_{LO}-\omega)(\Delta\tau_g+\tau_C+\tau_A)+\theta_1-\theta_2]} \times \mathcal{V}^*(\omega_{LO}-\omega) H_1^*(-\omega) H_2(-\omega), \quad (222)$$

where notations are the same with those in equations (217) and (218).

3.3 Expected Correlator Output with Clock and Atmospheric Effects

Now we can derive theoretical expressions of the expected correlation (or expected correlator output) after the DT and FS, by taking inverse Fourier transform, with zero argument, of the cross-power spectrum, as we did earlier in Chapter 3:

$$R_{v_{F1}v_{F2}}(0) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_{v_{F1}v_{F2}}(\omega) d\omega,$$

but now newly taking into account the clock and atmospheric effects.

3.3.1 Expected Correlator Output, General Expression

The inverse Fourier transformation of equations (221) and (222) gives, for the USB reception case,

$$R_{v_{F1}v_{F2}}^{USB}(0) = \frac{A_0}{4\pi} \Re \left\{ e^{-i[\omega_{LO}(\Delta\tau_g+\tau_C+\tau_A)+\theta_1-\theta_2]} \times \int_0^{\infty} \mathcal{V}(\omega_{LO}+\omega) e^{-i\omega(\Delta\tau_g+\tau_C+\tau_A)} H_1(\omega) H_2^*(\omega) d\omega \right\}, \quad (223)$$

while for the LSB reception case,

$$R_{v_{F1}v_{F2}}^{LSB}(0) = \frac{A_0}{4\pi} \Re \left\{ e^{i[\omega_{LO}(\Delta\tau_g+\tau_C+\tau_A)+\theta_1-\theta_2]} \times \int_0^{\infty} \mathcal{V}^*(\omega_{LO}-\omega) e^{-i\omega(\Delta\tau_g+\tau_C+\tau_A)} H_1^*(-\omega) H_2(-\omega) d\omega \right\} \quad (224)$$

where notations are the same with those in equations (217) and (218), except for the clock offset τ_C , the atmospheric delay τ_A , and the LO initial phase offset $\theta_1 - \theta_2$ which we introduced in subsequent discussions.

These are general expressions for the expected correlations after the DT and FS, with the clock and atmospheric effects taken into account.

3.3.2 Correlator Output for Continuum Spectrum Source

For a continuum spectrum source with almost constant complex visibility across the receiving band, the expected correlations after the DT and FS given by equations (223) and (224) are reduced to

$$R_{v_{F1}v_{F2}}^{USB}(0) = A_0 |\mathcal{V}^U| |\mathcal{B}_{12}^U| \cos[(\omega_{LO} + \omega_I) (\Delta\tau_g + \tau_C + \tau_A) + \theta_1 - \theta_2 - \Phi_v^U - \Phi_B^U], \quad (225)$$

for the USB reception case, and

$$R_{v_{F1}v_{F2}}^{LSB}(0) = A_0 |\mathcal{V}^L| |\mathcal{B}_{12}^L| \cos[(\omega_{LO} - \omega_I) (\Delta\tau_g + \tau_C + \tau_A) + \theta_1 - \theta_2 - \Phi_v^L + \Phi_B^L], \quad (226)$$

for the LSB reception case, respectively, as we saw in Chapter 3, where ω_I is the central frequency of the videoband, $|\mathcal{V}|$ and Φ_v are amplitude and phase of the complex visibility:

$$\mathcal{V}(\omega_{LO} + \omega_I) = |\mathcal{V}^U| e^{i\Phi_v^U}, \quad (227)$$

in the USB reception case, and

$$\mathcal{V}(\omega_{LO} - \omega_I) = |\mathcal{V}^L| e^{i\Phi_v^L}, \quad (228)$$

in the LSB reception case, while $|\mathcal{B}_{12}|$, and Φ_B are amplitude and phase of the bandwidth pattern \mathcal{B}_{12} with bandwidth $\Delta\omega$ of the two antennas 1 and 2:

$$\mathcal{B}_{12}^{USB}(B, \Delta\tau_g + \tau_C + \tau_A) = |\mathcal{B}_{12}^U| e^{i\Phi_B^U} \quad (229)$$

in the USB reception case, and

$$\mathcal{B}_{12}^{LSB}(B, \Delta\tau_g + \tau_C + \tau_A) = |\mathcal{B}_{12}^L| e^{i\Phi_B^L}, \quad (230)$$

in the LSB reception case, with $B = \Delta\omega / (2\pi)$. Here, the bandwidth patterns are given by

$$\mathcal{B}_{12}^{USB}(B, \Delta\tau_g + \tau_C + \tau_A) = \frac{1}{4\pi} \int_{-\frac{\Delta\omega}{2}}^{\frac{\Delta\omega}{2}} e^{-i\omega'(\Delta\tau_g + \tau_C + \tau_A)} H_1(\omega_I + \omega') H_2^*(\omega_I + \omega') d\omega', \quad (231)$$

in the USB reception case, and

$$\mathcal{B}_{12}^{LSB}(B, \Delta\tau_g + \tau_C + \tau_A) = \frac{1}{4\pi} \int_{-\frac{\Delta\omega}{2}}^{\frac{\Delta\omega}{2}} e^{-i\omega'(\Delta\tau_g + \tau_C + \tau_A)} H_1^*(-\omega_I - \omega') H_2(-\omega_I - \omega') d\omega', \quad (232)$$

in the LSB reception case, respectively.

3.3.3 Simplest Case of Rectangular Filters

If we assume the ideally simple case of rectangular filters with the gain factor G and bandwidth B , then the bandwidth patterns are given by equations

$$\mathcal{B}_{12}^{USB}(B, \Delta\tau_g + \tau_C + \tau_A) = \mathcal{B}_{12}^{LSB}(B, \Delta\tau_g + \tau_C + \tau_A) = \frac{GB}{2} \frac{\sin(\pi B(\Delta\tau_g + \tau_C + \tau_A))}{\pi B(\Delta\tau_g + \tau_C + \tau_A)},$$

as we saw in Chapter 3. Therefore, equations (225) and (226) of the expected correlations for a continuum spectrum source after the DT and FS are reduced to

$$R_{v_{F1}v_{F2}}^{USB}(0) = \frac{A_0GB}{2} |\mathcal{V}| \frac{\sin(\pi B(\Delta\tau_g + \tau_C + \tau_A))}{\pi B(\Delta\tau_g + \tau_C + \tau_A)} \times \cos((\omega_{LO} + \omega_I)(\Delta\tau_g + \tau_C + \tau_A) + \theta_1 - \theta_2 - \Phi_v), \quad (233)$$

for the USB reception case, and

$$R_{v_{F1}v_{F2}}^{LSB}(0) = \frac{A_0GB}{2} |\mathcal{V}| \frac{\sin(\pi B\Delta(\tau_g + \tau_C + \tau_A))}{\pi B(\Delta\tau_g + \tau_C + \tau_A)} \times \cos((\omega_{LO} - \omega_I)(\Delta\tau_g + \tau_C + \tau_A) + \theta_1 - \theta_2 - \Phi_v), \quad (234)$$

for the LSB reception case.

These expected correlations again show the “white fringe”, with the sinc function envelopes corresponding to bandwidth $B = \Delta\omega/(2\pi)$ and the fringe patterns in cosine terms.

Amplitudes and arguments of the cosine terms are the theoretical expressions of the fringe amplitude and the fringe phase as introduced in Chapter 3 for the case of rectangular filters and a continuum spectrum source. The above formulae are often used as simplified models of actual correlator outputs for a continuum spectrum source. Note that the clock and atmospheric effects $(\omega_{LO} \pm \omega_I)(\tau_C + \tau_A)$ are directly contained in the fringe phase.

3.4 Limitations Imposed by Atmosphere and Clock Effects

Irregular phase fluctuations caused by the atmospheric delay and the clock offset impose two major difficulties in VLBI:

1. it is mostly impossible to directly use the important observable of the interferometry, the fringe phase, and
2. the coherence time, during which we can integrate observed data without significant loss of signal amplitude, is severely limited.

We will briefly describe these difficulties below.

3.4.1 Example of VLBI Fringe Phase

Figure 67 shows an example of time variation of the VLBI fringe phase which was obtained for a bright maser spot (single-line component) in the water maser source W3(OH) observed at 22 GHz band with Kitt Peak–Pie Town 417 km baseline of the VLBA (Very Long Baseline Array). Left panel shows a raw phase which is confined within -180° to $+180^\circ$. Right panel shows a connected phase. The large time variation of the fringe phase is mostly due to the effect of irregular atmospheric delay contained therein.

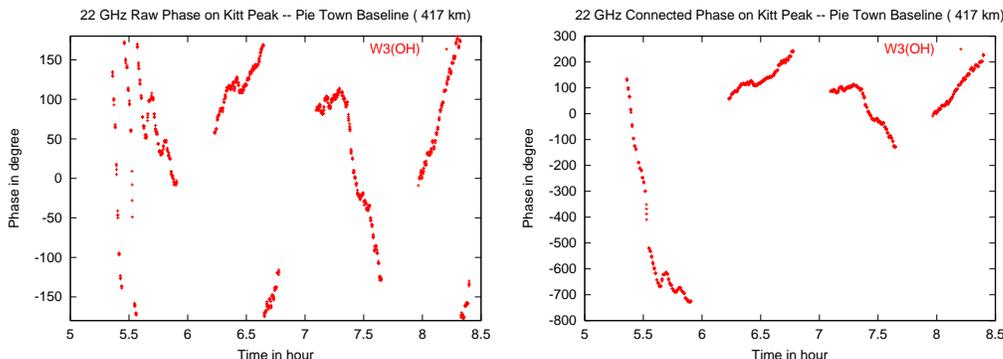


Figure 67: VLBI fringe phase of a maser spot in W3(OH) at 22 GHz obtained with Kitt Peak - Pie Town baseline. Left: raw phase confined within -180 deg to $+180$ deg. Right: connected phase. Horizontal axis shows time in hour.

This figure clearly illustrates how short the coherence time could be at 22 GHz during which the phase variation should not exceed 1 radian ($\sim 60^\circ$).

Also, it is evident that solving the $2\pi n$ cycle ambiguity is extremely difficult in view of the large irregular phase variation due to the atmosphere and clocks which easily exceeds 360° . As a matter of fact, it was totally impossible to connect the phase in the right panel of Figure 67 across the three gaps from 5:55 to 6:10, from 6:45 to 7:05, and from 7:35 to 7:55, resulting from intermittent observations of a calibrator source other than W3(OH).

Note that we can trace the phase values with a few second interval and connect them beyond $\pm 180^\circ$ in this Figure, solely owing to very high signal-to-noise ratio due to exceptionally strong flux density of the maser spot. For most of less bright sources, we can detect the phase values only after integration for much longer time intervals. Therefore, we usually cannot see the rapid phase variation in this detail.

Figure 67 is brought from unpublished interim results by Kazuya Hachisuka in his PhD thesis work in 2001 (Hachisuka, 2001).

3.4.2 Difficulty in Direct Use of VLBI Phase

It is quite difficult for current VLBI, especially at high frequencies, to use an important observable of the radio interferometry — the phase, i.e. the phase spectrum of the cross-correlation (i.e. phase of the cross-power spectrum), or the fringe phase (phase of the sinusoidal fringe pattern) in the correlator output.

This difficulty is often attributed to the irregular phase errors due to the clock-offset and atmospheric effects. Although such a statement is generally correct, the most important point here is that **we mostly cannot uniquely determine even a polluted phase**, i.e. a phase containing the clock-offset and atmospheric effects. What does this statement mean? This means that

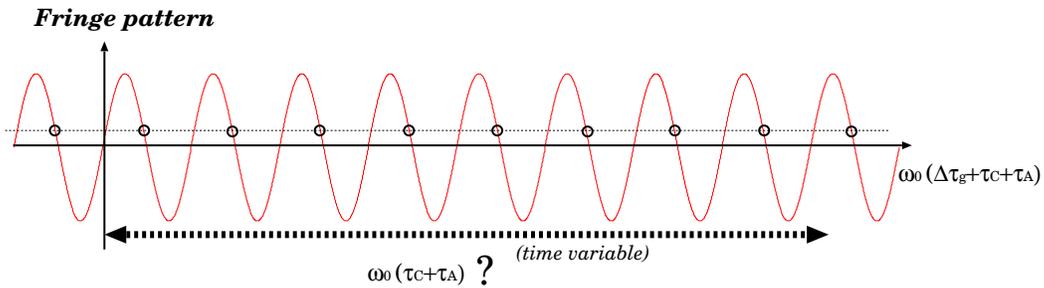


Figure 68: Unpredictable and irregularly variable clock-offset and atmospheric delay do not allow us to solve the $2\pi n$ “cycle ambiguity” in the VLBI phase. ω_0 is the center frequency of the RF band, $\Delta\tau_g$ is the residual geometric delay, τ_C is the clock offset, and τ_A is the atmospheric delay.

phase variations due to the unpredictable and irregularly variable clock offset and atmospheric delay, which easily exceed 360 degrees at high frequencies, do not permit us to solve the $2\pi n$ “cycle ambiguity”, inherent in the phase. Thus, we cannot tell if the observed phase is -200 degrees, or 160 degrees, or 520 degrees, \dots (Figure 68). Better to say, in this regard, that the phase data are completely corrupted or broken, rather than just contaminated. Consequently, it is almost impossible, at least at present, to directly use the VLBI phase in scientific analyses for structures of radio sources, or for precise positions of radio sources and VLBI stations.

Note that this statement must be treated with some care. We are talking here about the difficulty for **direct** use of the VLBI phase. There are powerful techniques such as “phase referencing”, “closure phase”, and others, which partially recover the phase information from VLBI data and allow us to use the information for the hyper-fine radio source imaging, and/or the high-precision astrometry and geodesy. We will discuss these topics later.

Also, we mainly talked about **high frequency** cases. In lower frequency observations with wavelengths not much shorter than scales of variations of the atmospheric excess path length (a few to several tens cm), it is sometimes possible to connect the fringe phase time series avoiding $2\pi n$ phase jumps. Therefore, there is a possibility for us to succeed to directly use the fringe phase as an observable at lower frequencies. Such a possibility is intensively studied in modern geodetic VLBI.

3.4.3 Limitation to Coherence Time

The irregular time variation of the phase imposed by the propagation delay due to the turbulent wet troposphere is rather rapid, and limits the interval of the time, during which we can coherently integrate the correlated VLBI data without significantly losing the signal amplitude. We called such a time interval the “coherence time”.

As we saw earlier, the coherence time τ_c is roughly estimated by the “accumulated phase error should not exceed one radian” condition:

$$\omega_0 \sigma_y \tau_c \cong 1,$$

where $\omega_0 = \omega_{LO} \pm \omega_I$ is the central frequency of the RF band (+ and – correspond to USB and LSB cases, respectively), and σ_y is the Allan standard deviation of the fractional frequency deviation $y = \delta\omega / \omega_0$.

While the stability of the hydrogen maser frequency standard is now better than 10^{-15} , the stability of the atmospheric delay in the troposphere is $\sim 10^{-13}$, in a time scale around 100 sec, as we saw before (Figure 66). If we assume a constant value of the Allan standard deviation $\sigma_y \sim 10^{-13}$ in a range of timescale from 10 sec to 1000 sec, the coherence times at various frequencies will be:

Observing Frequency (GHz)	2	8	10	22	43	86	129
Coherence Time (sec)	800	200	160	73	37	19	12

It is evident from this table that the atmospheric effect really severely limits the coherence time. Note that, at frequency higher than 100 GHz, the coherence time is as short as < 16 sec!

The short coherence time causes a serious limitation to the sensitivity in VLBI, which does not exist in single dish radio telescopes, or short-baseline and low-frequency connected-element interferometer arrays.

In fact, although the signal-to-noise ratio S/N is generally improved in proportion to the square root of the integration time τ_a , i.e. $S/N \propto \sqrt{\tau_a}$, in radio astronomical observations, we have to multiply here another factor η_{cP} to $\sqrt{\tau_a}$ in case of VLBI, i.e. $S/N \propto \eta_{cP} \sqrt{\tau_a}$, where η_{cP} is the coherence factor due to fringe-phase fluctuations introduced in Subsection 1.2.9. η_{cP} could rapidly drop to zero if we integrate correlated data beyond the coherence time. Therefore, it is meaningless in VLBI to simply increase the integration time, since, if we do so beyond the coherence time, we will lose the signal amplitude and will no longer be able to improve the S/N .

Therefore, in VLBI, we can directly detect only those radio sources which are strong enough to give sufficiently high signal-to-noise ratio **via integration within the coherence time**.

This limitation is especially severe in high-frequency VLBI observation, since the coherence time is the shorter, the higher the frequency, and there are other factors, such as the roughness of the antenna surface, or the receiver noise temperature, which degrade the sensitivity with increasing frequency. Thus, we can usually observe only exceptionally strong radio sources with mm-wave VLBI.

Again, we must treat this statement with some care. We are talking about **direct** detection of a source with **simple** integration. If we can well detect some bright component of a radio source within the coherence time, we can perform many hours of integration for detecting fainter components of the same source, by calibrating the fluctuating phases of the fainter components using similarly fluctuating phase of the detected bright component. This technique is effectively used in VLBI image synthesis techniques with sophisticated iterative processes for separating the complex visibility from the atmospheric and instrumental effects. The same sort of technique is used for detecting a weak source via long-time integration using an adjacent strong source, well detectable within the coherence time, as a phase calibrator in the “phase referencing”. Related topics will be discussed later.

4 Correlation Processing in VLBI

4.1 Features of Correlation Processing in VLBI

So far, we described the correlator in radio interferometry as the “multiplier + integrator”. However, correlation processing in VLBI has following specific features which are not limited to the simple multiplication and averaging.

- **Processing of played back (or transmitted via networks) records.**

Usually, VLBI data are correlated not in real time, but first recorded to magnetic tapes or disks, or transmitted via high-speed networks, and then processed by VLBI correlators. Therefore, precise “synchronization” of data streams from different stations is an important function to be built in VLBI correlators.

- **Delay tracking and fringe stopping in correlation processing.**

In ordinary VLBI systems, no delay tracking, nor fringe stopping, is carried out at station sites. They are performed in the correlation processing. Therefore, we can regard the VLBI correlator as the “**multiplier + integrator, with radio source tracking capability**”.

- **Hardware processing, and software processing (fringe fitting).**

For a number of reasons, which we will examine below, accuracy of theoretical prediction of a VLBI geometric delay is limited. Consequently, duration of time, over which the correlator hardware can integrate data without significantly losing coherence is limited to a certain short interval of time due to incompleteness of the predicted delay. This time interval is usually much shorter than the coherence time due to the atmospheric effects or instability of frequency standards as discussed above. We call this time interval of hardware integration, which is limited by the incompleteness of the theoretical model, the “**accumulation period**” (or “parameter period”) of the correlator.



Figure 69: 10-station FX type correlator (NAOJ, Mitaka, Japan).

In these circumstances, the correlation processing of VLBI data is usually divided into two steps.

In the first step, the hardware correlator processes segments of data one by one, where each segment spans an interval of time equal to the accumulation period. Though short, the accumulation period is still millions to hundreds of millions times longer than the sampling interval of the original data. Thus the data volume is greatly reduced through this hardware integration. Of course, a high processing speed, exceeding or equal to the recording speed, is usually required in the hardware correlator.



Figure 70: K4 based XF type correlator used in the CRL (now NICT) Key Stone Project (NICT, Koganei, Japan).

In the second step, which is often called the “**fringe search**” or “**fringe fitting**”, we further integrate outputs of the hardware correlator in computers for some duration of time up to the coherence time, searching correct parameters for the delay tracking and fringe stopping, specifically the group delay and the fringe frequency which we introduced in Chapter 3. We will discuss details of this integration later. Since the accumulation period is usually too short for detecting the white fringe with sufficient signal-to-noise ratio, it is mostly in the second step where the correlation peak of the white fringe is detected.

- **Powerful digital spectrometer.**

In order to search the correct delay-tracking and fringe-stopping parameters (group delay and fringe frequency) in the second step of the correlation processing, the VLBI correlators adopt either multi-lag (so called XF type correlators, see Figure 70), or multi-frequency-channel (so called FX type correlators, see Figure 69) designs.

These designs are equivalent to having many instrumental delays with different lengths around the theoretically predicted one, so that the white fringe may be found among them.

At the same time, the multi-lag or multi-frequency-channel designs make it possible to use the VLBI correlators as powerful digital spectrometers for VLBI spectral-line observations.

The XF correlator is an ordinary correlator, as we talked about earlier, which multiplies and averages received signals. The output from an XF correlator is the cross-correlation. The cross-power spectrum is calculated from this cross-correlation via Fourier transformation.

On the other hand, the FX correlator multiplies and averages **Fourier transformed received signals** obtained by special digital signal processing units at its front gate. The cross-power spectrum is directly obtained at the output of the FX correlator. The cross-correlation is calculated from the cross-power spectrum by inverse Fourier transformation.

The Fourier transformation theory guarantees that the XF and FX correlators give equivalent results. Owing to clever use of the Fast Fourier Transform (FFT) algorithm, the FX correlator becomes increasingly effective as the number of stations increases.

4.2 First Step—Hardware Processing

4.2.1 Accumulation Period

How do we determine the accumulation period?

We saw earlier that the expected correlation is given by equation (225) in the USB reception case, or by equation (226) in the LSB reception case, for a continuum source. So, let us consider

$$\Delta\phi = \omega_0(\Delta\tau_g + \tau_C + \tau_A), \quad (235)$$

as a main term in the residual phase of the fringe pattern in the expected correlation, where $\omega_0 = \omega_{LO} \pm \omega_I$ is the center frequency of the RF band

(+ and – correspond to USB and LSB cases, respectively), ω_{LO} is the local oscillator frequency, ω_I is the center frequency of the IF (video) band, $\Delta\tau_g = \tau_{g0} - \tau_i$ is the residual delay, $\tau_{g0} = \mathbf{D} \cdot \mathbf{s}_0 / c$ is the geometric delay, with baseline vector \mathbf{D} , source direction vector \mathbf{s}_0 , and the light velocity c , τ_i is the theoretically predicted “instrumental delay”, τ_C is the clock offset, and τ_A is the atmospheric delay.

If the delay model τ_i is not accurate enough, or τ_C and/or τ_A are large, time variation $\Delta\dot{\phi}$ of the residual phase $\Delta\phi$ could be large. Then the accumulation time t_{AP} is determined by the “ ≤ 1 radian” condition:

$$t_{AP} \leq \frac{1}{|\Delta\dot{\phi}|}. \quad (236)$$

Usually, the accumulation periods are determined by following factors.

- If frequency settings of the frequency standards in VLBI stations have a slight offset of $\sim 10^{-12}$, say, we have $\sim 10^{-12}$ sec/sec = 1 psec/sec error in delay rate, which is called “clock–rate error”.
- Position error of $\Delta s \sim 0.2$ arcsec, say, of a radio source causes delay–rate error of

$$\Delta\dot{\tau}_g = \frac{\dot{\mathbf{D}} \cdot \Delta \mathbf{s}}{c} = \frac{(\boldsymbol{\omega} \times \mathbf{D}) \cdot \Delta \mathbf{s}}{c} \sim 2.3 \text{ psec/sec},$$

for $D = 10000$ km baseline (here $\boldsymbol{\omega}$ is the Earth’s angular velocity). Note that actual errors in source coordinates could be much larger than this value in radio source catalogs based on single–dish observations. For example, errors in cataloged positions of some maser sources could exceed 10 arcsec!

- Baseline error of $\Delta D \sim 10$ m, say, causes delay–rate error of

$$\Delta\dot{\tau}_g = \frac{(\boldsymbol{\omega} \times \Delta \mathbf{D}) \cdot \mathbf{s}}{c} \sim 2.3 \text{ psec/sec}.$$

So, if we take $\Delta\dot{\phi} \sim \omega_0 \times 2.3$ psec/sec as a typical case of the plausible phase rate error, then we have:

Obs. Frequency (GHz)	8	22	43	86	129
Max. Accumulation period (sec)	8.6	3.1	1.5	0.8	0.5
Coherence time (sec) at $\sigma_y \cong 10^{-13}$	200	73	37	19	12

The AP's are really much shorter than the coherence times.

Usually, AP values are selectable in VLBI correlators. If you are uncertain about clock rate, source position, or station position, it is safer to select as short AP as possible. Otherwise, you could select the maximum AP as estimated above in order to reduce data volume of correlator output.

4.2.2 “Classical” XF-Type Correlator

Figure 71 shows basic components of a “classical” magnetic-tape-based MK-3 type XF correlator, for a single baseline composed of two stations.

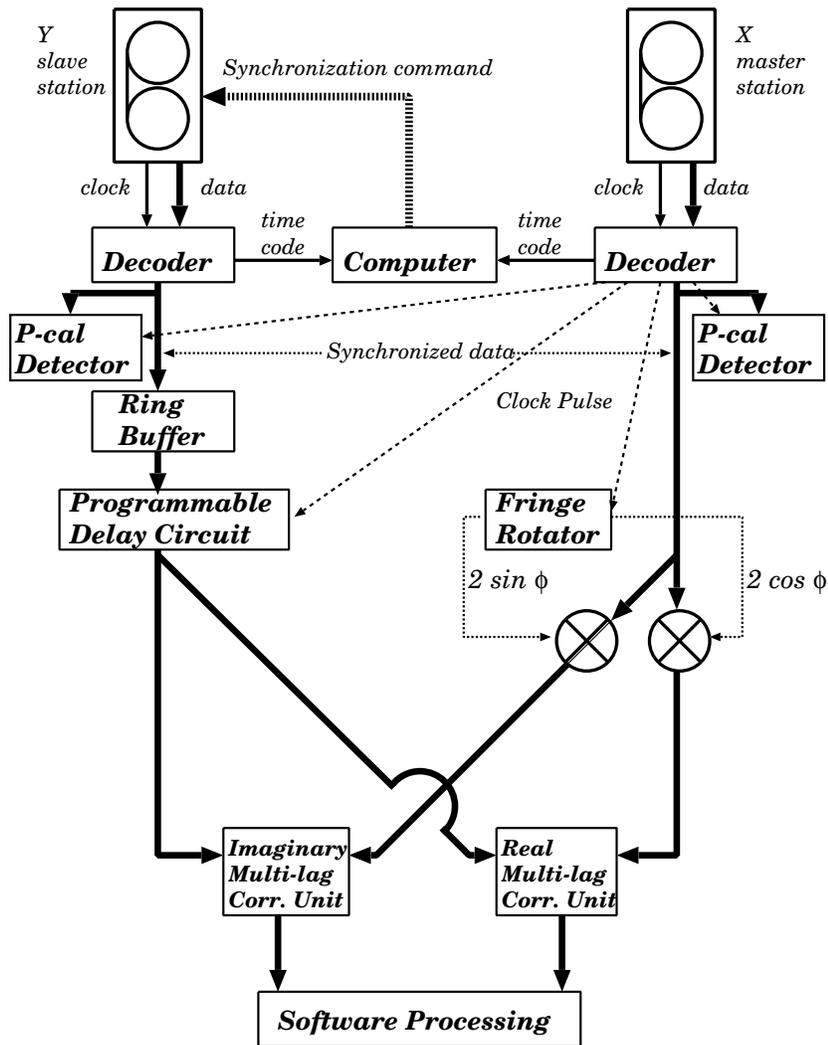


Figure 71: An example of the XF-type correlator.

This can be regarded as a single-baseline unit in a multi-baseline correlator. In fact, in case of N_s -station correlator, $N_s \times (N_s - 1) / 2$ (number of baselines) such units are used. Also, MK-3 VLBI systems usually adopt the multi-IF-channel design. If number of IF channels (or BB channels) is N_{ch} , all subunits after the decoders are in actuality composed of N_{ch} identical devices.

In the following, we will briefly examine functions of a hardware correlator, taking this classical XF-type correlator as an example, which is a little obsolete but convenient for illustrating basics of the VLBI correlator.

4.2.3 Data Synchronization

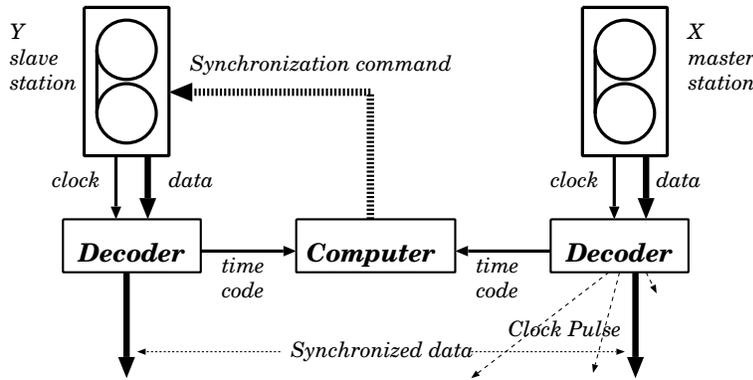


Figure 72: Data synchronization scheme.

In the correlation processing, data from two stations must be synchronized, in the sense that they must be read out with exactly the same speed. For this purpose, units called the “decoders” pick up time mark codes from the played back data and send them to a control computer.

The computer compares these codes, and then issues commands to the tape handling units, in order to control the tape speed of the “slave station”, usually denoted as the “Y-station”, until its time codes are read synchronously with those of the “master station”, denoted as the “X-station” (Figure 72). Of course, this “speed matching” function is not necessary for systems with hard-disk recorders.

We can always represent a theoretically predicted instrumental delay τ_i as a sum of time-constant and time-variable parts for a certain duration of time. The time-constant part of the instrumental delay during a scan of a source, which usually can be selected to be much larger than the time-variable part, is inserted at this stage by adjusting the time mark codes of the

simultaneously played back records to be offset by the time-constant delay. Thus, a part of the delay tracking is performed in the data synchronization stage.

At the same time, the decoder of the master X-station generates timing pulses (or clock pulses), which are adjusted to the data rate of the master record, and sends these timing pulses to all units of the correlator, in order to synchronize their operations to the data rate.

4.2.4 Detection of Phase Calibration Tones

As we described in Subsection 2.1.8, the antenna unit of the phase and delay calibrator system injects comb-tones into the received signal during VLBI observations, so that at least one comb-tone signal falls into each of the videoband (or baseband) channels. In the MK-3 system and its successors, local oscillator frequencies of the video converters (or BBC's) are selected in such a way that the sinusoidal comb-tone signal appears at a frequency of 10 kHz in each videoband.

The P-cal detector unit in the VLBI correlator picks up phases of the comb-tones in different videobands, in order to estimate, and then remove from the observed data, the instrumental phase offsets due mainly to the initial phases added by the local oscillators of the video converters.

Figure 73 schematically shows the principle of the P-cal detection.

Let us consider that the record V from one of the stations is composed of the received signal V_r and a sinusoidal comb-tone signal of frequency ω_{pcal} (at 10 kHz, say):

$$V = V_r + A \cos(\omega_{pcal} t - \phi_{VB}) = V_r + a \cos(\omega_{pcal} t) + b \sin(\omega_{pcal} t), \quad (237)$$

where A is the amplitude of the comb-tone, ϕ_{VB} is the instrumentally added phase at each videoband, and coefficients a and b are related to ϕ_{VB} by

$$a = A \cos(\phi_{VB}), \quad \text{and} \quad b = A \sin(\phi_{VB}).$$

The P-cal detector unit generates standard cosine and sine signals of frequency $\omega_{pcal} t$ according to the timing pulses provided from the master decoder, and then multiply them to each of bifurcated V record, and time-average (integrate) the products. The products of the V record with the standard cosine and sine signals are given by

$$\begin{aligned} V \cos(\omega_{pcal} t) &= V_r \cos(\omega_{pcal} t) + a \cos^2(\omega_{pcal} t) + b \sin(\omega_{pcal} t) \cos(\omega_{pcal} t) \\ &= V_r \cos(\omega_{pcal} t) + a \frac{1 + \cos(2\omega_{pcal} t)}{2} + b \frac{\sin(2\omega_{pcal} t)}{2}, \end{aligned}$$

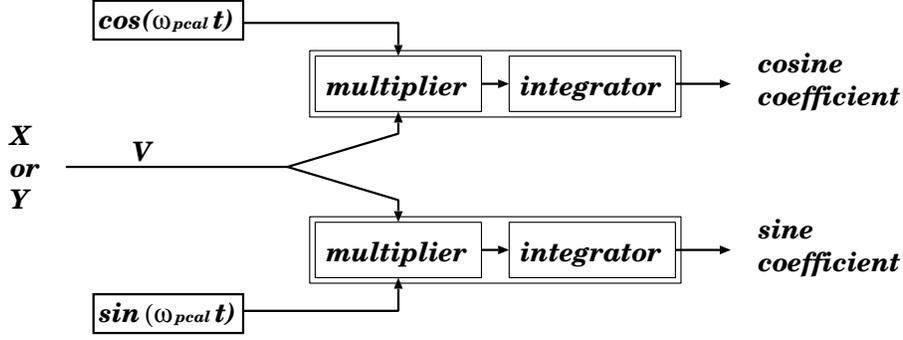


Figure 73: Schematic view of the P-cal detector unit.

$$\begin{aligned}
 V \sin(\omega_{pcal} t) &= V_r \sin(\omega_{pcal} t) + a \sin(\omega_{pcal} t) \cos(\omega_{pcal} t) + b \sin^2(\omega_{pcal} t) \\
 &= V_r \sin(\omega_{pcal} t) + a \frac{\sin(2\omega_{pcal} t)}{2} + b \frac{1 - \cos(2\omega_{pcal} t)}{2}.
 \end{aligned}$$

Since randomly varying and oscillating terms disappear after the time-averaging (integration), remaining integrated products are

$$\begin{aligned}
 \overline{V \cos(\omega_{pcal} t)} &= \frac{a}{2} = \frac{A}{2} \cos(\phi_{VB}), \\
 \overline{V \sin(\omega_{pcal} t)} &= \frac{b}{2} = \frac{A}{2} \sin(\phi_{VB}).
 \end{aligned}$$

Therefore, we can derive the instrumentally added phase ϕ_{VB} from these integrated products, and use it to calibrate the videoband phase.

In actual digital processing of MK-3 type data, the standard 10 kHz cosine and sine signals are often approximated by simpler patterns such as illustrated in Figure 46.

4.2.5 Ring Buffer Memory

Remaining part of the theoretically predicted instrumental delay, after the data synchronization, is inserted into one of the records (usually into that of the slave Y-station), to accomplish the delay tracking. A constant, during an accumulation period, part of the remaining delay, which is usually still much larger than the time-varying part within the accumulation period, is inserted by using a large ring buffer memory.

The data are written to and read from different addresses of the ring buffer, with a certain constant separation corresponding to the constant part of the predicted delay (Figure 74). Both the write and read addresses shift one by one (clock-wise in the Figure) at every arrival of a data sample.

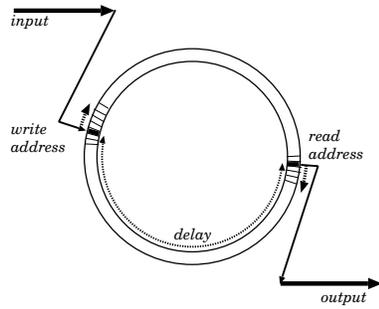


Figure 74: Ring buffer memory absorbs a delay constant during an accumulation period.

4.2.6 Delay Tracking by Programmable Delay Circuit

The time variable part of the instrumental delay is inserted into one of the records (usually the Y-station record) by a programmable delay circuit, which is a small ring buffer with programmable write address (Figure 75).

In order to follow the time variation of the theoretically predicted delay, which is usually well approximated by a linear variation within the short accumulation period, the write address is periodically shifted to the next one, after counting arrivals of a certain number of clock pulses.

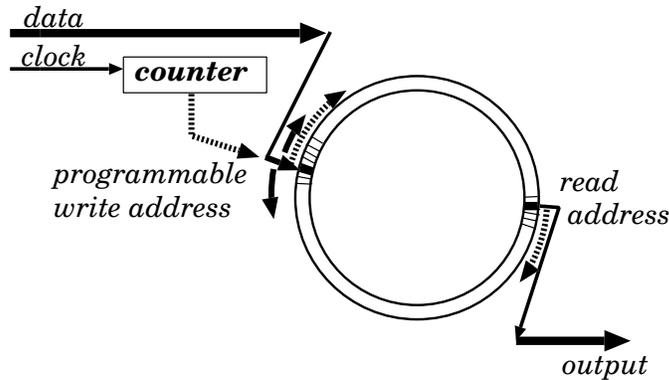


Figure 75: Programmable delay circuit for delay tracking in the videoband.

4.2.7 Address Shift Timing in Programmable Delay Circuit

In the digital circuit of the hardware correlator, the delay tracking can be carried out only discretely, through step-like jumps by an amount equal to the sampling interval t_s . Therefore, the shifts of the write address of the

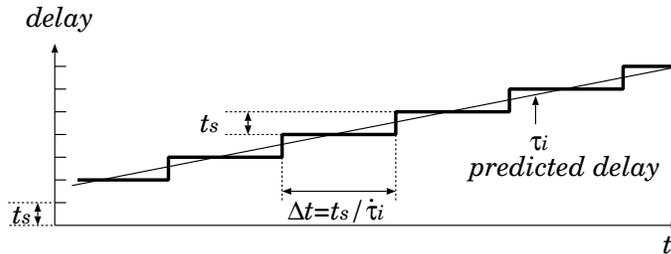


Figure 76: Step-like delay tracking in the programmable delay circuit.

programmable delay circuit occur with every passage of an interval of time Δt , during which the predicted delay τ_i changes by t_s , i.e. $\dot{\tau}_i \Delta t = t_s$. Thus, the address shift occurs once in

$$\frac{\Delta t}{t_s} = \frac{1}{\dot{\tau}_i} \text{ samples,}$$

(see Figure 76).

For the largest delay rate on the Earth's surface, which is $\dot{\tau}_i = 3.1 \mu\text{sec} / \text{sec}$, this shift interval is roughly 300,000 samples (in this regard, the vertical scale of Figure 76 is quite exaggerated). Such a step-like change of the instrumental delay does not exactly follow the smooth change of the actual delay, and hence causes some loss of coherence. In order to minimize this loss, the shift timing is adjusted to occur when the difference between the smooth change of the predicted delay and the step-like change reaches $t_s/2$.

4.2.8 “Fringe Rotator”: Fringe Stopping and Complex Correlation

A unit called the “fringe rotator” stops rapid time variation of the fringe phase (i.e. the phase of the fringe pattern in the expected correlation), by applying a theoretical correction based on the predicted delay.

For this purpose, the method of fringe stopping by multiplication of a sinusoidal function, which we saw in Chapter 3, is widely used in VLBI correlators. Using this method, we can also obtain the complex correlation, too, as we will see below (Figure 77).

The fringe rotator generates standard cosine and sine signals $2 \cos \Phi$ and $2 \sin \Phi$ with the argument

$$\Phi = \omega_{LO} \tau_i,$$

using theoretically predicted value of τ_i given by a control computer. Here ω_{LO} is the “total” local oscillator frequency which is the sum of LO frequen-

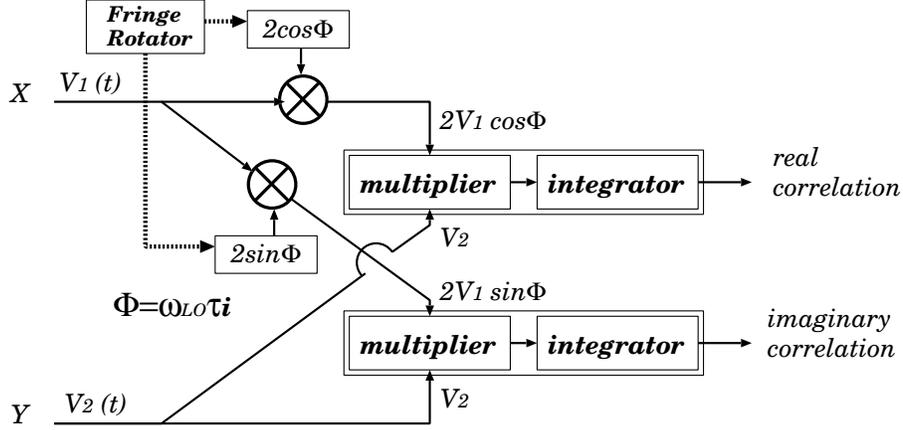


Figure 77: A circuit which realizes both fringe stopping and complex correlation.

cies of downconverters used for successive conversion of the RF signal to the videoband (or baseband) signal.

Each of the records, $V_1(t)$ and $V_2(t)$ from X- and Y-stations is bifurcated into two identical records. One of the bifurcated $V_1(t)$ records is multiplied by the cosine signal $2 \cos \Phi$, and the other bifurcated $V_1(t)$ record is multiplied by the sine signal $2 \sin \Phi$. The multiplication of cosine and sine signals realizes the 90-degrees phase shift which is essential in the complex correlation.

One of the $V_1(t)$ records from X-station, multiplied by the cosine signal, is led to a unit called the “real correlation unit”. This unit multiplies the cosine-multiplied $V_1(t)$ record with one of the bifurcated $V_2(t)$ records from Y-station and integrates the product. The other $V_1(t)$ record, multiplied by the sine signal, is led to the “imaginary correlation unit”, which multiplies this sine-multiplied record with the other one of the bifurcated $V_2(t)$ records and integrates the product.

In this stage, the delay tracking was already performed in the Y-station record $V_2(t)$, in terms of the ring buffers, but the fringe stopping is yet to be applied. Therefore, for a case of the USB reception of a continuum-spectrum source, expected correlation of the product of the $V_1(t)$ and $V_2(t)$ records is given by replacing $\omega_{LO} \Delta\tau_g$ in the cosine argument of equation (225) with $\omega_{LO} \tau_{g_0}$:

$$\begin{aligned}
 \langle V_1(t) V_2(t) \rangle &= R_{V_1 V_2}^{USB}(0) \\
 &= A_0 | \mathcal{V}^U | | \mathcal{B}_{12}^U | \cos[\omega_{LO} (\tau_{g_0} + \tau_C + \tau_A) + \omega_I (\Delta\tau_g + \tau_C + \tau_A) \\
 &\quad + \theta_1 - \theta_2 - \Phi_v^U - \Phi_B^U] \\
 &= \mathcal{A} \cos(\omega_{LO} \tau_{g_0} + \phi_r),
 \end{aligned} \tag{238}$$

where notations are the same with those in equation (225), except for newly introduced combined amplitude $\mathcal{A} = A_0 | \mathcal{V}^U | | \mathcal{B}_{12}^U |$ and combined phase $\phi_r = \omega_I \Delta\tau_g + (\omega_{LO} + \omega_I) (\tau_C + \tau_A) + \theta_1 - \theta_2 - \Phi_v^U - \Phi_B^U$ which is the sum of all phase terms in the cosine argument other than the $\omega_{LO} \tau_{g_0}$. For definiteness, we considered here the USB reception case for a continuum spectrum source only. However, the last line of equation (238) can be regarded as a general formula which holds for most of cases of our interest.

Thus, expected correlations between $2V_1(t) \cos \Phi$ and $V_2(t)$, and $2V_1(t) \sin \Phi$ and $V_2(t)$, are given by

$$\begin{aligned} 2\langle V_1(t) V_2(t) \rangle \cos \Phi &= 2 \mathcal{A} \cos(\omega_{LO} \tau_{g_0} + \phi_r) \cos(\omega_{LO} \tau_i) \\ &= \mathcal{A} \{ \cos[\omega_{LO} (\tau_{g_0} + \tau_i) + \phi_r] + \cos[\omega_{LO} (\tau_{g_0} - \tau_i) + \phi_r] \}, \\ 2\langle V_1(t) V_2(t) \rangle \sin \Phi &= 2 \mathcal{A} \cos(\omega_{LO} \tau_{g_0} + \phi_r) \sin(\omega_{LO} \tau_i) \\ &= \mathcal{A} \{ \sin[\omega_{LO} (\tau_{g_0} + \tau_i) + \phi_r] - \sin[\omega_{LO} (\tau_{g_0} - \tau_i) + \phi_r] \}. \end{aligned} \quad (239)$$

Then time averaging in integrators of the two correlation units almost nullifies the rapidly oscillating terms with $\omega_{LO} (\tau_{g_0} + \tau_i)$ argument, leaving terms with nearly stopped oscillations with $\omega_{LO} \Delta\tau_g$ argument only.

Therefore, expectations of the outputs of the real and imaginary correlation units will be

$$\begin{aligned} &\mathcal{A} \cos(\omega_{LO} \Delta\tau_g + \phi_r), \quad (\text{real correlation}), \\ &-\mathcal{A} \sin(\omega_{LO} \Delta\tau_g + \phi_r), \quad (\text{imaginary correlation}). \end{aligned} \quad (240)$$

Thus, the combination of the fringe rotator and the real and imaginary correlation units well executes the fringe stopping and complex correlation. We can really estimate the fringe amplitude \mathcal{A} and the residual fringe phase $\omega_{LO} \Delta\tau_g + \phi_r$ of the fringe pattern from the real and imaginary correlation results in equation (240) even when the fringe is completely stopped.

Note that the real correlator output in equation (240) is just the same with $R_{v_{F1}v_{F2}}^{USB}(0)$ in equation (225) which we obtained so far for the output of an ordinary (i.e. non-complex) correlator after DT and FS.

4.2.9 Multi-Lag Correlation Unit

In multi-lag correlators, which are widely used in VLBI, the real and imaginary correlation units actually consist of many multiplier & integrator units, as shown in Figure 78. Each of shift registers, connected sequentially in the V_2 record of the Y-station, say, inserts an additional delay, equal to the sampling interval t_s , to the record. After each insertion of the additional delay t_s , the delayed V_2 record is bifurcated to two identical records. One

of the bifurcated record is sent to one of the multiplier & integrator units, and multiplied with the $2V_1 \cos \Phi$ or $2V_1 \sin \Phi$ record of the X–station and integrated, while the other is sent to the next shift register. In this way, we obtain n correlated results (outputs of multipliers & integrators) with n different “instrumental delays” $\mathcal{R}(\tau_1), \dots, \mathcal{R}(\tau_l), \dots, \mathcal{R}(\tau_n)$, if the number of shift registers is $n - 1$. Here, $\tau_{l+1} = \tau_l + t_s$ for any l . The number of lags

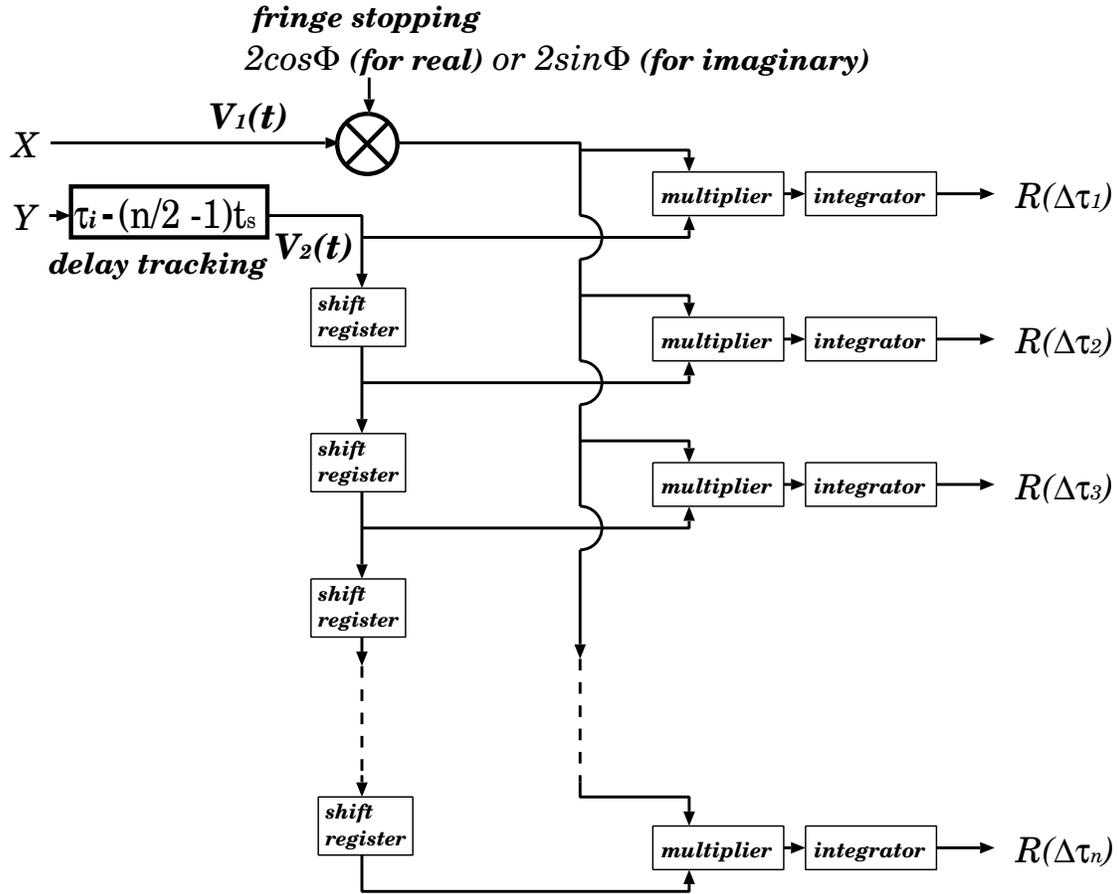


Figure 78: Multi-lag correlation unit (real or imaginary). Here $\Delta\tau_l = \tau_l - \tau_i$ is the time lag between the X and Y data as described in the text.

n is usually chosen to be an even number. Hereafter, we assume that n is even. The larger the n , the larger the chance for the actual delay to be found among these instrumental delays.

For convenience of ensuing software processing, the instrumental delay inserted by the data-synchronization and ring-buffer units is usually shifted from the theoretically predicted delay value τ_i by $(n/2 - 1)t_s$, so that the

theoretically predicted delay value τ_i is found just at the central lag of the multi-lag correlation unit. Thus, usually we have $\tau_1 = \tau_i - (n/2 - 1)t_s$ and $\tau_l = \tau_1 + (l - 1)t_s = \tau_i + (l - n/2)t_s$, and therefore $\tau_{n/2} = \tau_i$, as shown in Figure 78. Then, the correlation peak should be found at this central lag, if the theoretical prediction is correct.

In this circumstance, it is convenient to introduce a new time lag $\Delta\tau_l$ defined by an equation

$$\Delta\tau_l = \tau_l - \tau_i = \left(l - \frac{n}{2}\right)t_s = m t_s, \quad (241)$$

with a new lag number m :

$$m = l - \frac{n}{2}, \quad (242)$$

which takes 0 at the central lag $l = n/2$. We can conceive that the X and Y time series data are displaced one against another by $\Delta\tau_l = m t_s$, though their “alignment” when $m = 0$ should not be guaranteed as far as the theoretically predicted delay τ_i is incomplete. Hereafter, we will denote the correlated results as $\mathcal{R}(\Delta\tau_1), \dots, \mathcal{R}(\Delta\tau_l), \dots, \mathcal{R}(\Delta\tau_n)$.

4.2.10 Algorithm for 1-Bit Correlation

Correlation \mathcal{R} of 1-bit quantized data $\hat{x}[n]$ and $\hat{y}[n]$ within an accumulation period (AP) is calculated by a surprisingly simple logical circuitry. If N_t is the total number of samples within the AP, then the correlation (i.e. multiplication and integration) result is

$$\mathcal{R} = \frac{1}{N_t} \sum_{n=1}^{N_t} \hat{x}[n] \hat{y}[n].$$

In the 1-bit quantization case, $\hat{x}[n]$ and $\hat{y}[n]$ can take only +1 and -1, which correspond to 1 and 0, respectively, in the bit-assignment. Therefore, if we denote number of matched bits (i.e. 1 and 1, or 0 and 0) in the AP as N_m , then the number of unmatched bits (1 and 0 or 0 and 1) is $N_t - N_m$, and thus

$$\mathcal{R} = \frac{(+1) \times N_m + (-1) \times (N_t - N_m)}{N_t} = \frac{2N_m - N_t}{N_t}.$$

N_t and N_m are available from the following circuitry using the “exclusive NOR gate” shown in the right panel.

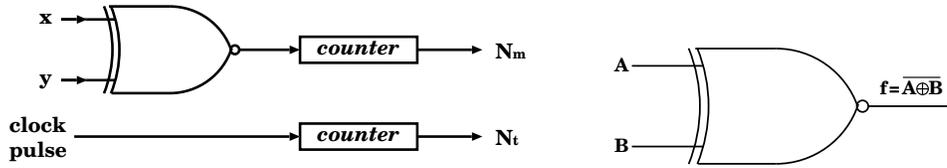


Figure 79: Logical circuitry for the 1-bit correlation (left), and exclusive NOR gate (right) with the following table of operation.

A	B	f
1	1	1
1	0	0
0	1	0
0	0	1

On the other hand, logical circuitries for correlation of 2-bit quantized data or floating-point data are more complicated and usually implemented by specially designed LSI's (Large Scale Integrated circuits).

4.2.11 Autocorrelation Unit

Many VLBI correlators include so-called "autocorrelation units" which calculate autocorrelations and then power spectra, or directly power spectra in case of FX-correlators, of records from individual observing stations. The records are bifurcated from those going to the cross-correlation units before delay-tracking and fringe stopping.

The power spectra thus obtained are quite useful in many respects, including calibration of bandwidth characteristics of receiving systems and rough estimation of sizes of spectral components in maser sources by comparing their power- and cross-power spectra (see Figure 28, as an example).

The autocorrelation units are usual multi-lag, or multi-frequency-channel, multiplier & integrator units which are not different from those used for cross-correlations. But they have neither delay-tracking and fringe-stopping units, nor imaginary correlation units, which are not necessary for autocorrelations.

Usually, N_s -station VLBI correlator includes N_s autocorrelation units.

4.2.12 Coherence Loss in Digital Correlation Processing

Besides the quantization loss, which we discussed before, the digital correlation processing may cause additional losses of the signal amplitude and

therefore the signal-to-noise ratio, which include

1. fringe rotation loss (coherence factor $\eta_R \geq 0.92$)
loss due to approximate representation of sinusoidal functions in the fringe rotator (ex. the 3-level approximation shown in Figure 80),

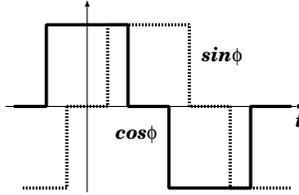


Figure 80: Sinusoidal patterns in the 3-level approximation used for fringe stopping in early models of hardware correlators.

2. fringe sideband rejection loss (coherence factor $\eta_S \geq 0.98$)
loss due to imperfection of digital filter which eliminates noise picked up by the fringe rotator from unnecessary sideband,
3. discrete delay step loss (coherence factor $\eta_D \geq 0.97$)
loss due to the step-wise tracking of continuously varying delay (see Figure 76).

Thus, in the worst case, the “coherence factor due to digital processing” $\eta_{c_D} = \eta_R \eta_S \eta_D$ could be as small as $\cong 0.87$.

However, modern correlators tend to adopt high-speed floating point computations, and large number of lags or frequency channels, which can make these factors close to 1.

When losses due to quantization, digital processing and fringe-phase fluctuations are taken into account, total coherence factor η_c should include a product $\eta_{c_Q} \eta_{c_D} \eta_{c_P}$, where η_{c_Q} and η_{c_P} are coherence factors due to quantization and fringe-phase fluctuations, respectively.

4.2.13 Output of Multi-Lag Complex Correlator

An n -lag complex correlation unit shown in Figure 78, which is a part of an XF-type complex correlator for a single baseline and a single videoband channel as illustrated in Figure 71, yields n complex outputs at each accumulation period (AP). Therefore, if the number of AP's in an observation is N , such a complex correlator generates a matrix of $n \times N$ complex outputs $\mathcal{R}_k(\Delta\tau_l)$,

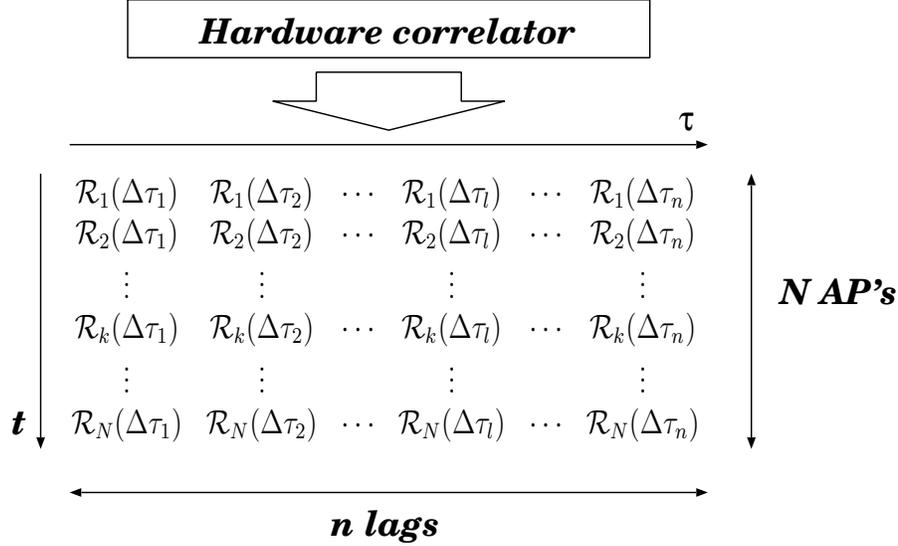


Figure 81: Outputs of a single-baseline single-channel unit of a multi-lag complex XF-type correlator.

with $l = 1, \dots, n$ and $k = 1, \dots, N$, during the observation, as schematically shown in Figure 81, where $\Delta\tau_l = \tau_l - \tau_i = m t_s$ with $m = l - n/2$ is the time lag between the X and Y time series data, as given in equation (241).

Hereafter, we call such an output $\mathcal{R}_k(\Delta\tau_l)$ of the complex correlator at l -th lag and k -th AP as “**complex cross-correlation**”.

This naming could be a little misleading since an actual output of each of the real and imaginary multiplier & integrator units shown in Figure 78 is a time average (actually a simple average in a digital correlator such as shown in equation (118)) of products of digitized samples of two voltage time series (“digital voltages” henceforth) during an AP. Thus, the output is not a statistical cross-correlation defined as an expectation (ensemble average) of product of two random variables. In practical correlation processings, correlator-handler softwares may divide the time average by a constant factor for later convenience, as we will see later. Anyway, the output $\mathcal{R}_k(\Delta\tau_l)$ is proportional to the time average of products of digital voltages, but not to their statistical cross-correlation. Hence, calling it as complex “cross-correlation” is a little inadequate, rigorously speaking. Nevertheless, the time average must be close to the statistical expectation, as far as the ergodicity is valid as we discussed in Chapter 3. We will use the term “complex cross-correlation” by convention in such an approximate sense.

Likewise, we call outputs of real and imaginary correlation units of a multi-lag complex correlator “**real cross-correlation**” and “**imaginary cross-correlation**”, respectively.

Thus, the single-baseline single-channel unit of a hardware correlator yields a matrix of $n \times N$ complex cross-correlations. If number of stations participating in an observation is N_s , we have $N_s(N_s - 1) / 2$ baselines. Furthermore, if number of videoband (or baseband) channels is N_{ch} , we need N_{ch} such correlator units for each baseline. Hence, a multi-station multi-channel hardware correlator yields $N_s(N_s - 1) N_{ch} / 2$ of such matrices. So, if $n = 512$, $N = 3600$, $N_s = 7$, $N_{ch} = 16$, and the real and imaginary cross-correlations are stored in 4-byte data each, say, then the data volume of an observation will be

$$4 \times 2 \times n \times N \times \frac{N_s \times (N_s - 1)}{2} \times N_{ch} \cong 5 \text{ Gbyte.}$$

Usually, we need a DVD disk or a digital cassette magnetic tape when we receive such correlated data from a correlation center.

4.2.14 Expected Output of Multiplier & Integrator

How can we theoretically describe an output $\mathcal{R}_k(\Delta\tau_l)$ of the hardware correlator at k -th AP and l -th lag which we called the complex cross-correlation?

The answer will be obtained by considering an expected form of the complex cross-correlation. The word “expected” again means here that we consider statistical expectation of the correlator output. Hereafter, we call this statistical expectation as “expected complex cross-correlation” and denote it by $\langle \mathcal{R}_k(\Delta\tau_l) \rangle$. So, how can we theoretically describe the expected complex cross-correlation? In order to answer to this question we will first examine expected output of a multiplier & integrator shown in Figure 78 as the final stage of the hardware correlator.

To be concrete, let us consider input digital voltages in k -th AP at l -th multiplier & intrgrator of either real- or imaginary correlation unit as shown in Figure 78. The digital voltages are displaced one against another by a time lag $\Delta\tau_l = m t_s$ inserted by shift registers in the n -lag correlation unit, where t_s is a sampling interval, and $m = l - n/2$ is the new lag number varying from $-n/2 + 1$ to $n/2$ introduced in equation (242). We will denote such digital voltages from X- and Y-stations as $\hat{x}[i]$ and $\hat{y}[i - m]$, respectively, where i is a sample number within an AP. Then, how does an output of the multiplier & integrator look like?

The output is a time average of products of the digital voltages $\hat{x}[i]$ and $\hat{y}[i - m]$ within an AP. Therefore, just like in the analog correlation case,

expectation of this output is equal to a statistical cross-correlation $R_{\hat{x}\hat{y}}[m]$ of the digital voltages as far as the assumption of the stationary random process is applicable, namely,

$$\left\langle \frac{1}{\mathcal{N}_t} \sum_{i=1}^{\mathcal{N}_t} \hat{x}[i] \hat{y}[i-m] \right\rangle = \frac{1}{\mathcal{N}_t} \sum_{i=1}^{\mathcal{N}_t} \langle \hat{x}[i] \hat{y}[i-m] \rangle = \frac{1}{\mathcal{N}_t} \sum_{i=1}^{\mathcal{N}_t} R_{\hat{x}\hat{y}}[m] = R_{\hat{x}\hat{y}}[m],$$

where \mathcal{N}_t is number of samples within an AP. Thus, the expected output of the multiplier & integrator at l -th lag (or m -th lag, in terms of the new lag number) is equal to a value at $j = m$ of a general cross-correlation $R_{\hat{x}\hat{y}}[j] = \langle \hat{x}[i] \hat{y}[i-j] \rangle$ of the digital voltages with an arbitrary integer argument j .

In order to get a theoretical expression of $R_{\hat{x}\hat{y}}[j]$, we first consider a cross-correlation $R_{\hat{x}'\hat{y}'}[j]$ of voltages before the DT(delay tracking) and FS(fringe stopping), i.e. “just digitized” voltages $\hat{x}'[i]$ and $\hat{y}'[i-j]$. In actual processing in a VLBI correlator, sampled and clipped voltages are brought from observing stations and then undergo the DT and FS before the multiplication & integration. So, we conceive that the “just digitized” $\hat{x}'[i]$ and $\hat{y}'[i-j]$ are digital voltages at input ports of the VLBI correlator.

We discussed in Subsections 1.1.20 and 1.1.24 that the cross-correlation $R_{\hat{x}'\hat{y}'}[j]$ of such “just digitized” voltages is proportional to a cross-correlation coefficient $r_{xy}(j t_s)$ of original analog voltages $x(t)$ and $y(t - j t_s)$ as given in equation (69):

$$r_{xy}(j t_s) = \frac{R_{xy}(j t_s)}{\sqrt{R_{xx}(0) R_{yy}(0)}}, \quad (243)$$

as far as $|r_{xy}(j t_s)| \ll 1$, where $R_{xy}(j t_s)$ is a cross-correlation at argument $\tau = j t_s$, while $R_{xx}(0)$ and $R_{yy}(0)$ are powers or dispersions (autocorrelations at zero arguments), of the analog voltages. The smallness condition $|r_{xy}(j t_s)| \ll 1$ is mostly satisfied in radio interferometry where system noise is usually much stronger than signal from a radio source. Specifically, equation (92):

$$R_{\hat{x}'\hat{y}'}[j] = \frac{2}{\pi} r_{xy}(j t_s),$$

for the 1-bit quantization case, as well as equation (106):

$$R_{\hat{x}'\hat{y}'}[j] = \frac{2}{\pi} [(\mathbf{n} - 1) E + 1]^2 r_{xy}(j t_s),$$

for the 2-bit quantization case, show the proportionality. For the 2-bit quantization case, we again assumed equal dispersions $R_{xx}(0) = R_{yy}(0) = \sigma^2$ of original analog voltages and used the same notations as those in Subsection 1.1.24. Note that the coefficient $(2/\pi) [(\mathbf{n} - 1) E + 1]^2$ is constant as far

as the analog dispersion σ^2 is constant during an integration time of the correlation processing.

Now the digital voltages undergo the DT and FS. Then, amplitude of their cross-correlation is further reduced by the coherence factor η_{c_D} due to digital processing as shown in Subsection 4.2.12, compared with a cross-correlation of ideal analog voltages for which the DT and FS are performed in a lossless way.

At the same time, the digital voltages enter to the complex correlation process through the bifurcations and the 90-degrees phase shift by means of the multiplications of cosine and sine signals provided by the fringe rotator as described in Subsection 4.2.8. This “transforming to complex” operation is sort of commutative with quantization and other digital operations. Therefore, we can conceive as if the bifurcations and the 90-degrees phase shift are performed in the analog stage, and then voltages for real- and imaginary correlation units are individually digitized and processed.

Thus, the expected real- and imaginary multiplier & integrator outputs at l -th lag with the new lag number m are now given by

$$\begin{aligned} R_{\hat{x}\hat{y}}^r[m] &= \eta_{c_D} \frac{2}{\pi} r_{xy}^r(m t_s), \\ R_{\hat{x}\hat{y}}^i[m] &= \eta_{c_D} \frac{2}{\pi} r_{xy}^i(m t_s), \end{aligned} \quad (244)$$

for the 1-bit quantization case, or

$$\begin{aligned} R_{\hat{x}\hat{y}}^r[m] &= \eta_{c_D} \frac{2}{\pi} [(\mathbf{n} - 1) E + 1]^2 r_{xy}^r(m t_s), \\ R_{\hat{x}\hat{y}}^i[m] &= \eta_{c_D} \frac{2}{\pi} [(\mathbf{n} - 1) E + 1]^2 r_{xy}^i(m t_s), \end{aligned} \quad (245)$$

for the 2-bit quantization case. Here $R_{\hat{x}\hat{y}}^r[m]$ and $R_{\hat{x}\hat{y}}^i[m]$ are cross-correlations of digital voltages in the real- and imaginary correlation units, while $r_{xy}^r(m t_s)$ and $r_{xy}^i(m t_s)$ are mutually 90-degrees-phase-shifted cross-correlation coefficients of ideal analog voltages after the DT and FS, which could be obtained if we had an analog complex correlator.

4.2.15 Cross-Correlation Coefficient as Expected Output

Early VLBI correlators, which processed mostly 1-bit quantized data, directly yielded the output of the multiplier & integrator as the correlator output. Therefore, expectation of this correlator output was the cross-correlation $R_{\hat{x}\hat{y}}[m]$ of digital voltages.

Nowadays, however, correlator–handler softwares in VLBI correlators tend to divide the output of the multiplier & integrator by a geometric mean of dispersions $R_{\hat{x}\hat{x}}[0]$ and $R_{\hat{y}\hat{y}}[0]$ of digital voltages, and yield the divided result

$$\frac{1}{\sqrt{R_{\hat{x}\hat{x}}[0] R_{\hat{y}\hat{y}}[0]}} \frac{1}{\mathcal{N}_t} \sum_{i=1}^{\mathcal{N}_t} \hat{x}[i] \hat{y}[i - m],$$

as a final correlator output. This is equivalent to consider that we correlate normalized digital voltages

$$\frac{\hat{x}[i]}{\sqrt{R_{\hat{x}\hat{x}}[0]}} \quad \text{and} \quad \frac{\hat{y}[i - m]}{\sqrt{R_{\hat{y}\hat{y}}[0]}}$$

instead of $\hat{x}[i]$ and $\hat{y}[i - m]$.

For 1–bit quantized data, the dispersions are identically equal to 1, as we discussed in Subsection 1.1.20, and hence no division operation is necessary in actuality.

For 2–bit quantized data, the dispersions can be either calculated based on equation (111) under the equal–dispersion assumption $R_{xx}(0) = R_{yy}(0) = \sigma^2$ for analog voltages:

$$R_{\hat{x}\hat{x}}[0] = R_{\hat{y}\hat{y}}[0] = \Phi + \mathbf{n}^2 (1 - \Phi), \quad (246)$$

where we used the same notations as those in Subsection 1.1.25, or can be derived from outputs of autocorrelation units mentioned in Subsection 4.2.11. Note that the dispersions as given by equations (111) and (112) are constant as far as the analog dispersion σ^2 is constant during an integration time of the correlation processing.

In such a case, expectation of the output of either real– or imaginary correlation unit is no longer equal to the cross–correlation, but equal to a cross–correlation coefficient $r_{\hat{x}\hat{y}}[m]$ of the digital voltages:

$$r_{\hat{x}\hat{y}}[m] = \frac{R_{\hat{x}\hat{y}}[m]}{\sqrt{R_{\hat{x}\hat{x}}[0] R_{\hat{y}\hat{y}}[0]}}, \quad (247)$$

as defined in equation (11).

Such a choice of the correlator output is meaningful since its expectation, the cross–correlation coefficient of the digital voltages, has following advantages:

1. from a point of view of symmetry, it is more appropriate to compare the cross–correlation coefficient $r_{xy}(m t_s)$ of the original analog voltages, which figures in the RHS of equations (244) and (245), with the cross–correlation coefficient $r_{\hat{x}\hat{y}}[m]$, rather than the cross–correlation $R_{\hat{x}\hat{y}}[m]$, of the digital voltages,

2. the ratio $r_{\hat{x}\hat{y}}[m] / r_{xy}(m t_s)$ of the cross-correlation coefficients of the digital and analog voltages is directly related to the signal-to-noise ratio of the correlator output as we saw in equation (126).

Henceforth, we will assume this case.

In view of equations (247), (244), (245), and (246), expectation of such a correlator output is given by

$$r_{\hat{x}\hat{y}}[m] = \eta_{c_D} \eta_{c_Q} r_{xy}(m t_s), \quad (248)$$

where η_{c_Q} is the coherence factor due to quantization given in equations (131) and (132), namely

$$\eta_{c_Q} = \frac{2}{\pi} \cong 0.64,$$

for the 1-bit quantization case, and

$$\eta_{c_Q} = \frac{2}{\pi} \frac{[(\mathbf{n} - 1) E + 1]^2}{[\Phi + \mathbf{n}^2 (1 - \Phi)]} \cong 0.88,$$

for the 2-bit quantization case. The numerical value 0.88 here corresponds to the optimum value discussed in Subsection 1.1.29.

We see two coherence factors in equation (248) for the expected correlator output. They are constants showing coherence losses occurring in quantization (η_{c_Q}) and digital processing (η_{c_D}). Generally speaking, there are other sources of coherence losses occurring through the integration in the correlation processing, such as the one due to fringe-phase fluctuations discussed in Subsection 1.2.9. However, coherence factors associated with these other effects are dependent on the integration time of the correlation processing, and usually very close to 1 for the short accumulation period (AP) of the hardware integration. Therefore, we take into account only η_{c_Q} and η_{c_D} in equation (248) for describing expected output of the hardware correlator within an AP.

4.2.16 Expected Complex Cross-Correlation

In the previous Subsection, we examined expected output of either real- or imaginary correlation unit shown in Figure 78. Thus, equation (248) is equally valid for cross-correlation coefficients of digital voltages both in the real- and imaginary correlation units. If we denote them as $r_{\hat{x}\hat{y}}^r[m]$ and $r_{\hat{x}\hat{y}}^i[m]$, respectively, we have

$$r_{\hat{x}\hat{y}}^r[m] = \eta_{c_D} \eta_{c_Q} r_{xy}^r(m t_s), \quad (249)$$

and

$$r_{\hat{x}\hat{y}}^i[m] = \eta_{c_D} \eta_{c_Q} r_{xy}^i(m t_s). \quad (250)$$

They are expected outputs of the real and imaginary correlation units. Therefore, the expected complex cross-correlation at k -th AP and l -th lag $\langle \mathcal{R}_k(\Delta\tau_l) \rangle$ is given by

$$\langle \mathcal{R}_k(\Delta\tau_l) \rangle = \tilde{r}_{\hat{x}\hat{y}}[m] = \eta_{c_D} \eta_{c_Q} \tilde{r}_{xy}(m t_s), \quad (251)$$

where

$$\tilde{r}_{\hat{x}\hat{y}}[m] = r_{\hat{x}\hat{y}}^r[m] + i r_{\hat{x}\hat{y}}^i[m], \quad (252)$$

and

$$\tilde{r}_{xy}(m t_s) = r_{xy}^r(m t_s) + i r_{xy}^i(m t_s), \quad (253)$$

are complex cross-correlation coefficients of digital and analog voltages, respectively. Here, we must be a little careful again about calling an output of a complex correlator as “complex cross-correlation”, since “expected complex cross-correlation” is actually a complex cross-correlation coefficient.

Now, we can rewrite equation (251) as

$$\langle \mathcal{R}_k(\Delta\tau_l) \rangle = \eta_{c_D} \eta_{c_Q} \frac{\tilde{R}_{xy}(m t_s)}{\sqrt{R_{xx}(0) R_{yy}(0)}},$$

with “cross-correlation of analog voltages”

$$\tilde{R}_{xy}(m t_s) = R_{xy}^r(m t_s) + i R_{xy}^i(m t_s), \quad (254)$$

where $R_{xy}^r(m t_s)$ and $R_{xy}^i(m t_s)$ are mutually 90-degrees-phase-shifted cross-correlations of ideal analog voltages after the DT and FS which could be obtained if we had an analog complex correlator, while $R_{xx}(0)$ and $R_{yy}(0)$ are powers or dispersions of analog voltages as before. Since η_{c_D} is essentially constant, while η_{c_Q} and $\sqrt{R_{xx}(0) R_{yy}(0)}$ usually do not appreciably vary during an integration time of the correlation processing, we will regard their combination $\eta_{c_D} \eta_{c_Q} / \sqrt{R_{xx}(0) R_{yy}(0)}$ as a constant coefficient of $\tilde{R}_{xy}(m t_s)$.

Then, the expected complex cross-correlation at k -th AP and l -th lag is given by

$$\langle \mathcal{R}_k(\Delta\tau_l) \rangle = \mathcal{C} \tilde{R}_{xy}(\Delta\tau_l), \quad (255)$$

where

$$\mathcal{C} = \frac{\eta_{c_D} \eta_{c_Q}}{\sqrt{R_{xx}(0) R_{yy}(0)}}, \quad (256)$$

is the constant coefficient, and $\Delta\tau_l = m t_s$ is the delay inserted by shift registers in the n -lag correlation unit as before. Equation (255) could be

interpreted as relating an expected output of an actual digital correlator to that of a virtual analog correlator which is functionally equivalent to the digital one but performing all operations in a lossless way.

Note that the cross-correlation of analog voltages $\tilde{R}_{xy}(\Delta\tau_l)$ in equation (255) contains only signal component coming from a radio source, and no system noise contribution. This is simply because noises from two stations do not correlate with each other, and do not correlate with signal, either. Therefore, the expected complex cross-correlation $\langle \mathcal{R}_k(\Delta\tau_l) \rangle$ in equation (255) also does not contain system noise contribution, except in the dispersions included in the constant coefficient \mathcal{C} . In this sense, we can regard the expected complex cross-correlation as a signal part of the output of the complex correlator $\mathcal{R}_k(\Delta\tau_l)$.

We are now ready to derive a theoretical form for the expected complex cross-correlation.

We saw in Subsections 3.3.1 and 3.3.2 that expected output of an ordinary analog correlator without multi-lag circuitry is given by a cross-correlation $R_{v_{F1}v_{F2}}^{USB}(0)$ of analog voltages v_{F1} and v_{F2} after perfect DT and FS at zero time delay $\tau = 0$. For definiteness, we assume here the USB reception case.

Therefore, the theoretical form of the real part of $\tilde{R}_{xy}(\Delta\tau_l)$, i.e. $R_{xy}^r(\Delta\tau_l)$ in equation (254), should be very close to that of $R_{v_{F1}v_{F2}}^{USB}(0)$. The only difference here is that we have an extra delay $\Delta\tau_l$ inserted into one of videoband voltages by the multi-lag correlator. We know that effect of such a delay is taken into account if we multiply $e^{i\omega\Delta\tau_l}$ to a cross-power spectrum of the voltages v_{F1} and v_{F2} , where ω is the videoband frequency. So, slightly modifying equation (223), we obtain

$$R_{xy}^r(\Delta\tau_l) = \frac{A_0}{4\pi} \Re \left\{ e^{-i[\omega_{LO}(\Delta\tau_g + \tau_C + \tau_A) + \theta_1 - \theta_2]} \times \int_0^\infty \mathcal{V}(\omega_{LO} + \omega) e^{-i\omega(\Delta\tau_g + \tau_C + \tau_A - \Delta\tau_l)} H_1(\omega) H_2^*(\omega) d\omega \right\}.$$

Thus, in view of equation (255), we now have

$$\Re \langle \mathcal{R}_k(\Delta\tau_l) \rangle = \mathcal{C} R_{xy}^r(\Delta\tau_l) = \frac{\mathcal{C} A_0}{4\pi} \Re \left\{ e^{-i[\omega_{LO}(\Delta\tau_g + \tau_C + \tau_A) + \theta_1 - \theta_2]} \times \int_0^\infty \mathcal{V}(\omega_{LO} + \omega) e^{-i\omega(\Delta\tau_g + \tau_C + \tau_A - \Delta\tau_l)} H_1(\omega) H_2^*(\omega) d\omega \right\} \quad (257)$$

as a general theoretical form for the expected real cross-correlation at k -th AP and l -th lag, where notations are the same with those in equation

(223), namely A_0 is the geometric mean of effective apertures of two antennas, $\Delta\tau_g = \tau_{g_0} - \tau_i$ is the residual geometric delay, $\tau_{g_0} = \mathbf{D} \cdot \mathbf{s}_0 / c$ is the geometric delay at a reference direction \mathbf{s}_0 in a radio source observed with an interferometer of baseline vector \mathbf{D} , τ_i is the theoretical prediction of the geometric delay used in the DT and FS (the “instrumental delay”), τ_C is the clock offset, τ_A is the atmospheric delay, $\theta_1 - \theta_2$ is the difference of the LO initial phases including the effect of the clock synchronization error between two stations as we defined in equation (216), $\mathcal{V}(\omega_{LO} + \omega)$ is the complex visibility, $H_1(\omega) H_2^*(\omega)$ is the “bandpass characteristics of combined IF filters” for the USB reception case, ω_{LO} is the local oscillator frequency, and ω is the IF (videoband) frequency. Values of all time-variable quantities are referred to the center time of k -th AP.

On the other hand, expected form of the imaginary cross-correlation is obtained by replacing $\omega_{LO} \Delta\tau_g$ in the above equation (257) by $\omega_{LO} \Delta\tau_g + \pi/2$. In fact, $\sin(\omega \tau_i) = \cos(\omega_{LO} \tau_i - \pi/2)$ is multiplied to the X-station record in the imaginary correlation unit, instead of $\cos(\omega_{LO} \tau_i)$ in the real correlation unit. Therefore, $\omega_{LO} \tau_{g_0} - \omega_{LO} \tau_i = \omega_{LO} \Delta\tau_g$ in the output of the real correlation unit should be replaced by $\omega_{LO} \tau_{g_0} - \omega_{LO} \tau_i + \pi/2 = \omega_{LO} \Delta\tau_g + \pi/2$ in the output of the imaginary correlation unit.

Since $\Re(e^{-i\frac{\pi}{2}} z) = \Re(-i z) = \Im z$ for any complex number z , we obtain

$$\begin{aligned} \Im\langle\mathcal{R}_k(\Delta\tau_l)\rangle &= \frac{\mathcal{C} A_0}{4\pi} \Im\left\{e^{-i[\omega_{LO}(\Delta\tau_g+\tau_C+\tau_A)+\theta_1-\theta_2]}\right. \\ &\quad \left.\times \int_0^\infty \mathcal{V}(\omega_{LO} + \omega) e^{-i\omega(\Delta\tau_g+\tau_C+\tau_A-\Delta\tau_l)} H_1(\omega) H_2^*(\omega) d\omega\right\}, \end{aligned} \quad (258)$$

for the expected imaginary cross-correlation,

Finally, the theoretical form for the expected complex cross-correlation $\langle\mathcal{R}_k(\Delta\tau_l)\rangle = \Re\langle\mathcal{R}_k(\Delta\tau_l)\rangle + i \Im\langle\mathcal{R}_k(\Delta\tau_l)\rangle$ is given by

$$\begin{aligned} \langle\mathcal{R}_k(\Delta\tau_l)\rangle &= \mathcal{C} \tilde{R}_{xy}(\Delta\tau_l) = \frac{\mathcal{C} A_0}{4\pi} \left\{e^{-i[\omega_{LO}(\Delta\tau_g+\tau_C+\tau_A)+\theta_1-\theta_2]}\right. \\ &\quad \left.\times \int_0^\infty \mathcal{V}(\omega_{LO} + \omega) e^{-i\omega(\Delta\tau_g+\tau_C+\tau_A-\Delta\tau_l)} H_1(\omega) H_2^*(\omega) d\omega\right\}. \end{aligned} \quad (259)$$

4.2.17 Expected Complex Cross-Correlation for White Spectrum Source

For a white spectrum source, i.e. a continuum source with constant visibility, equation (259) is reduced to

$$\langle\mathcal{R}_k(\Delta\tau_l)\rangle = \mathcal{C} A_0 |\mathcal{V}| |\mathcal{B}_{12}| e^{-i[(\omega_{LO}+\omega_I)(\Delta\tau_g+\tau_C+\tau_A)-\omega_I \Delta\tau_l+\theta_1-\theta_2-\Phi_v-\Phi_B]},$$

(260)

where ω_I is the central frequency of the IF band, $|\mathcal{V}|$ and Φ_v , and $|\mathcal{B}_{12}|$ and Φ_B are amplitudes and phases of the complex visibility \mathcal{V} and the bandwidth pattern \mathcal{B}_{12} with bandwidth $\Delta\omega = 2\pi B$, respectively. The complex visibility and the bandwidth pattern are given by equations (227), (229) and (231) with a slight modification due to the additional delay $\Delta\tau_l$:

$$\mathcal{V}(\omega_{LO} + \omega_I) = |\mathcal{V}| e^{i\Phi_v}, \quad \mathcal{B}_{12}(B, T) = |\mathcal{B}_{12}| e^{i\Phi_B},$$

$$\text{with } \mathcal{B}_{12}(B, T) = \frac{1}{4\pi} \int_{-\frac{\Delta\omega}{2}}^{\frac{\Delta\omega}{2}} e^{-i\omega'T} H_1(\omega_I + \omega') H_2^*(\omega_I + \omega') d\omega',$$

where we introduced a notation

$$\mathcal{T} = \Delta\tau_g + \tau_C + \tau_A - \Delta\tau_l.$$

In the simplest case of the IF band made by a rectangular filter (Figure 82), we again obtain the white fringe with the sinc-function envelope:

$$\langle \mathcal{R}_k(\Delta\tau_l) \rangle = \frac{\mathcal{C} A_0 G B |\mathcal{V}|}{2} \frac{\sin(\pi B \mathcal{T})}{\pi B \mathcal{T}} e^{-i[(\omega_{LO} + \omega_I)(\Delta\tau_g + \tau_C + \tau_A) - \omega_I \Delta\tau_l + \theta_1 - \theta_2 - \Phi_v]}. \quad (261)$$

Therefore, the expected real and imaginary cross-correlations are given by

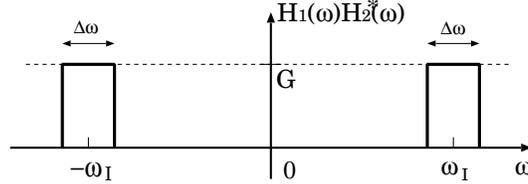


Figure 82: IF band made by a rectangular filter.

$$\Re\langle \mathcal{R}_k(\Delta\tau_l) \rangle = \frac{\mathcal{C} A_0 G B |\mathcal{V}|}{2} \frac{\sin(\pi B \mathcal{T})}{\pi B \mathcal{T}} \times \cos((\omega_{LO} + \omega_I)(\Delta\tau_g + \tau_C + \tau_A) - \omega_I \Delta\tau_l + \theta_1 - \theta_2 - \Phi_v), \quad (262)$$

$$\Im\langle \mathcal{R}_k(\Delta\tau_l) \rangle = -\frac{\mathcal{C} A_0 G B |\mathcal{V}|}{2} \frac{\sin(\pi B \mathcal{T})}{\pi B \mathcal{T}} \times \sin((\omega_{LO} + \omega_I)(\Delta\tau_g + \tau_C + \tau_A) - \omega_I \Delta\tau_l + \theta_1 - \theta_2 - \Phi_v). \quad (263)$$

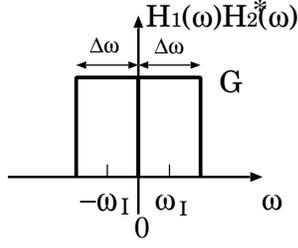


Figure 83: Rectangular videoband.

4.2.18 Rectangular Videoband Case

If the IF band is a videoband (baseband), which is usually the case in VLBI, and rectangular (Figure 83), we have $\omega_I = \Delta\omega/2 = \pi B$. Also, we have

$$t_s = \frac{1}{2B}, \quad \text{and therefore} \quad \omega_I = \pi B = \frac{\pi}{2t_s},$$

when the sampling is done with the Nyquist rate. Hence, if we introduce new notations:

$$\Delta\tau_D = \Delta\tau_g + \tau_C + \tau_A, \quad (\text{“residual delay”}), \quad (264)$$

$$\Phi_F = \omega_{LO}\Delta\tau_D + \theta_1 - \theta_2 - \Phi_v, \quad (\text{“residual phase”}), \quad (265)$$

then $\mathcal{T} = \Delta\tau_D - \Delta\tau_l$, and equations (262) and (263) become

$$\Re\langle\mathcal{R}_k(\Delta\tau_l)\rangle = \frac{\mathcal{C} A_0 G B |\mathcal{V}|}{2} \frac{\sin\left(\frac{\pi}{2} \frac{\Delta\tau_D - \Delta\tau_l}{t_s}\right)}{\frac{\pi}{2} \frac{\Delta\tau_D - \Delta\tau_l}{t_s}} \cos\left(\Phi_F + \frac{\pi}{2} \frac{\Delta\tau_D - \Delta\tau_l}{t_s}\right), \quad (266)$$

$$\Im\langle\mathcal{R}_k(\Delta\tau_l)\rangle = -\frac{\mathcal{C} A_0 G B |\mathcal{V}|}{2} \frac{\sin\left(\frac{\pi}{2} \frac{\Delta\tau_D - \Delta\tau_l}{t_s}\right)}{\frac{\pi}{2} \frac{\Delta\tau_D - \Delta\tau_l}{t_s}} \sin\left(\Phi_F + \frac{\pi}{2} \frac{\Delta\tau_D - \Delta\tau_l}{t_s}\right), \quad (267)$$

in the rectangular videoband case.

4.2.19 Behavior of Correlator Outputs

Let us examine behavior of the expected complex cross-correlation with respect to lag $\Delta\tau_l$ and time, using the simple model given by equations (266) and (267) for a white spectrum source with the rectangular videoband.

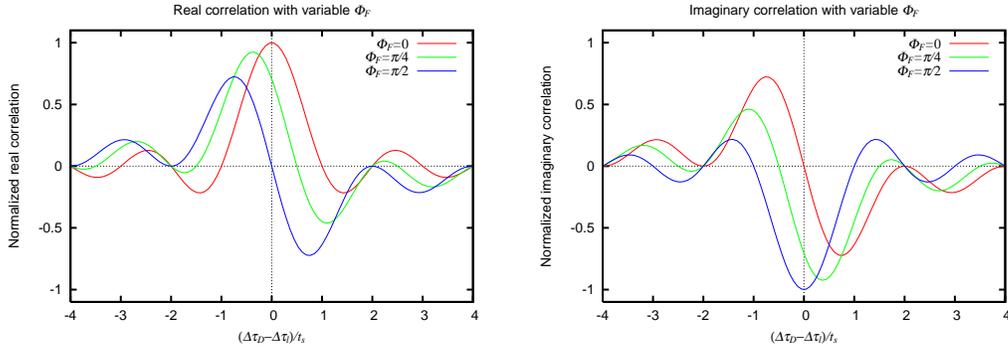


Figure 84: Normalized expected real (left) and imaginary (right) correlations as functions of $(\Delta\tau_g - \Delta\tau_l)/t_s$ with various Φ_F .

Figure 84 shows expected real and imaginary cross-correlations normalized by $\mathcal{C} A_0 G B |\mathcal{V}| / 2$ as functions of $(\Delta\tau_g - \Delta\tau_l)/t_s$, for $\Phi_F = 0, \pi/4$ and $\pi/2$. When the fringe stopping by the hardware correlator is not complete, the residual phase Φ_F varies in time. Therefore, the Figure can be interpreted as showing time variation of patterns of the expected correlations with varying Φ_F due to the imperfect fringe stopping.

In order to see the time variation more clearly, we use the new lag number $m = l - n/2$ introduced in equation (242), where n is the number of lags in a correlator. Then, we have $\Delta\tau_l = m t_s$, and the expected real and imaginary

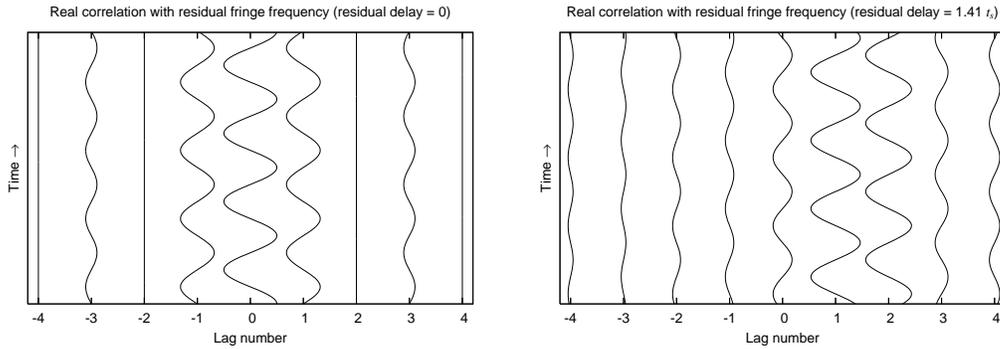


Figure 85: Time variations of expected real cross-correlation at various lag numbers due to the incomplete fringe stopping, when the residual delay is zero (left) and non-zero (right).

cross-correlations in equations (266) and (267) are reduced to

$$\begin{aligned}\Re\langle\mathcal{R}_k(\Delta\tau_l)\rangle &= \frac{\mathcal{C} A_0GB |\mathcal{V}|}{2} \frac{\sin\left(\frac{\pi}{2} \frac{\Delta\tau_D}{t_s} - \frac{\pi}{2} m\right)}{\frac{\pi}{2} \frac{\Delta\tau_D}{t_s} - \frac{\pi}{2} m} \cos\left(\Phi_F + \frac{\pi}{2} \frac{\Delta\tau_D}{t_s} - \frac{\pi}{2} m\right), \\ \Im\langle\mathcal{R}_k(\Delta\tau_l)\rangle &= -\frac{\mathcal{C} A_0GB |\mathcal{V}|}{2} \frac{\sin\left(\frac{\pi}{2} \frac{\Delta\tau_D}{t_s} - \frac{\pi}{2} m\right)}{\frac{\pi}{2} \frac{\Delta\tau_D}{t_s} - \frac{\pi}{2} m} \sin\left(\Phi_F + \frac{\pi}{2} \frac{\Delta\tau_D}{t_s} - \frac{\pi}{2} m\right).\end{aligned}\tag{268}$$

Figure 85 shows time variations of the expected real cross-correlation at lag numbers $m = -4, \dots, 0, \dots, 4$ due to the incomplete fringe stopping and resultant time-variable residual phase Φ_F . Left panel shows the case when the residual delay is zero, i.e. $\Delta\tau_D = 0$, while right panel shows the case with non-zero residual delay ($\Delta\tau_D \neq 0$).

Note that the phase of the cosine function changes by $\pi/2$ when we increment the lag number m by 1, as equation (268) requires. This is a consequence of our choice of the videoband and the Nyquist sampling.

Figure 85 illustrates that the expected correlator output at each lag oscillates in time (or just zero at certain lag numbers), whenever the fringe stopping with the hardware correlator is incomplete. This means that expected correlator outputs at successive AP (accumulation period) numbers also show such oscillations, since the hardware correlator periodically yields correlated results at every AP.

It is clearly meaningless to directly integrate (time-average) the correlator outputs over many APs, unless we stop such oscillations.

Figure 86 shows the oscillatory behavior of the expected real (left) and imaginary (right) cross-correlations over a lag-number vs AP-number plane. Top panel shows the case when the residual group delay is zero ($\Delta\tau_D = 0$), while bottom panel shows the non-zero case ($\Delta\tau_D \neq 0$).

We saw in Subsection 4.2.13 that an n -lag hardware correlator generates a matrix of $n \times N$ complex cross-correlations for an observation consisting of N AP's. Therefore, it is just straightforward to make a bird's eye plot of real or imaginary cross-correlations over the lag-number vs AP-number plane, generally speaking.

For a radio source which is strong enough to give a high signal-to-noise ratio even with a very short integration time during an AP, we can see systematic oscillatory behavior of actual outputs of real and imaginary correlators over the lag-number vs AP-number plane similar to the one shown in Figure 86. Later, we will see examples of such correlator outputs for strong

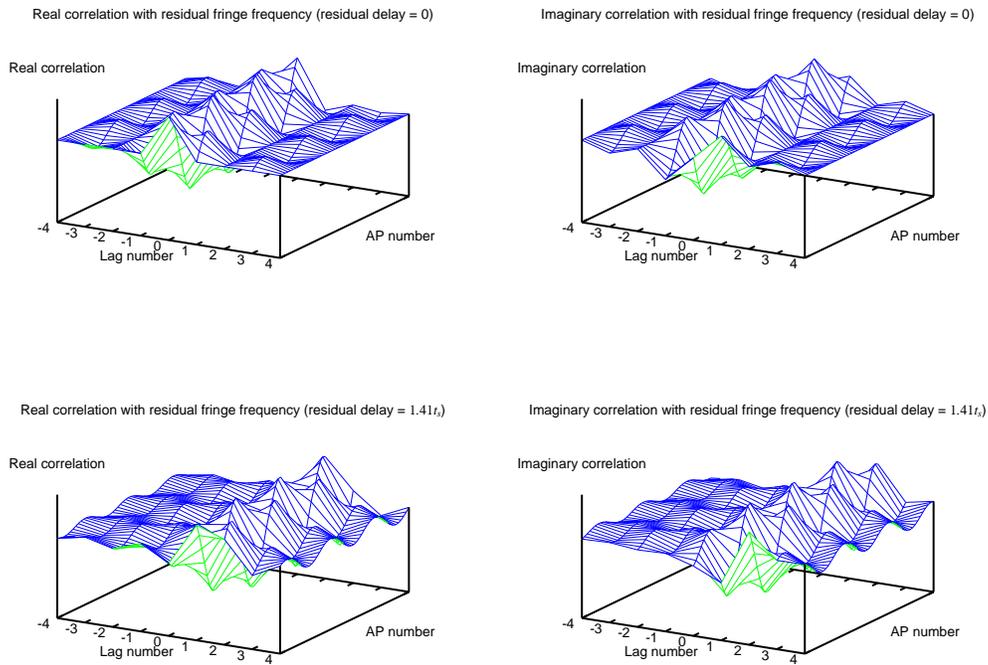


Figure 86: Expected real (left) and imaginary (right) cross-correlations over the lag-number vs AP-number plane when the fringe stopping with the hardware correlator is incomplete. Top and bottom panels correspond to cases when the residual group delay is zero (top) and non-zero (bottom).

sources in Subsection 4.3.18. However, for most of sources, we can hardly see any systematic behavior of the correlator outputs in the lag–number vs AP–number plane due to overwhelming noise which easily masks weak signal from the radio source.

4.2.20 PC–Based Correlators

Nowadays, a number of institutions in the world successfully developed and use PC–based correlators for processing VLBI data. This is a technology to realize functions of a hardware correlator by means of commercially available general purpose high–speed PC’s. All elements of VLBI correlation processing including the DT and FS are performed in ordinary PC’s, without any specially made digital circuitry, just following software commands. In this sense, the PC–based correlators can be regarded as “software correlators”. The hard disk recording and/or network transfer of observed data

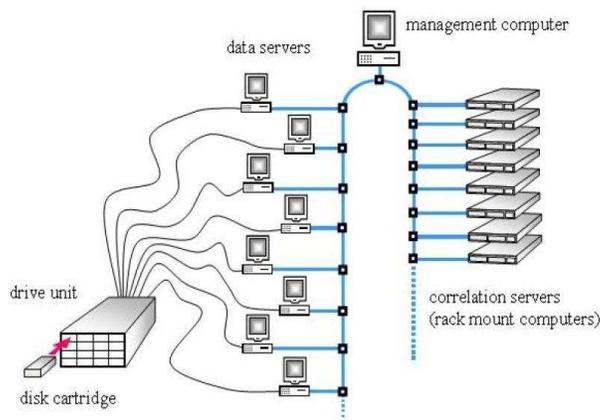


Figure 87: Schematic view of PC–based K5/VSSP software correlator developed by the NICT and used in the GSI (from Machida et al., 2006, IVS GM).

are highly suited to this kind of processing. As an example, Figure 87 shows a schematic diagram of K5/VSSP software correlator developed by the NICT (National Institute of Information and Communications Technology) at Kashima, Japan, and being used in the GSI (Geodetic Survey Institute) at Tsukuba, Japan (from Machida et al., 2006, IVS GM).

The PC–based correlators are quite inexpensive, flexible and easy to expand or update. Also, the PC–based correlators can be easily combined with high–speed transfer of observed data via information networks. Therefore, they offer very cost–effective and user–friendly data processing in international VLBI observations.

CRL 鹿島 VLBI ニュース

第94号 平成15年8月21日
 独立行政法人高度社会科学研究施設 宇宙通信研究センター
 〒901-8542 鹿島 鹿島 4-1-1
 TEL 0239-84-7137 FAX 0239-84-7159
 WWW URL http://www.crl.go.jp/ku/rdcastr/index_e_3.html

汎用PCを用いた2Gbps-VLBIシステム完成 国際基線での実証実験に成功！

日本とフィンランドで独立に開発されたVSIデータを直接PCで記録する装置を用いて、2003年6月13日2Gbpsという世界最速のVLBI実験に成功しました。観測にはCRLで開発された1Gbps-Zot A/Dコンブリが各観測所に配備され、日本側はCRLが開発した60MHz/66bitのPCバスを用いたデータ取得システムによりVLBIデータをPCへ転送し、FPGA技術を用いたハードディスクへ連続的に記録しました。フィンランド側で記録されたデータは鹿島局へ汎用ネットワークを利用して伝送され、PC上で動作する高速相関ソフトウェアで迅速に処理されました。この2Gbpsという速度は最速のハードウェアVLBIシステムと同等であり、今後のVLBIシステムの主流がPC主体へと変化する大きなマイルストーンを築きました。(M.K.記)



写真1: VSIデータをPCへ取得するPCIボード
Linux上で動作し、66MHz/66bitのPCバスを利用することで最大2Gbpsの連続データ取得が可能。将来的にはVSIデータ出力にも対応する。



写真2: 開発に協力してくれたメーカーの方々と、今回開発に用いられた2台のPC。

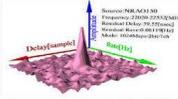


表: 相関処理時に量子化ビット数を変化させた時のSNR変化。天体は NRAO150 で積分時間は16秒とした。ビット観測によりSNRが約4倍程度向上した。

	Yohkoh-1M	Yohkoh-2M
Kashima-1M	22.8	27.4
Kashima-2M	27.2	31.8

図1: 2Gbps-VLBIシステムで最初に観測された NRAO150 の二次元フリンジャー手続。

Figure 88: “Kashima VLBI News” issued in 2003 by the Radio Astronomy Application Group at Kashima branch of the CRL (Communications Research Laboratory, now the NICT: National Institute of Information and Communications Technology), reporting successful correlation processing of 2 Gbps international VLBI data, transferred via an information network, with a PC-based FX correlator.

PC-based correlators are now gaining very high performance even comparable with the newest hardware correlators. An example is a PC-based FX correlator developed by the CRL (Communications Research Laboratory, now the NICT) shown in Figure 88. In 2003, this FX software correlator successfully processed network transferred data of an international VLBI observation with an ultra high speed of 2 Gbps. Now the correlator is upgraded to 4 station Gbps correlator for the VERA (VLBI Exploration of Radio Astrometry) array.

4.2.21 New 16-Station FX-Type Correlator in Seoul

For processing wide-band data of a large VLBI array consisting of many stations, hardware correlators with specially designed circuitries are still more advantageous than PC-based correlators. Table 10 shows main specifications of a newly developed 16-station FX-type correlator in Seoul. Figure 89 shows a block diagram of the correlator. The Seoul Correlator will be used for processing data of East Asian VLBI Network composed of VLBI arrays in China, Japan and Korea, including a future spaceborne VLBI antenna VSOP-2.

Maximum number of antennas	16
Number of input ports/antenna	4

Maximum data rates/antenna	Total 8 Gbps
Input interface specification	2 Gbps VSI-H
Input data Number of bits/sample	2 bits/sample
Sampling rate	1024 MHz
Total bandwidth	512 MHz
Substreams	10 modes of channelization
Maximum delay	± 0.11 sec (VSOP-2)
Maximum delay rate	25 μ sec/sec (VSOP-2)
Spectral resolution	4~128 kHz (≥ 0.05 km/sec at 22 GHz)
Integration time	25.6 msec ~ 1 sec
Maximum number of outputs	480 cross- and 64 auto-correlations
Maximum rate of outputs	About 1.4 GB/sec

Table 10: Main specifications of 16-station FX-type correlator in Seoul.

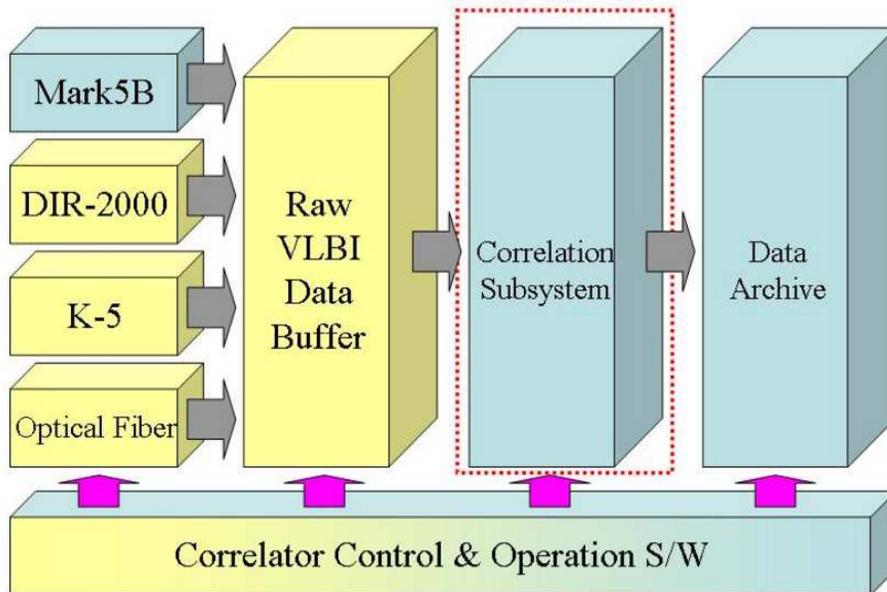


Figure 89: Block diagram of a new 16-station FX-type hardware correlator developed in Seoul for East Asian VLBI Network.

4.3 Second—Step, Software Processing (Fringe Fitting)

The purpose of the second-step software processing, which is often called the “fringe search” or “fringe fitting”, is to detect the white fringe, and determine the correlation peak with sufficiently high signal-to-noise ratio, by integrating the correlator outputs further in time, i.e. over many accumulation periods (AP’s), flexibly searching correct parameters for the delay tracking (DT) and fringe stopping (FS). These parameters are the group delay τ_G and the fringe frequency F_r which we introduced in Chapter 3. Since we already applied theoretically predicted instrumental delay τ_i and its time rate $\dot{\tau}_i$ for the DT and FS in the hardware processing, what we actually need to search are the residual group delay $\Delta\tau_G = \tau_G - \tau_i$ and the residual fringe frequency $\Delta F_r = F_r - \omega_0\dot{\tau}_i$, where ω_0 is a center frequency of the RF band.

This second step is mandatory for geodetic VLBI, since the parameters estimated in this step, the group delay and the fringe frequency, are the basic observables of current geodetic VLBI, as we will discuss later. Therefore, the second-step processing is usually performed in correlation centers, and resultant group delay and fringe frequency values are provided to scientific users in the geodetic VLBI community.

For astrophysical imaging VLBI, this step is usually left to users, who directly analyse outputs of hardware correlators for observed astronomical radio sources of their interest. When necessary, they can use appropriate fringe fitting tasks in existing image synthesis softwares, such as the one called “FRING” in the “AIPS” (Astronomical Image Processing System) developed by the NRAO (National Radio Astronomy Observatory), USA.

4.3.1 Some Words on Terminology

Perhaps, it is worthwhile here to put a variety of technical terms in order, which are frequently used in discussions on the fringe fitting but in a little confusing way.

1. We defined in Chapter 3 the group delay τ_G and the fringe frequency F_r as derivatives in frequency and time, respectively, of the phase spectrum $\Phi(\omega, t)$ of the cross-correlation (i.e. phase of the cross-power spectrum) of received voltages. Generally speaking, quantities thus defined are functions of frequency ω and time t , i.e.

$$\tau_G(\omega, t) = \frac{\partial\Phi(\omega, t)}{\partial\omega},$$
$$F_r(\omega, t) = \frac{\partial\Phi(\omega, t)}{\partial t}.$$

In the context of the fringe fitting, however, the same terms “group delay” and “fringe frequency” are mostly used to indicate their values $\tau_G(\omega_0, t_0)$ and $F_r(\omega_0, t_0)$ at certain reference frequency ω_0 and reference time t_0 . The reference frequency and reference time are usually selected at a center frequency of an observing band and at a midpoint of an integration time in a correlation processing of a radio source, respectively. We will follow this convention in subsequent discussions.

2. Sometimes, we talk about “delay rate” instead of the fringe frequency, since the fringe frequency is just proportional to the delay rate. Specifically, the delay rate $\dot{\tau}_\phi$ is a quantity which is equal to the fringe frequency divided by the RF center frequency ω_0 :

$$\dot{\tau}_\phi = \frac{F_r(\omega_0, t_0)}{\omega_0}.$$

3. A term “fringe rate” is frequently used as another name of the fringe frequency F_r .
4. We introduced the fringe frequency F_r as an angular frequency in Chapter 3, which is measured in units of rad/sec. However, the fringe frequency is often defined as a physical frequency in unit of Hz. We will denote this as f_r in order to distinguish from ours. Then, we have

$$f_r = \frac{F_r}{2\pi}, \quad \text{and} \quad \dot{\tau}_\phi = \frac{f_r(\nu_0, t_0)}{\nu_0},$$

where $\nu_0 = \omega_0/(2\pi)$ is the center physical frequency of the RF band.

4.3.2 Two-Dimensional Search in Software Processing

We now consider how the signal part of the correlator output is processed in this second step, disregarding the noise contribution. For this purpose, we use the expected complex cross-correlation $\langle \mathcal{R}_k(\Delta\tau_l) \rangle$ at k -th AP and l -th lag, which contains only the signal contribution coming from a radio source, as we saw in Subsection 4.2.16.

As we saw in Subsection 4.2.19, the matrix of expected complex cross-correlations show some systematic oscillatory pattern over the lag-number (l) vs AP-number (k) plane, as illustrated once again in Figure 90 (this Figure shows expected real cross-correlation only).

Although theoretically predicted fringe frequency (or delay rate) must be good enough within the short timescale of one accumulation period (AP) for the hardware integration, it is not necessarily so over many AP's. Thus, the expected complex cross-correlations usually show oscillations along the AP-number axis due to the residual fringe frequency.

Also, predicted delay usually contains some error. Therefore, the maximum amplitude of the oscillation appears not at the central lag, but somewhere in the lag-number axis shifted from the central lag by an amount of the residual group delay.

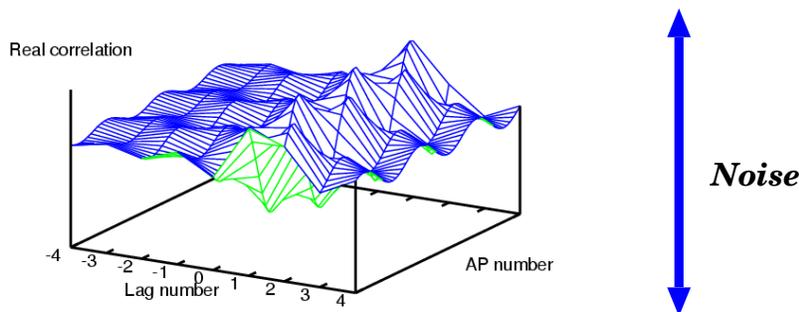


Figure 90: Expected real cross-correlation on the lag-number vs AP-number plane. Lag number shown here is $m = l - n/2$ which takes zero at the central lag. In actual correlator outputs, this kind of systematic oscillatory pattern of the correlated signal is usually masked by the overwhelming noise.

The lag number where the along-AP oscillation shows the maximum amplitude must correspond to the correlation peak of the white fringe, which we are looking for. So, at first sight, it appears rather easy to find the residual group delay corresponding to the correlation peak by just searching for a point in the lag-number axis where the along-AP oscillation shows maximum amplitude.

In actuality, however, the signal component, which shows the systematic pattern, is usually deeply buried in the overwhelming noise, as indicated in Figure 90. Then, we cannot see any systematic pattern at the beginning. We can find the maximum correlation with a high signal-to-noise ratio only after further integration of the hardware-correlator outputs during some interval of time, well longer than a single AP and sufficient to depress the noise.

In order to make the integration meaningful, however, we need a correction for the residual fringe frequency, which is accurate enough to stop the oscillation of the expected complex cross-correlation along the AP-number axis. Otherwise, the integration would give no result other than zeros.

But, we do not know the correct residual fringe frequency value beforehand. Therefore, the residual fringe frequency must be also searched for, along with the residual group delay, on the basis of the hardware–correlator outputs themselves.

Thus, we must simultaneously search for two parameters: the residual group delay and the residual fringe frequency, which allow us to find the correlation peak. In other words, we need **a two–dimensional search for the correlation peak in the group–delay vs fringe–frequency plane.**

Figure 91 illustrates the need of the two–dimensional search.

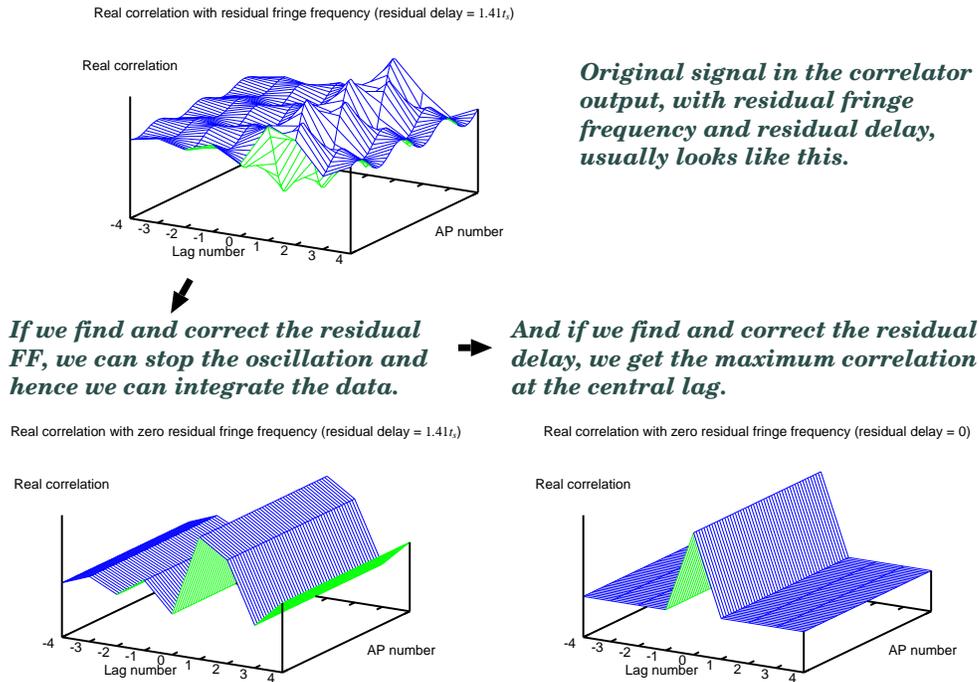


Figure 91: We can integrate the correlator outputs over many AP’s depressing the strong noise and find the correlation peak, if we successfully search and correct for the residual fringe frequency and the residual group delay.

4.3.3 How to Apply Corrections to Correlator Outputs?

In the two–dimensional search, we need to apply corrections for the residual group delay and the residual fringe frequency to the complex cross–correlations in order to improve incomplete predictions of the delay and fringe frequency which we applied in the hardware processing.

However, it is not easy to directly apply corrections to the expected complex cross-correlation $\langle \mathcal{R}_k(\Delta\tau_l) \rangle$. In fact, even in the simplest case of the white-spectrum source with rectangular videoband, which we saw in Subsection 4.2.18, the quantities to be corrected, i.e. the residual delay in equation (264):

$$\Delta\tau_D = \Delta\tau_g + \tau_C + \tau_A,$$

and time variation of the residual phase in equation (265):

$$\Phi_F = \omega_{LO}\Delta\tau_D + \theta_1 - \theta_2 - \Phi_v,$$

due to the residual fringe frequency, are contained in the expected complex cross-correlation in a little complicated way. As we see from equations (266) and (267), the expected complex cross-correlation in this case is given by

$$\langle \mathcal{R}_k(\Delta\tau_l) \rangle = \frac{\mathcal{C} A_0 GB |\mathcal{V}|}{2} \frac{\sin\left(\frac{\pi}{2} \frac{\Delta\tau_D - \Delta\tau_l}{t_s}\right)}{\frac{\pi}{2} \frac{\Delta\tau_D - \Delta\tau_l}{t_s}} e^{-i\left(\Phi_F + \frac{\pi}{2} \frac{\Delta\tau_D - \Delta\tau_l}{t_s}\right)}.$$

We could apply a correction $\delta\Phi_F$ to Φ_F by multiplying $e^{i\delta\Phi_F}$ to $\langle \mathcal{R}_k(\Delta\tau_l) \rangle$. But, evidently, such a simple method does not work for the residual delay $\Delta\tau_D$ appearing in the sinc function. Therefore, we need to find a better way.

An elegant solution of this difficulty comes out, if we take the spectrum of the complex cross-correlation.

4.3.4 Spectrum of Complex Cross-Correlation

A very convenient form for applying corrections to the predicted delay and fringe frequency is obtained when we calculate a spectrum of the expected complex cross-correlation.

In the USB reception case, a general form for the expected complex cross-correlation at l -th lag and k -th AP was given in equation (259):

$$\begin{aligned} \langle \mathcal{R}_k(\Delta\tau_l) \rangle &= \mathcal{C} \tilde{R}_{xy}(\Delta\tau_l) = \frac{\mathcal{C} A_0}{4\pi} \left\{ e^{-i[\omega_{LO}(\Delta\tau_g + \tau_C + \tau_A) + \theta_1 - \theta_2]} \right. \\ &\quad \left. \times \int_0^\infty \mathcal{V}(\omega_{LO} + \omega) e^{-i\omega(\Delta\tau_g + \tau_C + \tau_A - \Delta\tau_l)} H_1(\omega) H_2^*(\omega) d\omega \right\}, \end{aligned}$$

where notations are the same with those used in Subsection 4.2.16.

We can calculate a spectrum $\langle S_k(\omega) \rangle$ of this expected complex cross-correlation at k -th AP by Fourier transforming the above equation with

respect to the delay $\tau = \Delta\tau_l$ inserted by the multi-lag correlator:

$$\langle S_k(\omega) \rangle = \int_{-\infty}^{\infty} \langle \mathcal{R}_k(\tau) \rangle e^{-i\omega\tau} d\tau.$$

We can do this, because VLBI correlators have many lags. Yes, VLBI correlators are powerful spectrometers, indeed!

We denoted here, and will denote in following discussions, the delay $\Delta\tau_l$ simply as τ , and regard it as a continuous variable extended from $-\infty$ to $+\infty$, so that we can use familiar Fourier integrals. Of course, actual complex cross-correlations are given at finite number of discrete delay values $\Delta\tau_1, \dots, \Delta\tau_n$ only. In this sense, the above procedure is not very rigorous. Nevertheless, we will go on using Fourier integrals because the discrete Fourier transformation should yield essentially the same results, at a discrete set of frequency values, with those obtained by the continuous Fourier transformation.

Then, the Fourier transformation gives

$$\begin{aligned} \langle S_k(\omega) \rangle &= \int_{-\infty}^{\infty} \langle \mathcal{R}_k(\tau) \rangle e^{-i\omega\tau} d\tau = \frac{\mathcal{C} A_0}{4\pi} e^{-i[\omega_{LO}(\Delta\tau_g + \tau_C + \tau_A) + \theta_1 - \theta_2]} \\ &\times \int_0^{\infty} \mathcal{V}(\omega_{LO} + \omega') e^{-i\omega'(\Delta\tau_g + \tau_C + \tau_A)} H_1(\omega') H_2^*(\omega') \int_{-\infty}^{\infty} e^{-i(\omega' - \omega)\tau} d\tau d\omega', \end{aligned} \quad (269)$$

where the inner integral is nothing but one of definitions of the delta function:

$$\int_{-\infty}^{\infty} e^{-i(\omega' - \omega)\tau} d\tau = 2\pi \delta(\omega' - \omega). \text{ Therefore, we obtain}$$

$$\begin{aligned} \langle S_k(\omega) \rangle &= \begin{cases} \frac{1}{2} \mathcal{C} A_0 \mathcal{V}(\omega_{LO} + \omega) H_1(\omega) H_2^*(\omega) e^{-i[(\omega_{LO} + \omega)(\Delta\tau_g + \tau_C + \tau_A) + \theta_1 - \theta_2]} & \text{for } \omega \geq 0, \\ 0 & \text{for } \omega < 0. \end{cases} \end{aligned} \quad (270)$$

This is the spectrum of the expected complex cross-correlation at k -th AP in the USB case. Note that all terms to be corrected got really nicely gathered in the argument of the exponential function!

In the positive frequency range $\omega \geq 0$, the obtained spectrum in equation (270) is just equal to $2\mathcal{C}$ times the positive frequency part of the ordinary

cross-power spectrum of the received signals in the USB reception:

$$S_{v_{F1}v_{F2}}^{USB}(\omega) = \frac{1}{4} A_0 \mathcal{V}(\omega_{LO} + \omega) H_1(\omega) H_2^*(\omega) e^{-i[(\omega_{LO} + \omega)(\Delta\tau_g + \tau_C + \tau_A) + \theta_1 - \theta_2]},$$

for $\omega \geq 0$,

which we obtained in equation (221) (see Figure 92).

Needless to say that the word “ $2\mathcal{C}$ times” here refers to the amplitudes only. The phase spectra in the positive frequency range $\omega \geq 0$ are exactly

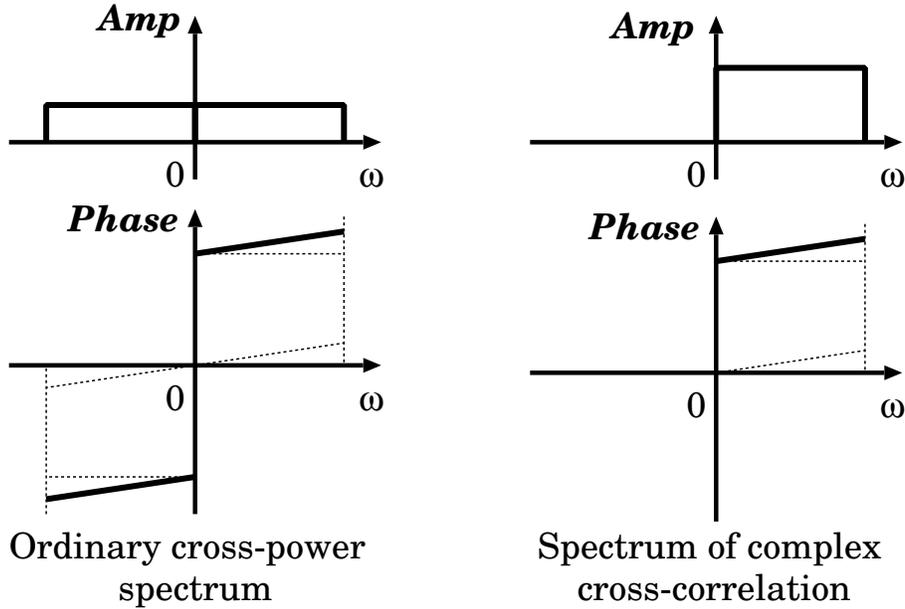


Figure 92: Hermitian symmetric ordinary cross-power spectrum of the received signals (left), and spectrum of the complex cross-correlation (right) in the USB reception case. We use here a simple model of a white-spectrum source in a rectangular videoband.

the same in the two spectra of Figure 92. Of course, this coincidence is not surprising at all, since we used this spectral form of equation (221) to derive the expected real and imaginary cross-correlations.

Unlike the ordinary cross-power spectrum, which is Hermitian symmetric, the spectrum of the expected complex cross-correlation is localized (or folded) in the positive frequency range only (see Figure 92). This is another merit of this spectrum for applying corrections.

4.3.5 Amplitude and Phase Spectra of Complex Cross–Correlation

Let us now express equation (270) in an amplitude–phase form:

$$\langle S_k(\omega) \rangle = \begin{cases} \mathcal{S}(\omega) e^{-i\Delta\Phi(\omega, t_k)} & \text{for } \omega \geq 0, \\ 0 & \text{for } \omega < 0, \end{cases} \quad (271)$$

introducing notations for amplitude spectrum $\mathcal{S}(\omega)$ and residual phase spectrum $\Delta\Phi(\omega, t)$ of the expected complex cross–correlation at central time t_k of k -th AP:

$$\begin{aligned} \mathcal{S}(\omega) &= \frac{\mathcal{C} A_0}{2} | \mathcal{V}(\omega_{LO} + \omega) | | H_1(\omega) H_2^*(\omega) |, \\ \Delta\Phi(\omega, t) &= (\omega_{LO} + \omega)(\Delta\tau_g + \tau_C + \tau_A) + \theta_1 - \theta_2 - \Phi_v - \Phi_b + 2\pi m, \end{aligned} \quad (272)$$

where \mathcal{C} is the constant coefficient given in equation (256), amplitude spectrum $| \mathcal{V}(\omega_{LO} + \omega) |$ and phase spectrum $\Phi_v(\omega, t)$ of the complex visibility are given by

$$\mathcal{V}(\omega_{LO} + \omega) = | \mathcal{V}(\omega_{LO} + \omega) | e^{i\Phi_v(\omega, t)},$$

and amplitude spectrum $| H_1(\omega) H_2^*(\omega) |$ and phase spectrum $\Phi_b(\omega, t)$ of the bandwidth characteristics are given by

$$H_1(\omega) H_2^*(\omega) = | H_1(\omega) H_2^*(\omega) | e^{i\Phi_b(\omega, t)},$$

while $2\pi m$ term stands for the cycle ambiguity. We approximate hereafter the amplitude spectrum $\mathcal{S}(\omega)$ to be constant in time.

We regard equation (271) as a general form for the spectrum of the expected complex cross–correlation at k -th AP in the USB reception case.

We called $\Delta\Phi(\omega, t)$ as the residual phase spectrum, because this is the difference:

$$\Delta\Phi(\omega, t) = \Phi(\omega, t) - \Phi_i(\omega, t),$$

between the full phase of the cross–power spectrum of the received signals:

$$\Phi(\omega, t) = (\omega_{LO} + \omega)(\tau_{g_0} + \tau_C + \tau_A) + \theta_1 - \theta_2 - \Phi_v - \Phi_b + 2\pi m,$$

and its theoretical prediction:

$$\Phi_i(\omega, t) = (\omega_{LO} + \omega)\tau_i.$$

4.3.6 Two-Dimensional Taylor Expansion of Residual Phase

Now we assume a continuum spectrum source.

Let us approximate the residual phase spectrum $\Delta\Phi(\omega, t)$ of the expected complex cross-correlation of the source by a linear function with respect to both frequency ω and time t , around a reference point P in the ω vs t plane located at $\omega = \omega_I$ and $t = t_0$. The reference frequency ω_I and reference time t_0 are usually selected at a central frequency ω_I of the IF (video) band and at a central time t_0 of an integration, respectively.

For this purpose, we expand the residual phase spectrum into the two-dimensional Taylor series, and leave only linear terms in the expansion:

$$\Delta\Phi(\omega, t) \cong \Delta\Phi(\omega_I, t_0) + \left. \frac{\partial\Delta\Phi}{\partial t} \right|_P (t - t_0) + \left. \frac{\partial\Delta\Phi}{\partial\omega} \right|_P (\omega - \omega_I), \quad (273)$$

where $|_P$ means a value at the reference point P with $\omega = \omega_I$ and $t = t_0$.

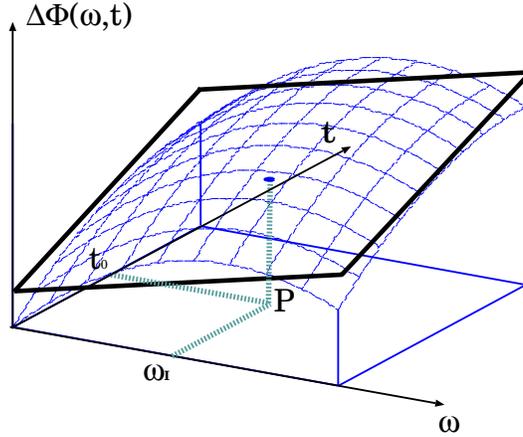


Figure 93: The linear approximation of the residual phase with respect to the time and frequency.

This is equivalent to approximating a curved surface of the residual phase spectrum $\Delta\Phi(\omega, t)$ by a plane which is tangent to the curved surface at the point P with $\omega = \omega_I$ and $t = t_0$ (Figure 93).

We can justify such approximation, provided that the second and higher order terms in the Taylor series expansion with respect to $t - t_0$ and $\omega - \omega_I$ are well smaller than 1 radian.

This condition is mostly satisfied for the $\omega - \omega_I$ dependence. In fact, the major phase terms due to the delays, i.e. $(\omega_{LO} + \omega)(\Delta\tau_g + \tau_C + \tau_A)$, are almost exactly linear with respect to $\omega - \omega_I$, except for the term due to

the ionospheric delay. Also, frequency dependence of other phase terms are intrinsically weak. The phase term due to the ionospheric delay is inversely proportional to the RF frequency $\omega_{LO} + \omega$. Therefore, n -th order term in the Taylor series expansion is proportional to $[(\omega - \omega_I)/(\omega_{LO} + \omega_I)]^n$. This must rapidly decrease with increasing n , since the RF frequency $\omega_{LO} + \omega_I$ is usually much larger than πB , where B is the bandwidth, and, of course, $|\omega - \omega_I| \leq \pi B$. Hence, the second and higher order terms are mostly negligible in the ionospheric delay effect, as well.

As for the $t - t_0$ dependence, the term $(\omega_{LO} + \omega)\Delta\tau_g$ due to the residual geometric delay, and also the visibility phase term $\Phi_v(\omega, t)$, vary in time usually following the Earth's diurnal rotation. Therefore, timescale of their variation is rather long compared with the integration time, which is mostly shorter than the coherence time. Thus, unless $\Delta\tau_g$ is large due, for example, to imperfect theoretically predicted delay, and/or the integration time is too long, we can keep the second and higher order terms to be well smaller than 1 radian. Also, non-linear time variations of phase due to the clock offset and atmospheric delay $(\omega_{LO} + \omega)(\tau_C + \tau_A)$, which remain after subtraction of usually dominating linear trends, are likely to be well smaller than 1 radian, as far as the integration time is shorter than the coherence time.

Consequently, the linear approximation in equation (273) should be good enough for most of ordinary VLBI observations, with integration times shorter than the coherence time.

However, the linear approximation may not be adequate if the coherence time is unexpectedly short due to a bad weather condition or a clock failure, and/or our theoretically predicted delay is incorrect. We will consider such a case later in coherence loss discussions.

4.3.7 Residual Group Delay and Residual Fringe Frequency

We introduced in Chapter 3 the group delay τ_G and the fringe frequency F_r as derivatives with respect to frequency and time, respectively, of the phase spectrum $\Phi(\omega, t)$ of the cross-power spectrum of the received signals. Therefore, the residual group delay $\Delta\tau_G$ and the residual fringe frequency ΔF_r which we are searching for are given by

$$\Delta\tau_G = \frac{\partial\Delta\Phi}{\partial\omega}, \quad \text{and} \quad \Delta F_r = \frac{\partial\Delta\Phi}{\partial t},$$

with a residual phase spectrum

$$\Delta\Phi = \Phi(\omega, t) - \Phi_i,$$

where $\Phi_i = (\omega_{LO} + \omega_I)\tau_i$ is the predicted phase, τ_i is the theoretical prediction of the geometric delay (the ‘‘instrumental delay’’) used for the DT and

FS, ω_{LO} is the local oscillator frequency, and ω_I is the IF (videoband) center frequency.

Now, we can use for $\Delta\Phi$ above the residual phase spectrum $\Delta\Phi(\omega, t)$ of the spectrum $\langle S_k(\omega) \rangle$ of the expected complex cross-correlation $\langle \mathcal{R}_k(\Delta\tau_l) \rangle$ given in equation (272), since this residual phase spectrum is just the same with that of the ordinary cross-power spectrum of received voltages after the DT and FS, as we saw in Subsection 4.3.4. In this case, τ_i is the instrumental delay used for the DT and FS performed in the hardware processing.

Thus, the partial derivatives

$$\left. \frac{\partial \Delta\Phi}{\partial t} \right|_P, \quad \text{and} \quad \left. \frac{\partial \Delta\Phi}{\partial \omega} \right|_P,$$

in the expression of the linearized residual phase given in equation (273) are nothing but the residual fringe frequency ΔF_r and the residual group delay $\Delta\tau_G$, respectively, at the point P with $\omega = \omega_I$ and $t = t_0$:

$$\Delta F_r = \left. \frac{\partial \Delta\Phi}{\partial t} \right|_P = \left. \frac{\partial(\Phi - \Phi_i)}{\partial t} \right|_P = F_r - (\omega_{LO} + \omega_I) \dot{\tau}_i, \quad (274)$$

$$\Delta\tau_G = \left. \frac{\partial \Delta\Phi}{\partial \omega} \right|_P = \left. \frac{\partial(\Phi - \Phi_i)}{\partial \omega} \right|_P = \tau_G - \tau_i. \quad (275)$$

Using the residual fringe frequency ΔF_r and the residual group delay $\Delta\tau_G$, we can rewrite the linearized residual phase spectrum of equation (273) as

$$\Delta\Phi(\omega, t) \cong \Delta\Phi(\omega_I, t_0) + \Delta F_r (t - t_0) + \Delta\tau_G (\omega - \omega_I). \quad (276)$$

Then, equation (271) for the spectrum of the expected complex cross-correlation at k -th AP is now reduced to

$$\langle S_k(\omega) \rangle = \begin{cases} \mathcal{S}(\omega) e^{-i[\Delta\Phi(\omega_I, t_0) + \Delta F_r (t_k - t_0) + \Delta\tau_G (\omega - \omega_I)]} & \text{for } \omega \geq 0 \\ 0 & \text{for } \omega < 0, \end{cases} \quad (277)$$

as far as our linear approximation model of the residual phase spectrum of the expected complex cross-correlation is valid.

4.3.8 Trial Corrections to Residual Phase

Let us now apply a set of trial corrections δF_r and $\delta\tau_G$ for the residual fringe frequency and the residual group delay, respectively, in the spectrum $\langle S_k(\omega) \rangle$ of the expected complex cross-correlation at k -th AP.

What is needed for this is really simple. It is just to multiply an exponential function

$$e^{i[\delta F_r (t_k - t_0) + \delta \tau_G (\omega - \omega_I)]},$$

to the spectrum given in equation (277)!!

In fact, we will then obtain a “corrected spectrum” $\langle S'_k(\omega) \rangle$:

$$\begin{aligned} \langle S'_k(\omega) \rangle &= \langle S_k(\omega) \rangle e^{i[\delta F_r (t_k - t_0) + \delta \tau_G (\omega - \omega_I)]} \\ &= \begin{cases} \mathcal{S}(\omega) e^{-i[\Delta \Phi(\omega_I, t_0) + (\Delta F_r - \delta F_r)(t_k - t_0) + (\Delta \tau_G - \delta \tau_G)(\omega - \omega_I)]} & \text{for } \omega \geq 0 \\ 0 & \text{for } \omega < 0. \end{cases} \end{aligned} \quad (278)$$

This equation implies that if the trial corrections δF_r and $\delta \tau_G$ happen to be close enough to the actual residual fringe frequency ΔF_r and residual group delay $\Delta \tau_G$, respectively, then we can effectively eliminate the time and frequency dependence of the “corrected residual phase spectrum”:

$$\begin{aligned} \Delta \Phi'(\omega, t_k) &= \Delta \Phi(\omega_I, t_0) + (\Delta F_r - \delta F_r)(t_k - t_0) + (\Delta \tau_G - \delta \tau_G)(\omega - \omega_I) \\ &\cong \Delta \Phi(\omega_I, t_0), \end{aligned} \quad (279)$$

as schematically illustrated in Figure 94. Thus we can stop oscillations both

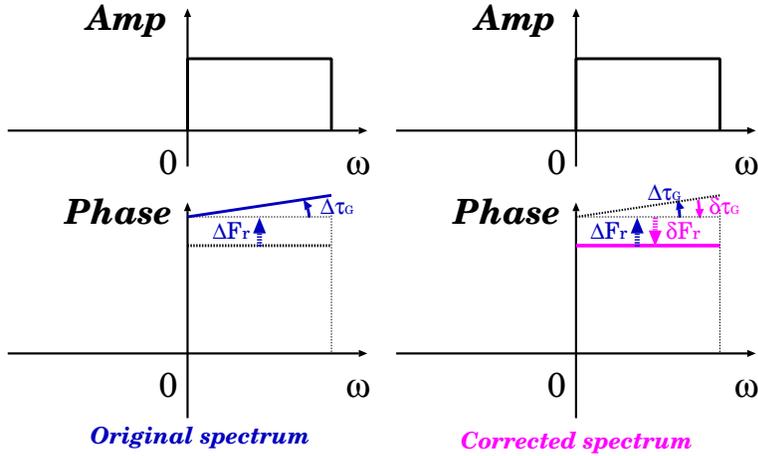


Figure 94: The residual fringe frequency shifts the phase spectrum by ΔF_r within a unit time, and the residual group delay causes inclination of the phase spectrum by $\Delta \tau_G$ (left). Suitable corrections δF_r and $\delta \tau_G$ can compensate these changes and keep the phase spectrum constant both in time and frequency (right).

in time and frequency in the corrected spectrum.

4.3.9 Finding Correlation Peak

Now we consider how to search for the correlation peak of the white fringe, assuming that our linear approximation model in equation (277) is good enough. For this purpose, we form a function $\mathcal{P}(\delta F_r, \delta \tau_G)$ of trial corrections δF_r and $\delta \tau_G$, which we call “peak finder”, by integrating the corrected spectrum $\langle S'_k(\omega) \rangle$ at k -th AP with frequency ω , and averaging corrected spectra over all AP's (i.e. in time):

$$\begin{aligned} \mathcal{P}(\delta F_r, \delta \tau_G) &= \frac{1}{2\pi N} \sum_{k=1}^N \int_{-\infty}^{+\infty} \langle S'_k(\omega) \rangle d\omega = \frac{1}{2\pi N} \sum_{k=1}^N \int_0^{2\pi B} \langle S'_k(\omega) \rangle d\omega \\ &= \frac{1}{2\pi N} \sum_{k=1}^N \int_0^{2\pi B} \mathcal{S}(\omega) e^{-i[\Delta\Phi(\omega_I, t_0) + (\Delta F_r - \delta F_r)(t_k - t_0) + (\Delta\tau_G - \delta\tau_G)(\omega - \omega_I)]} d\omega, \end{aligned} \quad (280)$$

where B is the bandwidth, and N is the number of accumulation periods.

It is evident that oscillations in time and in frequency of the exponential function in the integrand of equation (280) cease, and therefore the amplitude of the peak-finder function $|\mathcal{P}(\delta F_r, \delta \tau_G)|$ takes its maximum value, when the trial corrections are equal to the actual residuals, i.e. $\delta F_r = \Delta F_r$ and $\delta \tau_G = \Delta \tau_G$. For all other trial correction values $\delta F_r \neq \Delta F_r$ and/or $\delta \tau_G \neq \Delta \tau_G$, we get smaller amplitude due to oscillations of the integrand.

As the simplest example, we consider a case when $\mathcal{S}(\omega)$ is a white spectrum, i.e. $\mathcal{S}(\omega) = \mathcal{S} = \text{const}$. We assume, for definiteness, that N is an even number, $t_{N/2} = t_0$, and $t_k - t_0 = j t_{AP}$ with $j = k - N/2$, where t_{AP} is the accumulation period. Also, we note that $\omega_I = \pi B$ here. In this case, equation (280) yields

$$|\mathcal{P}(\delta F_r, \delta \tau_G)| = \mathcal{S}B \left| \frac{\sin[\pi B (\Delta\tau_G - \delta\tau_G)]}{\pi B (\Delta\tau_G - \delta\tau_G)} \frac{\sin(N t_{AP} \frac{\Delta F_r - \delta F_r}{2})}{N \sin(t_{AP} \frac{\Delta F_r - \delta F_r}{2})} \right|.$$

Figure 95 shows normalized amplitude $|\mathcal{P}(\delta F_r, \delta \tau_G)|/(\mathcal{S}B)$ of the peak-finder function against normalized trial corrections $\delta F_r t_{AP}$ and $\delta \tau_G / t_s$, where t_s is the Nyquist interval $t_s = 1/(2B)$. In this figure, we gave $\Delta F_r = 0.54/t_{AP}$ and $\Delta \tau_G = 2.12 t_s$ for values of the residual fringe frequency ΔF_r and the residual group delay $\Delta \tau_G$, respectively. Figure 95 clearly shows that the amplitude of the peak-finder function $|\mathcal{P}(\delta F_r, \delta \tau_G)|$ indeed takes its maximum when the trial corrections satisfy $\delta F_r = \Delta F_r$ and $\delta \tau_G = \Delta \tau_G$.

In other words, if we find a set of corrections δF_{r_m} and $\delta \tau_{G_m}$, which makes the amplitude of the peak-finder function $|\mathcal{P}(\delta F_r, \delta \tau_G)|$ largest, they must

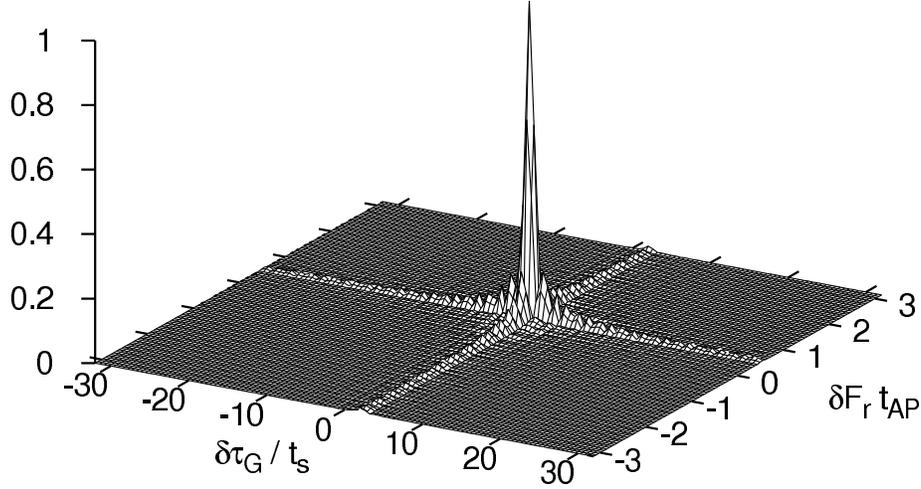


Figure 95: A plot of model normalized amplitude of peak-finder function $\mathcal{P}/(SB)$ against normalized trial corrections $\delta\tau_G / t_s$ and $\delta F_r t_{AP}$ in case of a white spectrum $\mathcal{S}(\omega) = \mathcal{S} = \text{const}$. Residual group delay and residual fringe frequency are assumed to be at $\Delta\tau_G = 2.12 t_s$ and $\Delta F_r = 0.54 / t_{AP}$, respectively.

be equal to the residual fringe frequency ΔF_r and the residual group delay $\Delta\tau_G$, i.e. $\delta F_{r_m} = \Delta F_r$ and $\delta\tau_{G_m} = \Delta\tau_G$, respectively, at the band center $\omega = \omega_I$ and at the center of the integration time $t = t_0$.

This means that we successfully found correct delay tracking (DT) and fringe stopping (FS) parameters, which are the residual fringe frequency ΔF_r and the residual group delay $\Delta\tau_G$.

What we did above with the peak-finder function is not limited to just finding the parameters only. We also applied the right parameters to correct the spectrum $S_k(\omega)$ of the expected complex cross-correlation at a point $\delta F_r = \delta F_{r_m} = \Delta F_r$ and $\delta\tau_G = \delta\tau_{G_m} = \Delta\tau_G$. Therefore, we achieved **perfect** DT and FS there, as far as our linear approximation model is appropriate.

Moreover, when the fringe-frequency correction δF_r is equal to the right residual fringe frequency $\delta F_{r_m} = \Delta F_r$, the corrected spectrum becomes constant in time and therefore the time averaging (averaging over all AP's) does not change anything. Then, equation (280) is reduced to

$$\begin{aligned} \mathcal{P}(\delta F_{r_m}, \delta\tau_G) &= \frac{1}{2\pi} \int_0^{2\pi B} \mathcal{S}(\omega) e^{-i[\Delta\Phi(\omega_I, t_0) + (\Delta\tau_G - \delta\tau_G)(\omega - \omega_I)]} d\omega \\ &= \frac{\mathcal{C} A_0}{4\pi} e^{-i\Delta\Phi(\omega_I, t_0)} \int_0^{2\pi B} |\mathcal{V}(\omega_{LO} + \omega)| |H_1(\omega) H_2^*(\omega)| e^{-i(\Delta\tau_G - \delta\tau_G)(\omega - \omega_I)} d\omega, \end{aligned}$$

where we took into account equation (272). It is evident that this equation shows an envelope pattern of a white fringe with respect to the “corrected residual group delay” $\Delta\tau_G - \delta\tau_G$ in the linear approximation given by equation (276). In particular, when $\mathcal{S}(\omega)$ is white, this yields the familiar sinc function envelope pattern. Therefore, the peak at $\delta F_r = \delta F_{r_m}$ and $\delta\tau_G = \delta\tau_{G_m}$ is the correlation peak of the white fringe. Thus, we finally found the correlation peak of the white fringe at the peak of the peak-finder function!

4.3.10 Fringe Amplitude and Residual Fringe Phase in VLBI

At the correlation peak $\delta F_r = \delta F_{r_m} = \Delta F_r$ and $\delta\tau_G = \delta\tau_{G_m} = \Delta\tau_G$, the peak-finder function takes a complex value given by

$$\mathcal{P}(\delta F_{r_m}, \delta\tau_{G_m}) = \frac{1}{2\pi} \int_0^{2\pi B} \mathcal{S}(\omega) e^{-i\Delta\Phi(\omega, t_0)} d\omega, \quad (281)$$

as far as our linear approximation model of the residual phase spectrum introduced in Subsection 4.3.6 is appropriate.

Integrand of this equation is the corrected spectrum of the expected complex cross-correlation as given in equation (278) at the correlation peak, where $\Delta F_r - \delta F_r = 0$ and $\Delta\tau_G - \delta\tau_G = 0$. Strictly speaking, this should be the time-averaged corrected spectrum as equation (280) demands. However, the time averaging does not change the spectrum when $\Delta F_r - \delta F_r = 0$, as far as we assume that the linear approximation model is valid. The range of integration can be formally extended to $-\infty \leq \omega \leq \infty$ since the spectrum itself is band-limited within 0 to $2\pi B$.

Therefore, equation (281) stands for an inverse Fourier transformation of the corrected spectrum of the expected complex cross-correlation at the correlation peak with zero argument ($\tau = 0$). Consequently, the $\mathcal{P}(\delta F_{r_m}, \delta\tau_{G_m})$ is a value of a “corrected” expected complex cross-correlation at $\tau = 0$ and at any time t ($t = t_0$, say), when the correlation peak is achieved with $\delta F_r = \delta F_{r_m}$ and $\delta\tau_G = \delta\tau_{G_m}$.

Of course, this corrected expected complex cross-correlation is no longer the expectation $\langle \mathcal{R}_k(\Delta\tau_l) \rangle$ of the output of our hardware correlator at k -th AP and l -th lag. This is a new corrected one which shows no along-AP oscillation and shows the maximum amplitude at the central lag, just like the one illustrated in the last panel of Figure 91.

We would get a new expected complex cross-correlation which is similar to the above corrected one, if we improve the model for the predicted delay τ_i taking into account correct information on the residual fringe frequency ΔF_r and the residual group delay $\Delta\tau_G$, and rerun the hardware correlator

using the improved model. Note however that the similarity would not be perfect. In fact, in the fringe fitting, we correct the residual fringe frequency and residual group delay but leave the constant part of the residual phase $\Delta\Phi(\omega_I, t_0)$ unaltered, while in the hardware reprocessing $\Delta\Phi(\omega_I, t_0)$ should be renewed by the improved predicted delay τ_i .

Thus, in view of equations (251) and (255), we describe the peak-finder function at the correlation peak $\mathcal{P}(\delta F_{r_m}, \delta\tau_{G_m})$ as

$$\mathcal{P}(\delta F_{r_m}, \delta\tau_{G_m}) = r_{\hat{x}\hat{y}}(0) = \mathcal{C} \mathcal{R}_{xy}(0), \quad (282)$$

where $r_{\hat{x}\hat{y}}(0)$ is the “corrected version” of the cross-correlation coefficient of digital voltages $\tilde{r}_{\hat{x}\hat{y}}[m]$ given in equation (252) and $\mathcal{R}_{xy}(0)$ is the “corrected version” of the cross-correlation of analog voltages $\tilde{R}_{xy}(m t_s)$ given in equation (254) both at zero argument $\tau = 0$. They are “corrected” through the fringe fitting, which is a floating point calculation in a computer using values of the expected complex cross-correlations at all k ’s and m ’s:

$$\langle \mathcal{R}_k(\Delta\tau_l) \rangle = \tilde{r}_{\hat{x}\hat{y}}[m] = \mathcal{C} \tilde{R}_{xy}(m t_s), \quad (k = 1, \dots, N, m = -\frac{n}{2} + 1, \dots, \frac{n}{2}).$$

Therefore, the result of the calculation is no longer a sampled quantity. This is why we use a notation $r_{\hat{x}\hat{y}}(0)$, in particular, instead of $r_{\hat{x}\hat{y}}[0]$.

The fringe fitting effectively performs as an additional “perfector of the DT and FS”. Therefore, $\mathcal{R}_{xy}(0)$ here should be equivalent to an expected output \mathcal{R} of a complex analog correlator with perfect DT and FS, as far as our fringe fitting is perfect.

We discussed such an output of a complex analog correlator in Chapter 3. We saw there that the expected output shows a sinusoidal fringe pattern near the correlation peak of a white fringe:

$$\mathcal{R} = \mathcal{A} e^{-i\Delta\Phi},$$

and called the amplitude \mathcal{A} of the fringe pattern the “fringe amplitude” (or “correlation amplitude”), and the phase $\Delta\Phi$ of the fringe pattern the “residual fringe phase” (or “residual correlation phase”). Therefore, it is natural to define the fringe amplitude \mathcal{A} in VLBI as the amplitude of $\mathcal{R}_{xy}(0)$:

$$\mathcal{A} = |\mathcal{R}_{xy}(0)|. \quad (283)$$

Then, we will call the amplitude of the peak-finder function at the correlation peak $\mathcal{P}(\delta F_{r_m}, \delta\tau_{G_m})$ as given in equations (281) and (282) the “**un-calibrated fringe amplitude**” and denote it as \mathcal{A}_U , namely

$$|\mathcal{P}(\delta F_{r_m}, \delta\tau_{G_m})| = |r_{\hat{x}\hat{y}}(0)| = \mathcal{C} |\mathcal{R}_{xy}(0)| = \mathcal{A}_U. \quad (284)$$

We call the amplitude as “uncalibrated” primarily because this is not the amplitude of the cross-correlation of analog voltages $\mathcal{R}_{xy}(0)$, but that of the cross-correlation coefficient of digital voltages $r_{\hat{x}\hat{y}}$. There is the coefficient \mathcal{C} between them, and therefore

$$\mathcal{A}_U = \mathcal{C} \mathcal{A}. \quad (285)$$

On the other hand, the phase of the peak-finder function is the same with that of $\mathcal{R}_{xy}(0)$, simply because \mathcal{C} is a real coefficient. So, we will call it the “residual fringe phase” and denote it as $\Delta\Phi$, namely

$$-\arg[\mathcal{P}(\delta F_{r_m}, \delta\tau_{G_m})] = \Delta\Phi. \quad (286)$$

Now, equation (281) gives

$$\mathcal{A}_U = |\mathcal{P}(\delta F_{r_m}, \delta\tau_{G_m})| = \frac{1}{2\pi} \int_0^{2\pi B} \mathcal{S}(\omega) d\omega, \quad (287)$$

and

$$\Delta\Phi = -\arg[\mathcal{P}(\delta F_{r_m}, \delta\tau_{G_m})] = \Delta\Phi(\omega_I, t_0), \quad (288)$$

Adopting equation (272) as a theoretical expression of the spectrum of the expected complex cross-correlation, we use equations (287), (288), (274), and (275) to yield theoretical formulae for the uncalibrated fringe amplitude, the residual fringe phase, the residual fringe frequency, and the residual group delay, correspondingly, in the USB reception case. All of them are given at the reference time t_0 and the reference frequency ω_I of the fringe fitting. They are *the uncalibrated fringe amplitude*:

$$\mathcal{A}_U = \frac{\mathcal{C} A_0}{4\pi} \int_0^{2\pi B} |\mathcal{V}(\omega_{LO} + \omega)| |H_1(\omega) H_2^*(\omega)| d\omega, \quad (289)$$

the residual fringe phase:

$$\Delta\Phi = (\omega_{LO} + \omega_I)(\Delta\tau_g + \tau_C + \tau_A) + \theta_1 - \theta_2 - \Phi_v - \Phi_b + 2\pi m, \quad (290)$$

the residual fringe frequency:

$$\Delta F_r = (\omega_{LO} + \omega_I)(\Delta\dot{\tau}_g + \dot{\tau}_C + \dot{\tau}_A) + \dot{\theta}_1 - \dot{\theta}_2 - \dot{\Phi}_v - \dot{\Phi}_b, \quad (291)$$

and *the residual group delay*:

$$\Delta\tau_G = \Delta\tau_g + \tau_C + \tau_A - \frac{\partial\Phi_v}{\partial\omega} - \frac{\partial\Phi_b}{\partial\omega}, \quad (292)$$

where notations are the same as before. Specifically, \mathcal{C} is the constant coefficient given in equation (256), A_0 is the geometric mean of effective apertures of two antennas, \mathcal{V} is the complex visibility, $H_1(\omega) H_2^*(\omega)$ is the “bandpass characteristics of combined IF filters”, $\Delta\tau_g = \tau_{g_0} - \tau_i$ is the residual geometric delay, $\tau_{g_0} = \mathbf{D} \cdot \mathbf{s}_0 / c$ is the geometric delay at a reference direction \mathbf{s}_0 of a source observed with an interferometer of baseline vector \mathbf{D} , τ_i is the theoretical prediction of the geometric delay used in the DT and FS (“the instrumental delay”), τ_C is the clock offset, τ_A is the atmospheric delay, $\theta_1 - \theta_2$ is the difference of the LO initial phases including the effect of the clock synchronization error as we defined in equation (216), $\Phi_v(\omega, t)$ is the phase spectrum of the complex visibility, $\Phi_b(\omega, t)$ is the phase spectrum of the “bandpass characteristics of combined IF filters”, ω_{LO} is the local oscillator frequency, and ω_I is the IF (videoband) center frequency.

The search procedure described above forms the basis of the software processing.

4.3.11 Coherence Factor Due to Imperfect Delay Model

We so far assumed that our linear approximation model for the residual phase spectrum of the expected complex cross-correlation given in Subsection 4.3.6 is appropriate, and therefore the DT and FS in the fringe fitting are perfect. In particular, the theoretical formula for the uncalibrated fringe amplitude as given in equation (289) is based on the assumption.

If not, however, we cannot perform a coherent integration in the fringe fitting. Then derived amplitude of the peak-finder function becomes smaller than that given in equation (289) due to the coherence loss.

In general, there are several sources of coherence losses in VLBI data processing which we have to address when we discuss an output of a VLBI correlator.

We previously discussed coherence factors describing coherence losses due to quantization η_{cQ} in Subsection 1.1.28, and due to digital processing η_{cD} in Subsection 4.2.12. They are just constants associated with quantization and digital operations. They are already taken into account in our discussions since they are included in the coefficient \mathcal{C} of the expected complex cross-correlation, as shown in equation (256).

We also discussed a coherence factor due to fringe-phase fluctuation η_{cP} in Subsection 1.2.9. This quantity is dependent on integration time τ_a of the correlation processing. So, $\eta_{cP} = \eta_{cP}(\tau_a)$, if we explicitly describe the dependence. Usually, η_{cP} is very close to 1 within a short AP of hardware integration. In the fringe fitting, however, η_{cP} could be appreciably smaller than 1, when the integration time is longer than the coherence time at the

observing date. This could happen if the coherence time is unexpectedly short due for example to a bad weather condition or a clock failure.

There is another source of coherence loss in VLBI which is associated with imperfection of theoretically predicted delay (called also as “instrumental delay” or “delay model”) used in the correlation processing.

We discussed in Subsection 4.3.6 that we can well approximate time variation of the residual phase by a linear function of time as far as our delay model is good and integration time is short enough. However, if the delay model has a large error, or the integration time is too long, second or higher order terms with respect to time could cause phase variation larger than 1 radian during the fringe fitting. This must induce coherence loss since the ordinary fringe–frequency search in the fringe fitting can estimate and correct only a linear trend of the phase variation, as we saw in preceding Subsections. For avoiding this inconvenience, some of modern fringe–fitting softwares have an option to search for the second–order term with respect to time.

In order to take into account this kind of coherence loss, we introduce a new coherence factor η_{c_M} and call it “coherence factor due to imperfect delay model”. This factor is a function of the integration time τ_a , i.e. $\eta_{c_M} = \eta_{c_M}(\tau_a)$, similarly to η_{c_P} . Usually, we do not pay much attention to this kind of loss in ordinary VLBI observations of astronomical radio sources, since mostly we can make η_{c_M} very close to 1 by applying reasonably good delay models and selecting short enough integration times. However, when we observe a new radio source or a human-made space vehicle with poorly known celestial position, η_{c_M} may significantly decrease with increasing integration time.

In the worst case, imperfect delay model may cause coherence loss even in the hardware processing stage if we select a too long AP (accumulation period). Therefore, it is highly recommended to select as short as possible AP when one is uncertain about the delay model.

At this stage, we can think that all main sources of the coherence loss have been already picked up. Then, total coherence factor η_c is given by

$$\eta_c = \eta_{c_Q} \eta_{c_D} \eta_{c_P} \eta_{c_M}. \quad (293)$$

4.3.12 Theoretical Form of Uncalibrated Fringe Amplitude

When the coherence loss due to fringe phase fluctuation and/or imperfect delay model are not negligible, the uncalibrated fringe amplitude, that is the amplitude of the peak–finder function at the correlation peak, is further reduced by factors $\eta_{c_P}(\tau_a)$ and $\eta_{c_M}(\tau_a)$ compared with the RHS of (285), i.e.

$$\mathcal{A}_U = \eta_{c_P}(\tau_a) \eta_{c_M}(\tau_a) \mathcal{C} \mathcal{A}. \quad (294)$$

Therefore, if we introduce a new coefficient \mathcal{C}_C which is defined by

$$\mathcal{C}_C = \eta_{c_P}(\tau_a) \eta_{c_M}(\tau_a) \mathcal{C} = \frac{\eta_{c_P}(\tau_a) \eta_{c_M}(\tau_a) \eta_{c_Q} \eta_{c_D}}{\sqrt{R_{xx}(0) R_{yy}(0)}} = \frac{\eta_c}{\sqrt{R_{xx}(0) R_{yy}(0)}}, \quad (295)$$

then from equations (283) and (294) we have

$$\mathcal{A}_U = \mathcal{C}_C \mathcal{A} = \frac{\eta_c}{\sqrt{R_{xx}(0) R_{yy}(0)}} \mathcal{A} = \eta_c \frac{|\mathcal{R}_{xy}(0)|}{\sqrt{R_{xx}(0) R_{yy}(0)}}, \quad (296)$$

and equation (289) now should be modified to

$$\mathcal{A}_U = \frac{\mathcal{C}_C A_0}{4\pi} \int_0^{2\pi B} |\mathcal{V}(\omega_{LO} + \omega)| |H_1(\omega) H_2^*(\omega)| d\omega, \quad (297)$$

where notations other than \mathcal{C}_C are the same with those given in Subsection 4.3.10. This is a general theoretical form for the uncalibrated fringe amplitude \mathcal{A}_U after the fringe fitting.

4.3.13 Search Window

Now we will see how we can practically perform the fringe fitting.

In the fringe fitting, ranges of selectable values for the trial corrections δF_r and $\delta \tau_G$ are limited by conditions of the hardware processing.

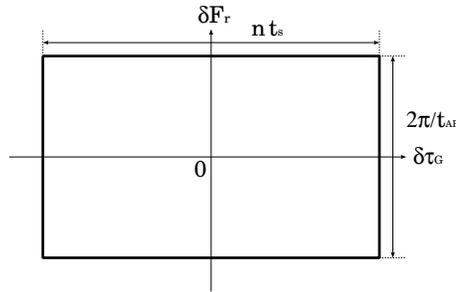


Figure 96: The search window. t_s is the sampling interval, t_{AP} is the accumulation period, and n is the number of lags.

For the fringe frequency, the selectable range is

$$-\frac{\pi}{t_{AP}} \leq \delta F_r < \frac{\pi}{t_{AP}}, \quad (298)$$

since the accumulation period t_{AP} can be the Nyquist interval, only when the physical fringe frequency spectrum $f_r = F_r/(2\pi)$ is band-limited within

a bandwidth (in the positive frequency range) $1/2t_{AP}$. This range is called the “fringe frequency window”.

When residual fringe frequency could be large due to unknown clock rate or imperfection of delay model, one should select a short AP for getting a wide enough fringe frequency window.

For the group delay, the selectable range is

$$-\frac{n}{2}t_s \leq \delta\tau_G < \frac{n}{2}t_s, \quad (299)$$

where t_s is the sampling interval and n is the number of lags, since the hardware correlator calculates the complex cross-correlations only in this range. This range is called the “delay window”.

The fringe frequency window and delay window, taken together, is called the “search window”. Figure 96 illustrates the search window.

4.3.14 Actual Software Processing

Now let us take into account noise contribution contained in the actual correlator output which we called the complex cross-correlation, $\mathcal{R}_k(\Delta\tau_l)$, where k is the AP number and l is the lag number. Then we have

$$\mathcal{R}_k(\Delta\tau_l) = \langle \mathcal{R}_k(\Delta\tau_l) \rangle + n_k(\tau_l), \quad (300)$$

where the expected complex cross-correlation $\langle \mathcal{R}_k(\Delta\tau_l) \rangle$ is the signal part, and $n_k(\tau_l)$ is the noise part, of the complex cross-correlation at k -th AP and l -th lag.

Typical procedure of the software processing consists of following steps.

1. Calculate spectra of correlator outputs $\mathcal{R}_k(\tau)$ at each AP.

First, we Fourier transform complex cross-correlations at every AP to get their spectra $S_k(\omega)$:

$$\begin{aligned} S_k(\omega) &= \int_{-\infty}^{\infty} \mathcal{R}_k(\tau) e^{-i\omega\tau} d\tau = \int_{-\infty}^{\infty} [\langle \mathcal{R}_k(\Delta\tau_l) \rangle + n_k(\tau_l)] e^{-i\omega\tau} d\tau, \\ &= \langle S_k(\omega) \rangle + N_k(\omega), \end{aligned} \quad (301)$$

where $\langle S_k(\omega) \rangle$ is the signal spectrum, i.e. the spectrum of the expected complex cross-correlation, having the amplitude $\mathcal{S}(\omega)$ and the residual phase $\Delta\Phi(\omega, t)$, as given in equations (271) and (272), and

$$N_k(\omega) = \int_{-\infty}^{\infty} n_k(\tau_l) e^{-i\omega\tau} d\tau,$$

is the noise spectrum.

Here we again adopt the familiar integral form of the Fourier transformation. In calculations of actual correlator outputs, the discrete Fourier transformation is carried out over equally spaced discrete delay values: $\Delta\tau_1 = (-n/2 + 1)t_s, \dots, \Delta\tau_{n/2} = 0, \dots, \Delta\tau_n = (n/2)t_s$, with the Fast Fourier Transform (FFT) algorithm.

Of course, this step of Fourier transformation can be skipped when we use an FX-type hardware correlator, since the correlator directly provides the complex cross-power spectra of received signals from the beginning.

2. Build “search function” on grid points.

Then we divide the search window into $n \times N$ equal meshes and select

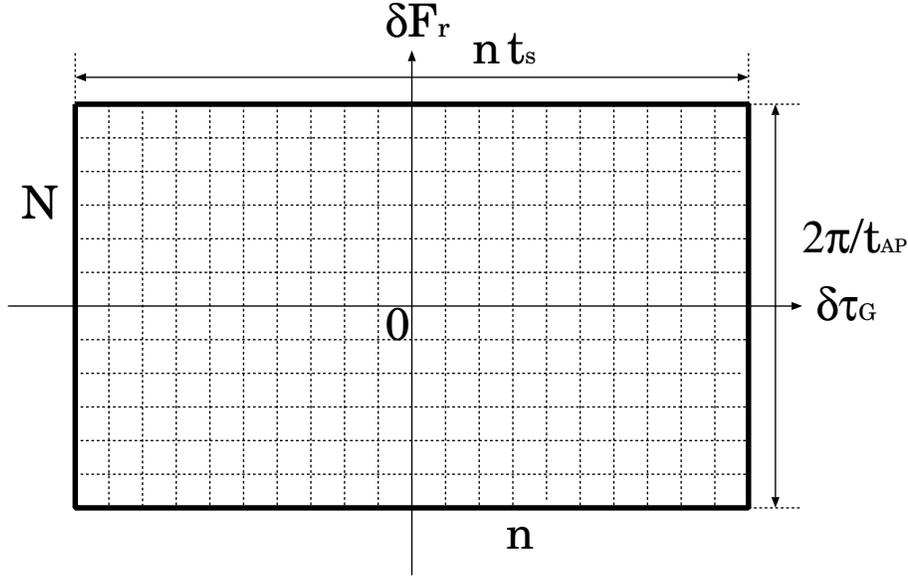


Figure 97: Search function is first calculated for trial corrections selected on grid points in the search window.

trial correction values for the fringe frequency and the group delay, δF_r and $\delta\tau_G$, respectively, first on the grid points (Figure 97).

For each set of trial correction values, we calculate so-called “**search function**” $\mathcal{F}(\delta F_r, \delta\tau_G)$ defined by:

$$\mathcal{F}(\delta F_r, \delta\tau_G) = \frac{1}{2\pi N} \sum_{k=1}^N \int_0^{2\pi B} S_k(\omega) e^{i[\delta F_r(t_k - t_0) + \delta\tau_G(\omega - \omega_I)]} d\omega. \quad (302)$$

This is an analog of the peak-finder function $\mathcal{P}(\delta F_r, \delta \tau_G)$, which we introduced for the noise-free case, in the present noise-existing case.

Note that the integration above is carried out in the positive frequency range only. This choice of the integration range does not distort signal information from an astronomical source, since the signal part of the spectrum of the complex cross-correlation is localized in the positive frequency range only, as we saw in Subsection 4.3.4. At the same time, this choice cuts off undesirable noise coming from the negative frequency side.

Separating the signal and noise parts, we have

$$\begin{aligned}
& \mathcal{F}(\delta F_r, \delta \tau_G) \\
&= \frac{1}{2\pi N} \sum_{k=1}^N \int_0^{2\pi B} [\langle S_k(\omega) \rangle + N_k(\omega)] e^{i[\delta F_r(t_k - t_0) + \delta \tau_G(\omega - \omega_I)]} d\omega \\
&= \frac{1}{2\pi N} \sum_{k=1}^N \int_0^{2\pi B} [\langle S_k(\omega) \rangle e^{i[\delta F_r(t_k - t_0) + \delta \tau_G(\omega - \omega_I)]} + N'_k(\omega)] d\omega, \\
&= \frac{1}{2\pi N} \sum_{k=1}^N \int_0^{2\pi B} \left[\mathcal{S}(\omega) e^{-i[\Delta \Phi(\omega_I, t_0) + (\Delta F_r - \delta F_r)(t_k - t_0) + (\Delta \tau_G - \delta \tau_G)(\omega - \omega_I)]} \right. \\
&\quad \left. + N'_k(\omega) \right] d\omega, \tag{303}
\end{aligned}$$

where we used equation (277) for the signal spectrum $\langle S_k(\omega) \rangle$, and introduced a modified noise spectrum:

$$N'_k(\omega) = N_k(\omega) e^{i[\delta F_r(t_k - t_0) + \delta \tau_G(\omega - \omega_I)]}.$$

The actual integration is again carried out on discrete frequency values with the FFT algorithm.

3. Find grid point where amplitude of search function takes maximum value.

If the trial correction values δF_r and $\delta \tau_G$ are close enough to the actual residual values ΔF_r and $\Delta \tau_G$, the signal amplitude is kept nearly constant in time $\cong \mathcal{S}(\omega)$, while the noise $N'_k(\omega)$ is suppressed in proportion to $1/\sqrt{B\tau_a}$, in the course of integration with frequency and averaging in time of the search function in equation (303), where B is the bandwidth and τ_a is the integration time ($\tau_a = N t_{AP}$).

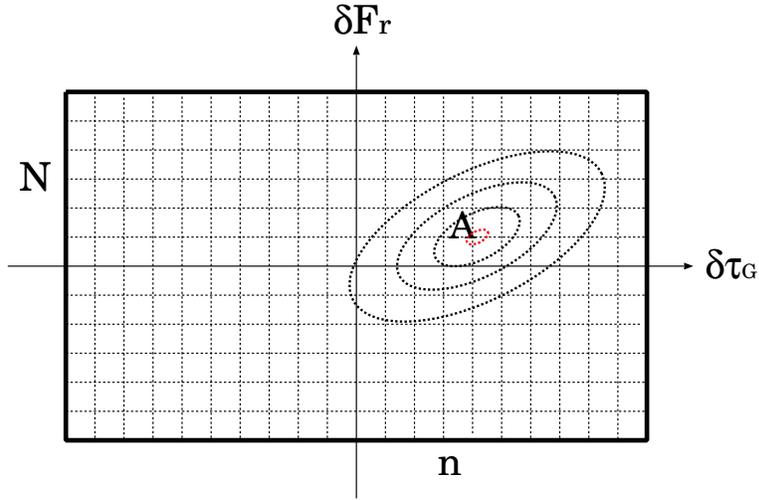


Figure 98: Peak of the search function on a grid point.

On the other hand, if the trial correction values are far from the actual residuals, the signal itself oscillates, both in time and frequency, and, therefore, is quickly depressed in the integration and averaging, and buried deeply under the noise.

Therefore, if the amplitude of the search function is markedly higher at, and around, a certain grid point A , than at other points, the true correlation peak must be found near this grid point (Figure 98).

4. Find correlation peak by ‘mountaineering method’.

The true correlation peak is further searched for, and eventually found, by a method called “mountaineering”. The trial correction values are varied little by little iteratively starting from the grid point A showing the highest amplitude, first in $\delta\tau_G$, then in δF_r and again in $\delta\tau_G$, \dots , searching for the points where the amplitude of the search function takes extremum values (B , C , D , E , \dots). This procedure is repeated until the iteration converges, and the correlation peak of the white fringe is finally found at $\delta F_r = \delta F_{r_m}$ and $\delta\tau_G = \delta\tau_{G_m}$ (Figure 99).

At the correlation peak, with correction values δF_{r_m} and $\delta\tau_{G_m}$, we obtain estimates of the residual fringe frequency:

$$\delta F_{r_m} \cong \Delta F_r,$$

the residual group delay:

$$\delta\tau_{G_m} \cong \Delta\tau_G,$$

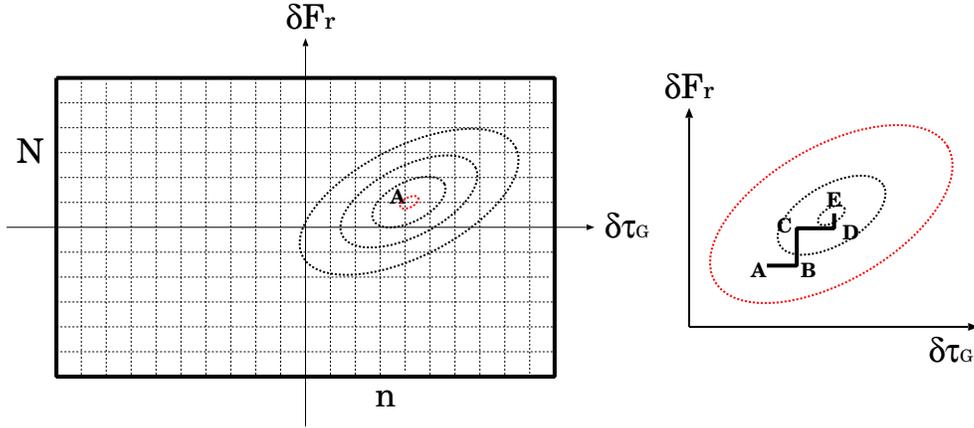


Figure 99: Finding the correlation peak by the ‘mountaineering method’.

the uncalibrated fringe amplitude:

$$|\mathcal{F}(\delta F_{r_m}, \delta \tau_{G_m})| \cong \mathcal{A}_U,$$

and the residual fringe phase:

$$-\arg[\mathcal{F}(\delta F_{r_m}, \delta \tau_{G_m})] \cong \Delta\Phi,$$

just like we saw in the discussion with the noise-free assumption. Of course, these estimates have only finite accuracies depending on the signal-to-noise ratio S/N of the correlation peak, when the noise exists. We will discuss the accuracy problem in the next section.

4.3.15 Correlation Peak of Continuum Source

Figure 100 shows a detected VLBI fringe of a continuum source observed on a baseline of the KSP (Keystone Project, 1993–2001) 4-station VLBI Array of CRL (now NICT) at X-band (8 GHz). Correlation processing was done with XF-type KSP correlator (Figure courtesy of NICT).

The Figure shows amplitude (vertical axis) of a search function which is made by combining data in 7 videoband channels of the K4 VLBI system over a search window (horizontal plane). Bandwidth of each videoband channel is 4 MHz. In the geodetic VLBI, fringe fitting based on single videoband channels is generally called the “coarse search”, as indicated in the Figure. This is distinguished from a “fine search” which uses full range of frequency distribution of videoband channels via the bandwidth synthesis technique.

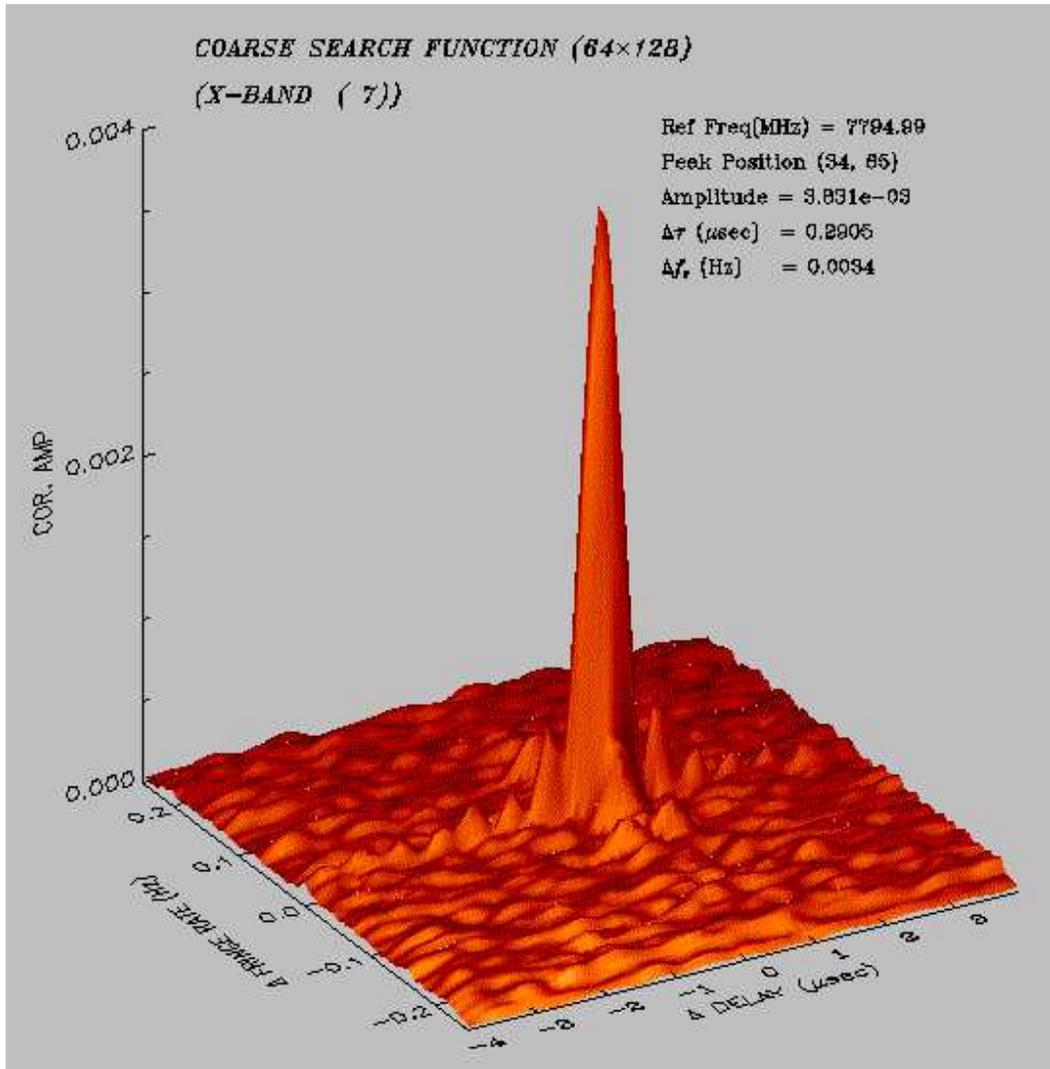


Figure 100: Detected fringe of a continuum source at 8 GHz. Axes in the horizontal plane show corrections for the group delay (labeled as “ Δ DELAY”) and physical fringe frequency (labeled as “ Δ FRINGE RATE”). Vertical axis shows amplitude of the search function. The correlation peak rises far above the “sea of noise” at a point (“peak position”) where the correction values are almost coincident with actual residual group delay and residual physical fringe frequency. Figure courtesy of NICT.

Axes in the horizontal plane show correction values for the group–delay $\delta\tau_G$ and physical fringe–frequency δf_r (or fringe–rate as labeled in the Figure). The search window covers 64 (group delay) \times 128 (fringe frequency) grid points. The correlation peak rises far above the “sea of noise” at 34–th in group delay and 65–th in fringe frequency grid (the “peak position”), where the correction values were closest to actual residual group delay and residual fringe frequency. Residual group delay and residual physical fringe frequency estimated through the “mountaineering” were 0.2905 μ sec and 0.0034 Hz, respectively.

Value of the amplitude of the search function at the correlation peak is shown as 0.003831 in this Figure. This is an estimate of the uncalibrated fringe amplitude, which corresponds to the amplitude of the complex cross–correlation coefficient after the DT and FS as equation (251) shows. Therefore, it is natural that the amplitude is smaller than 1.

Hereafter, we will denote the amplitude derived from the fringe fitting, which is an estimate of the uncalibrated fringe amplitude, as \mathbf{a} :

$$\mathbf{a} = | \mathcal{F}(\delta F_{r_m}, \delta\tau_{G_m}) | \cong \mathcal{A}_U. \quad (304)$$

4.3.16 Estimation of S/N of Detected Fringe

Now we derive a general formula for signal–to–noise ratio (S/N) of the detected fringe and estimate it using the amplitude \mathbf{a} obtained through the fringe fitting.

We saw in Subsection 1.1.28 that the S/N of a digital correlator output is given by

$$S/N = \left| \frac{r_{\hat{x}\hat{y}}[0]}{r_{xy}[0]} \right| |\rho| \sqrt{2B\tau_a},$$

where $|\rho|$ is the maximum amplitude of cross–correlation coefficient of analog voltages, $|r_{\hat{x}\hat{y}}[0]/r_{xy}[0]|$ is the amplitude ratio of cross–correlation coefficients of digital and original analog voltages, B is the bandwidth, and τ_a is the integration time (see equations (124) and (125)).

This equation should be directly applicable to a hardware correlator output if the DT and FS at the central lag happened to be perfect. In such a case, we would have $|r_{\hat{x}\hat{y}}[0]/r_{xy}[0]| = \eta_{c_Q} \eta_{c_D}$ in view of equation (251), and $\tau_a = t_{AP}$, where t_{AP} is the accumulation period (AP).

However, VLBI fringe is usually detected through the second step of the correlation processing, that is the fringe fitting. After the fringe fitting, τ_A increases to $N t_{AP}$, where N is the number of processed AP’s, and the amplitude ratio of the cross–correlation coefficients of digital and original (i.e. ideal loss–less) analog voltages is further reduced by $\eta_{c_P}(\tau_a) \eta_{c_M}(\tau_a)$

and becomes $\eta_{c_P}(\tau_a) \eta_{c_M}(\tau_a) \eta_{c_Q} \eta_{c_D} = \eta_c$. Also, after the fringe fitting, the maximum amplitude of the cross-correlation coefficient of analog voltages is given by

$$|\rho| = \frac{|\mathcal{R}_{xy}(0)|}{\sqrt{R_{xx}(0) R_{yy}(0)}}, \quad (305)$$

where $|\mathcal{R}_{xy}(0)|$ is the amplitude of the cross-correlation of analog voltages at the correlation peak, and $R_{xx}(0)$ and $R_{yy}(0)$ are powers or dispersions of the analog voltages.

On the other hand, we showed in equation (62) that the maximum amplitude $|\rho|$ is approximately given by

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

for a continuum spectrum source, where T_{A_i} and T_{S_i} , for $i = 1, 2$, are antenna temperature and system-noise temperature, respectively, of i -th element antenna.

Then, we obtain an equation for the S/N of the VLBI fringe:

$$S/N = \eta_c |\rho| \sqrt{2B\tau_a} = \eta_c \sqrt{\frac{T_{A_1} T_{A_2}}{T_{S_1} T_{S_2}}} \sqrt{2B\tau_a}. \quad (306)$$

Note that this equation has the general form which we gave in equation (133). We can use this equation to theoretically calculate an expected S/N value according to particular conditions of observation.

Now, in view of equations (296) and (305), we know that the uncalibrated fringe amplitude \mathcal{A}_U is given by

$$\mathcal{A}_U = \eta_c \frac{|\mathcal{R}_{xy}(0)|}{\sqrt{R_{xx}(0) R_{yy}(0)}} = \eta_c |\rho|. \quad (307)$$

Therefore, we obtain another form of the S/N :

$$S/N = \mathcal{A}_U \sqrt{2B\tau_a}. \quad (308)$$

Hence, we can estimate the S/N of the detected fringe using the derived amplitude \mathbf{a} by an equation:

$$S/N \cong \mathbf{a} \sqrt{2B\tau_a}, \quad (309)$$

since \mathbf{a} is an estimate of the uncalibrated fringe amplitude obtained through the fringe fitting. Fringe fitting softwares used in VLBI correlation processing often give S/N values of detected fringes calculated by equation (309).

Thus, if $B = 4$ MHz, $\tau_a = 120$ sec, and $\mathbf{a} = 0.003831$, then equation (309) yields $S/N \cong 119$.

4.3.17 Improvement of S/N through Integration

Figure 102 shows bird’s eye views of VLBI fringes detected with different integration times for a continuum source, which was observed on a baseline of the KSP Array at 8 GHz (Figure courtesy of NICT).

Panels show how results of the fringe fitting vary with increasing integration time from left–top to right–bottom. In each panel, amplitude of the search function is plotted for many points of trial correction values on the group–delay and fringe–frequency plane (the search window), just like in Figure 100. Correlation peak rises above the “sea of noise” at points where correction values are close to the residual group delay and residual fringe frequency.

We can see from the Figure that the peak amplitude \mathbf{a} is kept nearly constant while integration time is increased, except for first two or three panels, where the peak amplitude is severely affected by the still strong random noise. On the other hand, the noise is steadily suppressed, and therefore the S/N is markedly improved, with increasing integration time.

The constant peak amplitude means that the coherence factor η_c does not change with varying integration time, implying that the coherence factors $\eta_{cP}(\tau_a)$ and $\eta_{cM}(\tau_a)$ due to fringe–phase fluctuations and due to imperfect delay model, respectively, remain very close to 1 during all the integrations. Such a result is expected when both the frequency standards and the atmosphere are reasonably stable, and also the delay model is good enough, during the integration time.

However, this is not always the case in actual VLBI observations.

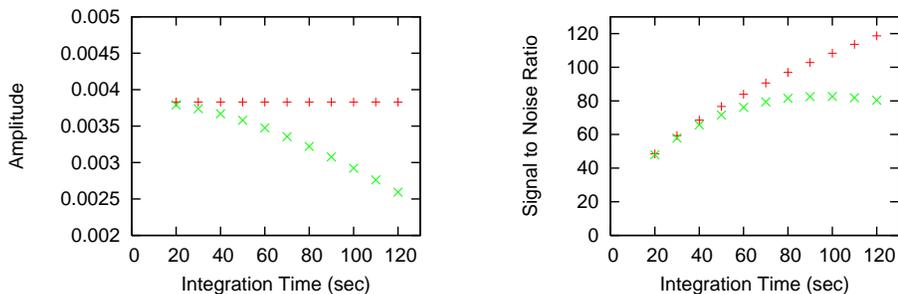


Figure 101: Schematic plots of peak amplitude (Left) and S/N (Right) derived from the fringe fitting as functions of integration time. (+) and (x) show cases without and with, respectively, a coherence loss increasing with integration time.

Therefore, it is highly recommended to check the coherence status at observing date before starting regular processing of large amount of observed

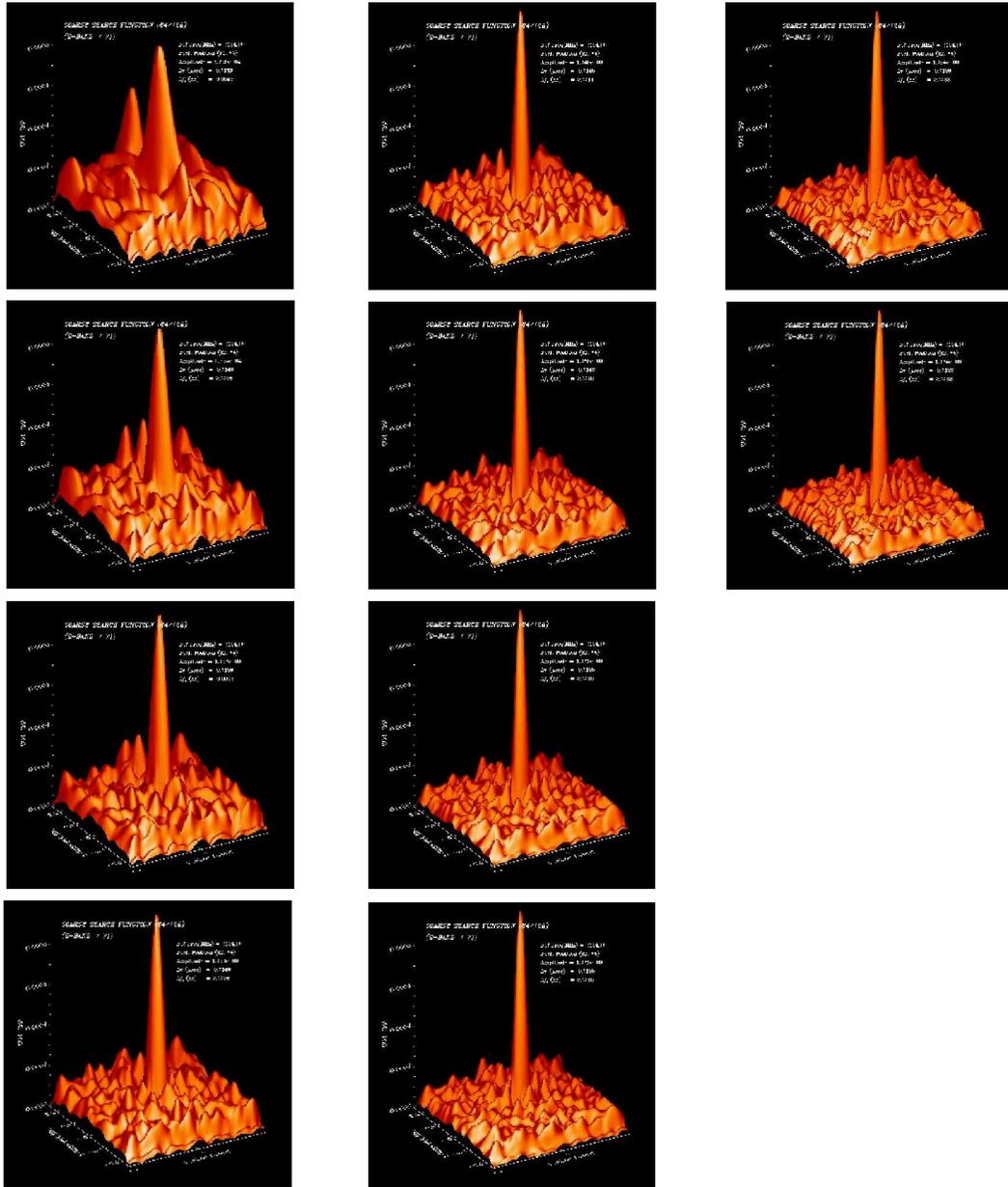


Figure 102: Amplitude of the search function on the group–delay and fringe–frequency plane, obtained from a VLBI observation. Panels show how signal–to–noise ratio of the correlation peak is improved with increasing integration time from left–top to right–bottom. Figure courtesy of NICT.

data. This can be done by repeating the fringe fitting for a bright source with varying integration time τ_a . Then, we can plot derived peak amplitude \mathbf{a} and/or S/N values against integration time τ_a as schematically illustrated in Figure 101.

If the coherence is well kept for a certain interval of time up to an expected coherence time, the derived amplitude \mathbf{a} should be nearly constant and the S/N should be improved in proportion to $\sqrt{\tau_a}$ with increasing integration time τ_a , as shown by (+) points in Figure 101. In such a case, we can safely proceed to the regular processing of rest of the data.

If the \mathbf{a} and S/N values fall down below the constant and $\sqrt{\tau_a}$ curve, respectively, as shown by (×) points in Figure 101, it is likely that actual coherence time is shorter than the expected. Then, we have to adopt shorter integration time.

If the \mathbf{a} and S/N values drop too fast, it is worth to suspect failure of the clock-LO system or large error in the delay model.

4.3.18 Examples of Detected Fringes for Continuum Sources

Figure 103 shows “historical” fringes detected with K3 correlator used from

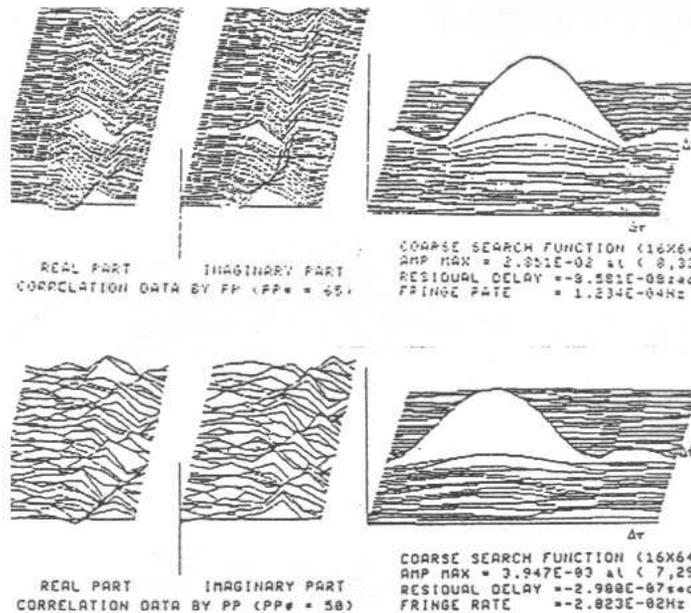


Figure 103: Fringes detected for 2 continuum spectrum sources (Top and Bottom) with 8-lag K3 correlator. Left 2 panels show real and imaginary cross-correlations. Right panel shows amplitude of the search function.

1980s to early 1990s which had only 8 lags.

Left 2 panels show outputs of the hardware correlator, i.e. the real and imaginary cross-correlations, respectively, as functions of lag-numbers and AP-numbers. On the other hand, right panel shows amplitude of the search function as a function of correction values for the group delay and fringe frequency, just as in Figure 100. Top panels are for a strong continuum spectrum source with derived peak amplitude α of 0.029, while Bottom panels are for a less strong source with derived peak amplitude of 0.0039.

In the real and imaginary cross-correlations of the strong source (Top, Left), we clearly see the characteristic patterns expected from the signal part of the cross-correlations, as we saw in Figure 86. On the other hand, such characteristic patterns are hardly seen in the Bottom-Left panels, due to the weaker signal level compared with the noise. After the integration through the fringe fitting, however, correlation peaks are well detected in both of the cases (Top and Bottom, Right).

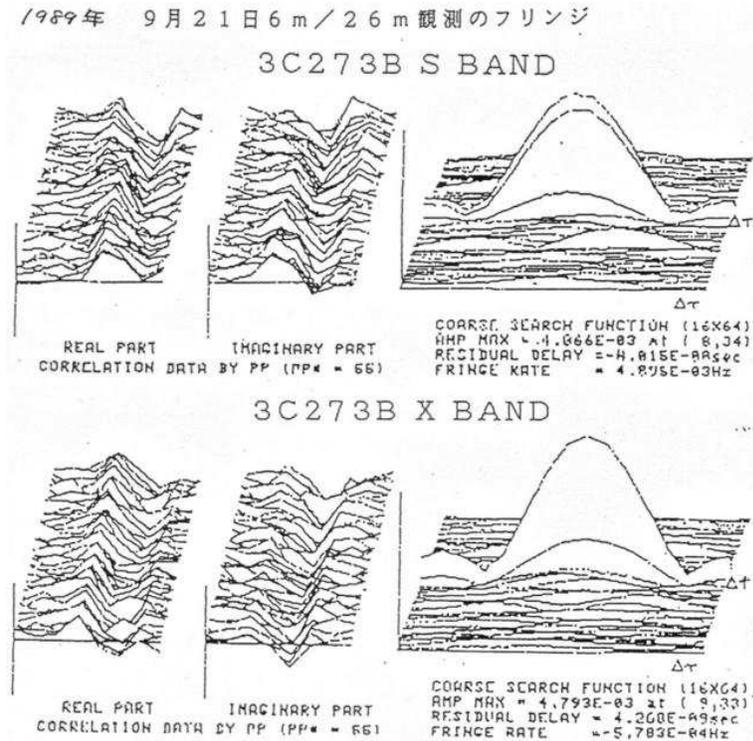


Figure 104: S-band and X-band fringes of 3C273B detected with K3 correlator. Observation was made on Kashima 26m - Nobeyama 6m baseline.

Figure 104 is another example of the result obtained with K3 correlator

which shows S-band and X-band fringes from a strong quasar 3C273B observed on a baseline connecting 26 m antenna in Kashima and 6 m antenna in Nobeyama.

Since 3C273B is very strong, we again see systematic patterns in the real and imaginary cross-correlations shown in Left panels of the Figure. In this case, we do not see any significant oscillations of the real and imaginary cross-correlations along the AP-number axis. Also, the maximum amplitudes of the cross-correlations appear nearly at the center of the delay window, i.e. at the central lag. This means that the residual group delay and residual fringe frequency were small, or otherwise speaking, the delay model given to the hardware correlator was fairly good. As a result, correlation peaks were detected nearly at the center of the search window in both S-band and X-band, as we actually see in Right panels.

Figure 105 shows 22 GHz fringe of 3C273B detected with 512-lag XF

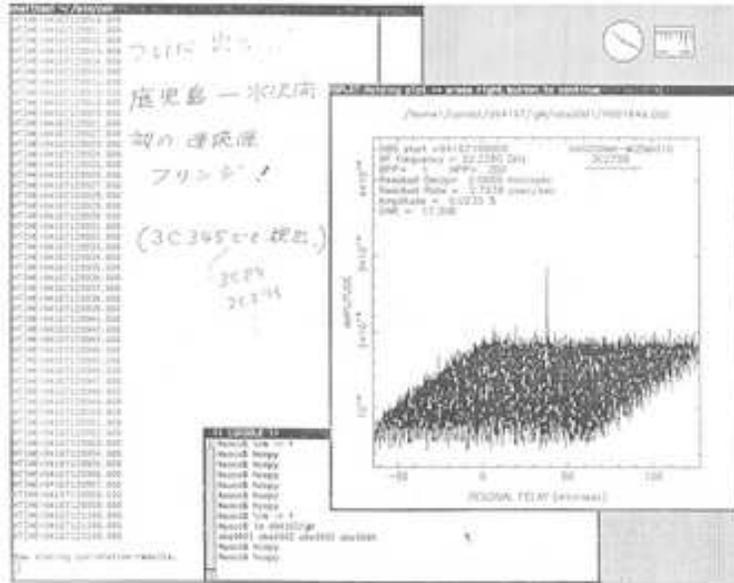


Figure 105: 22 GHz fringe of 3C273B on Mizusawa 10 m – Kagoshima 6 m baseline detected with 512-lag XF correlator NAOCO.

correlator NAOCO (New Advanced One-unit Correlator) used in 1990s. Observation was made on a 1300 km baseline between 10 m antenna in Mizusawa and 6 m antenna in Kagoshima. The Figure shows amplitude of the search function for a videoband with 2 MHz bandwidth on the group-delay-correction and fringe-frequency-correction plane, which is the same as those shown in Figures 100, 103, and 104. However, the correlation peak looks

very thin as a needle because of the large number of lags (512) in NAOCO, which gives a very wide delay window.

4.3.19 Software Processing for Spectralline Sources

So far, we discussed the software processing (fringe fitting, fringe search) for continuum spectrum sources.

In case of spectralline sources, such as astronomical maser sources, spectra of source signals consist of narrow lines. Their linewidths are usually much narrower than observing bandwidths.

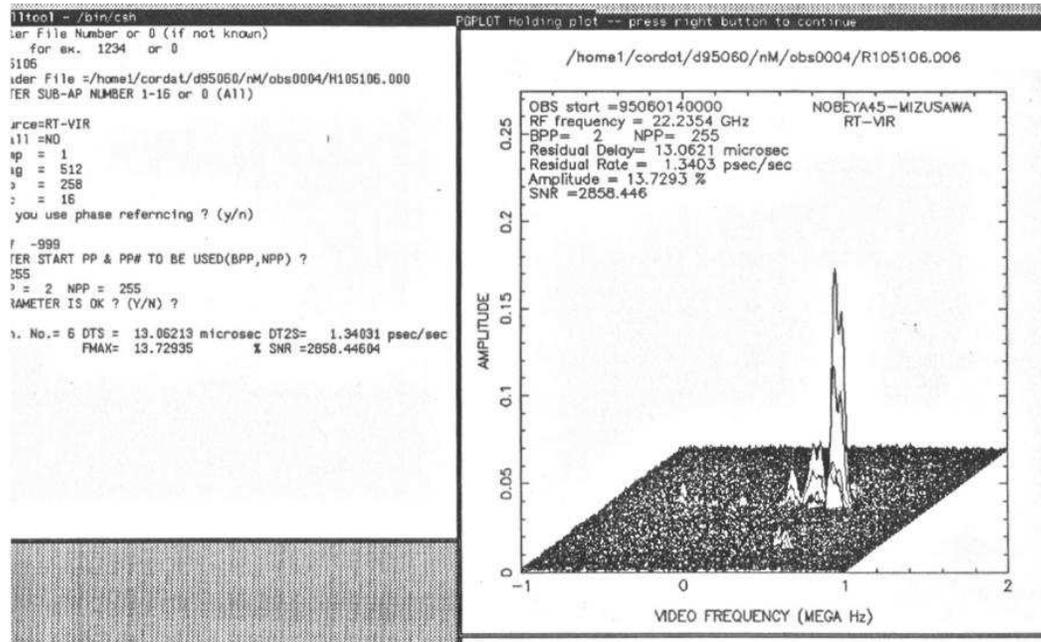


Figure 106: Fringe detected for a 22 GHz water maser source around an evolved star RT Vir observed on Nobeyama 45 m – Mizusawa 10 m baseline. Correlation processing was done with 512-lag XF correlator NAOCO. Two axes in the horizontal plane correspond to frequency and fringe–frequency correction, respectively. Vertical axis shows amplitude of the cross–power spectrum.

If we directly apply procedures of the software processing described above to such spectralline sources, amplitude of the search function as defined in equation (302) shows broad (in delay direction) peaks extending over wide ranges of the delay window. We can easily understand this by approximating the signal spectrum $\langle S_k(\omega) \rangle$ in equation (303) by a delta–function $\propto \delta(\omega - \omega_I)$. In this extreme case, the integration in equation (303) gives constant signal amplitude independent of the group–delay correction.

Therefore, naturally, we are not able to well estimate the residual group delay from fringes of spectralline sources. Wide bandwidth is essential for the group delay determination.

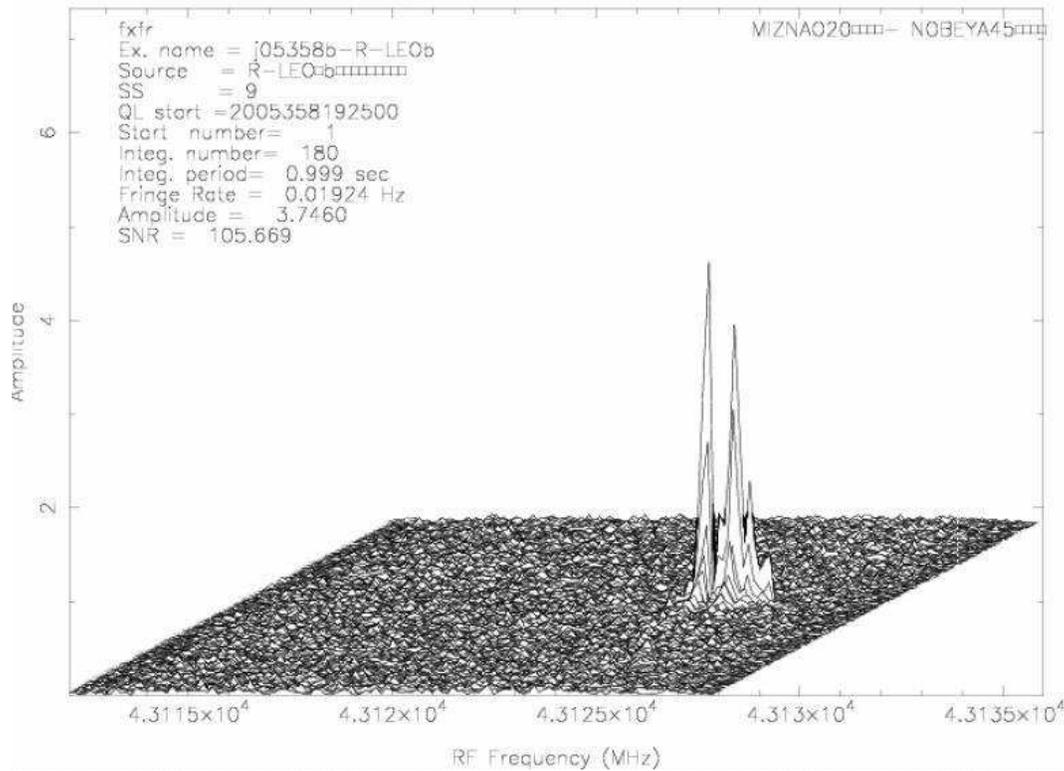


Figure 107: Fringe detected for a 43 GHz SiO maser source around an evolved star R Leo observed on Nobeyama 45 m – Mizusawa 20 m baseline in December 2005. Correlation processing was done with Mitaka FX correlator. Digitized Nobeyama data were transferred to Mitaka Correlation Center through an information network composed of broadband optical fiber cables.

On the other hand, the fact, that we can see the fringe of a spectralline source in a wide range of correction values for the group delay, implies that we do not need very accurate prediction of the group delay. In fact, a moderately good theoretical delay prediction which falls into the above wide range is sufficient to allow us to find the fringe.

Consequently, the residual group delay is not usually estimated, and the residual fringe frequency only is estimated, in the software processing of spectralline sources. Usually, the residual fringe frequencies at spectral peaks are searched on a frequency versus fringe–frequency–correction plane directly using the spectrum of the complex cross–correlation.

Figures 106 and 107 show examples of results of the “one–dimensional” search for fringes of maser sources. Note that two axes in horizontal planes of these Figures show frequency and fringe–frequency correction, respectively.

Maser sources usually consist of many compact maser features spread over small regions of the sky. Typical sizes of astronomical maser sources range from several to several tens of thousands of AUs (astronomical units), while sizes of individual maser features are a few AUs or smaller. Since the maser features are moving with high relative velocities of the order of $1 \sim 10$ km/sec, different features show different frequencies, which are offset from those measured in laboratories, due to the Doppler effect. Therefore, we see in these Figures many high and low spectral peaks at different frequencies which correspond to different maser features. Peaks often show different residual fringe frequency values since they are located at different positions on the sky.

4.3.20 If We Do Not Find Any Peak ...

If the theoretical prediction is accurate enough, the correlation peak should be found around the center of the search window, i.e. near $\delta F_r = 0$ and $\delta \tau_G = 0$. Such a case is usually a fortunate exception, however, in real VLBI observations, due mainly to a priori poorly known clock offset and clock rate offset. Although the frequency stability of hydrogen maser frequency standards is usually better than 10^{-15} in a timescale of around 1000 sec, it may drop to 10^{-13} level or worse in longer timescales of days or weeks. Therefore, it is not surprising if we encounter with a clock offset of μsec level and/or a clock rate offset of 10^{-12} sec/sec level. In such cases, peaks may be found at the edges of the search window or even may not be found at all!

Unfortunately, the worst case, when no peak is found within the search window, is not very rare. If this occurs, the first thing we have to do is to rerun the hardware correlator, changing the clock–offset and clock–rate–offset values in the correlator model. Usually, we have to repeat such naive search for several times. This process is often called the “clock search”.

At the same time, we have to carefully check if there was any mistake in the observation itself. Frequent mistakes include:

- large errors in source positions (including wrong epoch of equinox),
- large errors in station coordinates,
- wrong start and end times of recordings,
- wrong polarization,

- wrong bandwidth,
- wrong LO frequencies of VCs (BBCs),
- wrong hardware wirings,
- failure of frequency standard and/or phase-locked oscillator.

If one of the above is the case, it is almost hopeless to save the observation. This is typically the hardest time which every VLBI observer must confront.

If the first correlation peak is successfully found for a source in an experiment, then peaks of other sources are usually easily found, by applying the clock-offset and clock-rate-offset parameters estimated for the first source. Therefore, it is a usual practice in VLBI observations to include several bright sources with well-known positions in observing schedules, in order to use them as “fringe-finders”.

4.3.21 Threshold of Fringe Detection

Since the signal to noise ratio S/N is a kind of statistical concept, there is al-

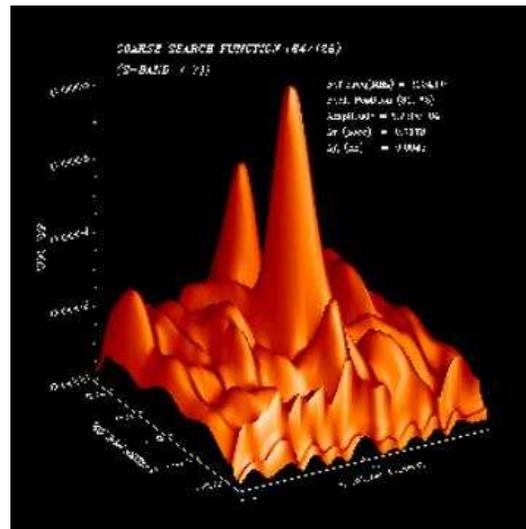


Figure 108: Which one is the true peak? Figure courtesy of NICT.

ways some possibility that amplitude of the noise exceeds by chance amplitude of the signal, even when S/N is larger than 1. In the VLBI software processing, we have to search for the fringe at many points of the search window (number of lags $n \times$ number of APs, or spectral channels, N). Therefore,

we have larger chance for a noise at some point exceeds the signal by chance. In fact, it is not easy to tell which one is the true correlation peak among noise-induced “peaks” when the signal to noise ratio is low (Figure 108).

In order to avoid misidentification of such a noise as the signal, the signal must be well larger than the noise level.

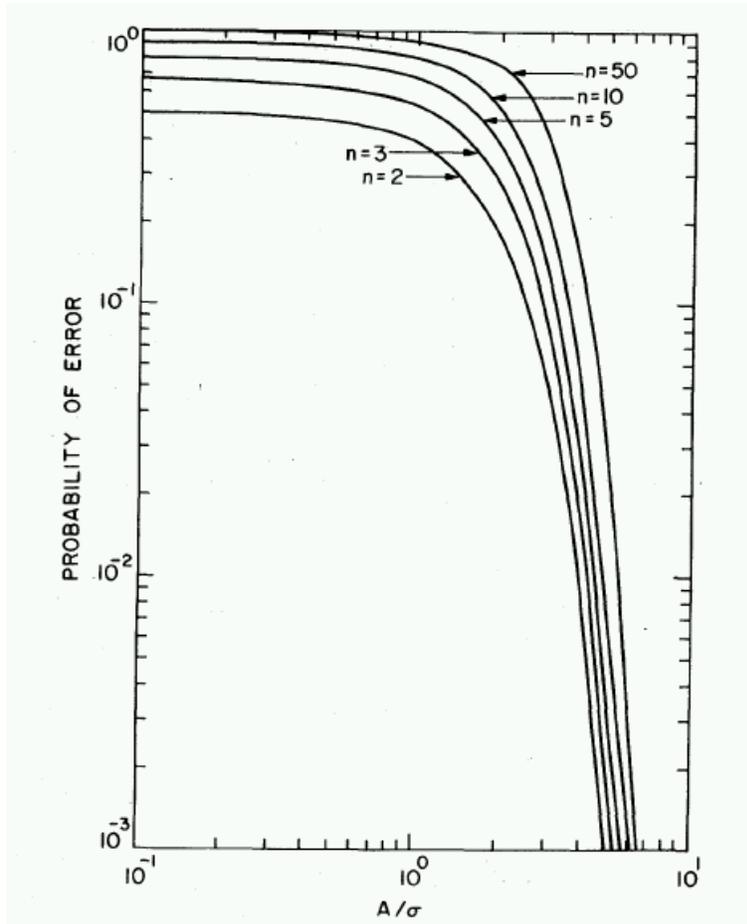


Figure 109: Probability that noise amplitude at some sample point will exceed by chance the signal amplitude, versus signal to noise ratio \mathcal{A}/σ . The curves are labeled according to the total number of samples. Figure brought from Moran (1976).

J.M. Moran (1976; see also Thompson, Moran, and Swenson, 2001) statistically investigated this problem and clarified that signal to noise ratio $S/N = \mathcal{A}/\sigma = 6 \sim 7$, where \mathcal{A} is the signal amplitude and σ is the standard deviation of the noise after the integration, is necessary for the probability of the misidentification to be smaller than 10^{-3} , if the number of search points (or samples) is around 50.

Figure 109, which is brought from Moran (1976), shows “probability of error”, i.e. the probability that one or more samples of amplitude of the search function among the samples with no signal will exceed by chance the amplitude of the sample with the signal, versus the signal to noise ratio \mathcal{A}/σ . The curves are labeled according to the total number n of samples, which corresponds to the number of independent samples in the search window $n \times N$ in our discussion. It is evident from this Figure that we need S/N of 6 or more in order to be sure that the probability of the misidentification is less than 10^{-3} , if the total number of samples is 50.

Therefore, it is natural to adopt $S/N = 6 \sim 7$ as a **threshold for fringe detection in VLBI**. This is the reason why we put $\mathcal{S}/\mathcal{N} = 7$ in our calculation of the minimum detectable flux density in VLBI.

5 Observables of VLBI

5.1 Four Basic Observables

5.1.1 Direct Products of Correlation Processing

In the course of the software processing (fringe fitting, or fringe search) described above, we obtained correction values for the fringe frequency δF_{r_m} and the group delay $\delta \tau_{G_m}$, respectively, which make the amplitude of the search function:

$$\begin{aligned} & \mathcal{F}(\delta F_r, \delta \tau_G) \\ &= \frac{1}{2\pi N} \sum_{k=1}^N \int_0^{2\pi B} S_k(\omega) e^{i[\delta F_r(t-t_k) + \delta \tau_G(\omega - \omega_I)]} d\omega, \end{aligned}$$

maximum. Then, at the peak of the amplitude of the search function, where $\delta F_r = \delta F_{r_m}$ and $\delta \tau_G = \delta \tau_{G_m}$, we obtained estimates

- δF_{r_m} for the residual fringe frequency ΔF_r , i.e.

$$\delta F_{r_m} \cong \Delta F_r, \quad (310)$$

- $\delta \tau_{G_m}$ for the residual group delay $\Delta \tau_G$, i.e.

$$\delta \tau_{G_m} \cong \Delta \tau_G, \quad (311)$$

- $|\mathcal{F}(\delta F_{r_m}, \delta \tau_{G_m})|$ for the uncalibrated fringe amplitude \mathcal{A}_U , i.e.

$$|\mathcal{F}(\delta F_{r_m}, \delta \tau_{G_m})| \cong \mathcal{A}_U, \quad (312)$$

- $-\arg[\mathcal{F}(\delta F_{rm}, \delta\tau_{Gm})]$ for the residual fringe phase $\Delta\Phi$, i.e.

$$-\arg[\mathcal{F}(\delta F_{rm}, \delta\tau_{Gm})] \cong \Delta\Phi, \quad (313)$$

at the band center $\omega = \omega_I$ and at the center of the observing time $t = t_0$.

These are direct products of the VLBI correlation processing which we obtain through the fringe fitting.

5.1.2 Calibration of Fringe Amplitude

The uncalibrated fringe amplitude \mathcal{A}_U is related to the fringe amplitude \mathcal{A} by equations (296) and (295):

$$\mathcal{A}_U = \mathcal{C}_C \mathcal{A}, \quad \text{with} \quad \mathcal{C}_C = \frac{\eta_c}{\sqrt{R_{xx}(0) R_{yy}(0)}},$$

where η_c is the total coherence factor and $R_{xx}(0)$ and $R_{yy}(0)$ are powers or dispersions of analog voltages just before the A/D conversion. The total coherence factor η_c is composed of four components

$$\eta_c = \eta_{c_Q} \eta_{c_D} \eta_{c_P} \eta_{c_M},$$

as we saw in equation (293), where η_{c_Q} , η_{c_D} , η_{c_P} , and η_{c_M} are coherence factors due to quantization, digital processing, fringe-phase fluctuation, and imperfect delay model, correspondingly.

Usually, we can readily calculate the coefficient \mathcal{C}_C . In fact, η_{c_Q} and η_{c_D} are just known constant factors inherent in the hardware correlator we used, while η_{c_P} and η_{c_M} must be very close to 1 as far as the integration time is properly selected and the theoretical delay model is good enough. On the other hand, powers $R_{xx}(0)$ and $R_{yy}(0)$ are regularly measured in VLBI stations for measuring the system noise temperatures. As we saw in Chapters 2 and 3, they are approximately given by

$$R_{ii}(0) = k T_{S_i} G_i B \quad (\text{for } i = x, y),$$

where k is the Boltzmann constant, T_{S_i} is the system noise temperature of i -th station, G_i is the gain factor in i -th station, and B is the bandwidth, correspondingly.

Using the coefficient \mathcal{C}_C thus calculated, we can calibrate the maximum amplitude of the search function $|\mathcal{F}(\delta F_{rm}, \delta\tau_{Gm})|$ to get an estimate of the fringe amplitude \mathcal{A} . For this purpose, we can just divide it by \mathcal{C}_C , i.e.

$$\frac{|\mathcal{F}(\delta F_{rm}, \delta\tau_{Gm})|}{\mathcal{C}_C} \cong \mathcal{A}. \quad (314)$$

5.1.3 Estimates and Expected Forms for Observables

The direct products of correlation processing contained residual values. Now we add back to them theoretically predicted values used in the hardware processing at $t = t_0$, i.e. $\omega_0 \dot{\tau}_i$ for the fringe frequency, τ_i for the group delay, and $\omega_0 \tau_i$ for the fringe phase, correspondingly, where ω_0 is the RF bandcenter frequency. In the USB reception case, ω_0 is equal to $\omega_0 = \omega_{LO} + \omega_I$, where ω_{LO} is the local oscillator frequency and ω_I is the IF (videoband) center frequency.

Then, from equations (310)–(314), we obtain following estimates and their expected forms. For the expected forms we use equations (291), (292), (297), and (290) which we derived in the noise-free case.

- $\delta F_{rm} + \omega_0 \dot{\tau}_i \cong F_r$ for the fringe frequency F_r , which has an expected form:

$$F_r = \omega_0 (\dot{\tau}_{g_0} + \dot{\tau}_C + \dot{\tau}_A) + \dot{\theta}_1 - \dot{\theta}_2 - \dot{\Phi}_v - \dot{\Phi}_b, \quad (315)$$

- $\delta \tau_{Gm} + \tau_i \cong \tau_G$ for the group delay τ_G , which has an expected form:

$$\tau_G = \tau_{g_0} + \tau_C + \tau_A - \frac{\partial \Phi_v}{\partial \omega} - \frac{\partial \Phi_b}{\partial \omega}, \quad (316)$$

- $|\mathcal{F}(\delta F_{rm}, \delta \tau_{Gm})| / \mathcal{C}_c \cong \mathcal{A}$ for the fringe amplitude \mathcal{A} , which has an expected form:

$$\mathcal{A} = \frac{A_0}{4\pi} \int_0^{2\pi B} |\mathcal{V}(\omega_{LO} + \omega)| |H_1(\omega) H_2^*(\omega)| d\omega, \quad (317)$$

- $-\arg[\mathcal{F}(\delta F_{rm}, \delta \tau_{Gm})] + \omega_0 \tau_i \cong \Phi(\omega_I, t_0)$ for the fringe phase $\Phi(\omega_I, t_0)$, which has an expected form:

$$\Phi(\omega_I, t_0) = \omega_0 (\tau_{g_0} + \tau_C + \tau_A) + \theta_1 - \theta_2 - \Phi_v - \Phi_b + 2\pi m. \quad (318)$$

Notations here are the same with those used in previous equations, i.e. $\Delta \tau_g = \tau_{g_0} - \tau_i$ is the residual geometric delay, $\tau_{g_0} = \mathbf{D} \cdot \mathbf{s}_0 / c$ is the geometric delay at a reference direction \mathbf{s}_0 in a radio source observed with an interferometer of baseline vector \mathbf{D} , τ_i is the theoretical prediction of the geometric delay used in the DT and FS (“the instrumental delay”), τ_C is the clock offset, τ_A is the atmospheric delay, $\theta_1 - \theta_2$ is the difference of the LO initial phases including the effect of the clock synchronization error between two stations as we defined in equation (216), $\Phi_v(\omega, t)$ is the phase spectrum of the complex visibility $\mathcal{V}(\omega_{LO} + \omega)$, $\Phi_b(\omega, t)$ is the phase spectrum of the “bandpass characteristics of combined IF filters” $H_1(\omega) H_2^*(\omega)$, B is the

bandwidth, A_0 is the geometric mean of effective apertures of two antennas, and $2\pi m$ stands for the cycle ambiguity.

These estimates for the fringe frequency, the group delay, the fringe amplitude, and the fringe phase are **four basic observables of the VLBI**.

5.2 Accuracies of VLBI Observables

The four observables of the VLBI, which we derived from the correlator outputs, contain noise contributions, besides the signals which have the expected forms given in equations (315) to (318). Now we will estimate accuracies of these observables.

5.2.1 Statistics of Amplitude and Phase

J.M. Moran developed a general statistical theory of amplitude and phase in his 1976 paper (Moran, 1976) which became the cornerstone for accuracy estimation of the VLBI observables. The threshold of the fringe detection, which we saw in Subsection 4.3.21, was derived by Moran in the same paper.

Moran considered a complex observed quantity:

$$\mathbf{Z} = \mathbf{A} + \boldsymbol{\varepsilon}, \quad (319)$$

consisting of the signal \mathbf{A} and the noise $\boldsymbol{\varepsilon}$. In our case, we regard as the signal \mathbf{A} the “complex fringe” with the fringe amplitude \mathcal{A} and the fringe phase Φ , i.e.

$$\mathbf{A} = \mathcal{A} e^{-i\Phi}. \quad (320)$$

To be definite, Moran considered the case when the real axis is chosen along the signal \mathbf{A} . Then, the fringe phase is zero ($\Phi = 0$), and the real component of \mathbf{Z} contains the signal \mathcal{A} and the noise ε_x , while the imaginary component contains only the noise ε_y :

$$\begin{aligned} Z_x &= \mathcal{A} + \varepsilon_x, \\ Z_y &= \varepsilon_y. \end{aligned} \quad (321)$$

Hereafter, we will denote the real and imaginary components with indices x and y , respectively. The noise components ε_x and ε_y are assumed to be mutually independent zero-mean Gaussian random variables of equal dispersions:

$$\langle \varepsilon_x^2 \rangle = \langle \varepsilon_y^2 \rangle = \sigma^2, \quad (322)$$

with the joint probability density:

$$f(\varepsilon_x, \varepsilon_y) = \frac{1}{2\pi\sigma^2} e^{-\frac{\varepsilon_x^2 + \varepsilon_y^2}{2\sigma^2}}. \quad (323)$$

We can easily derive the joint probability density of the real and imaginary components of the observed quantity \mathbf{Z} from equations (321) and (323):

$$f(Z_x, Z_y) = \frac{1}{2\pi\sigma^2} e^{-\frac{(Z_x - \mathcal{A})^2 + Z_y^2}{2\sigma^2}}. \quad (324)$$

The signal to noise ratio of the observed quantity is given by $S/N = \mathcal{A}/\sigma$.

Introducing amplitude Z and phase ϕ of the observed quantity \mathbf{Z} , we have

$$\mathbf{Z} = Z e^{-i\phi}, \quad (325)$$

where

$$Z = \sqrt{(\mathcal{A} + \varepsilon_x)^2 + \varepsilon_y^2}, \quad \text{and} \quad \phi = -\arctan\left(\frac{\varepsilon_y}{\mathcal{A} + \varepsilon_x}\right). \quad (326)$$

We call Z and ϕ the “observed fringe amplitude” and “observed fringe phase”, or just the “observed amplitude” and “observed phase”, respectively (Figure 110).

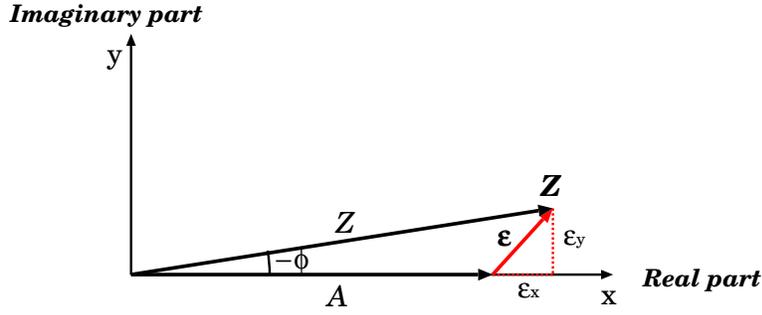


Figure 110: Amplitude Z and phase ϕ of the observed quantity \mathbf{Z} containing the real signal of amplitude \mathcal{A} and the noise ε .

We can derive probability density $f(Z)$ of the observed amplitude Z and probability density $f(\phi)$ of the observed phase ϕ from the joint probability density $f(Z_x, Z_y)$ in equation (324). The results are

$$f(Z) = \frac{Z}{\sigma^2} e^{-\frac{Z^2 + \mathcal{A}^2}{2\sigma^2}} I_0\left(\frac{\mathcal{A}Z}{\sigma^2}\right), \quad Z > 0, \quad (327)$$

$$f(\phi) = \frac{1}{2\pi} e^{-\frac{\mathcal{A}^2}{2\sigma^2}} \left\{ 1 + \sqrt{\frac{\pi}{2}} \frac{\mathcal{A} \cos \phi}{\sigma} e^{\frac{\mathcal{A}^2 \cos^2 \phi}{2\sigma^2}} \left[1 + \operatorname{erf}\left(\frac{\mathcal{A} \cos \phi}{\sqrt{2}\sigma}\right) \right] \right\}, \quad (328)$$

where $I_n(x)$ is the modified Bessel function of n -th order:

$$I_n(x) = \frac{1}{2\pi} \int_0^{2\pi} e^{x \cos \theta} \cos(n\theta) d\theta, \quad (329)$$

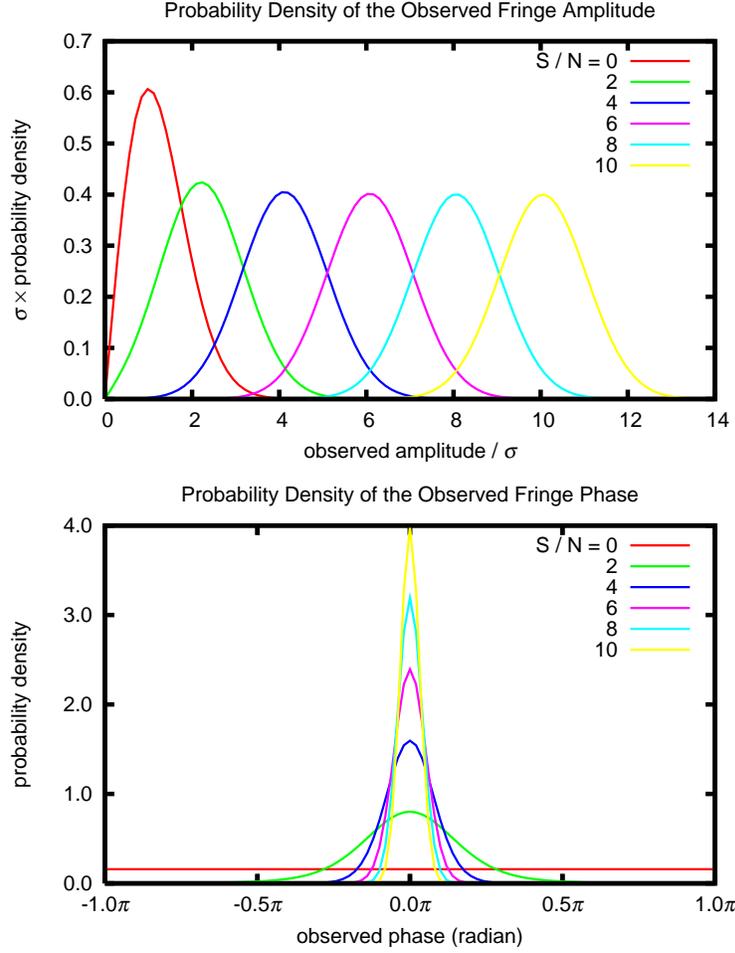


Figure 111: Standard deviation σ times probability density $f(Z)$ of the observed fringe amplitude Z versus Z/σ (top), and probability density $f(\phi)$ of the observed fringe phase ϕ versus ϕ . The curves are labeled according to the signal-to-noise ratio $S/N = \mathcal{A}/\sigma$.

for non-negative integer n , and $\text{erf}(x)$ is the error function defined by

$$\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-u^2} du, \quad (330)$$

with θ and u being arguments of integration (Abramowitz and Stegun, 1964). $f(Z)$ given in equation (327) is known as the Rice distribution.

One can verify these equations following, for example, discussions given in Papoulis (1984). Materials in the section titled “Two Functions of Two Random Variables” in Chapter 6 “Two Random Variables” of Papoulis (1984), especially discussions on “Joint Density” and examples therein, are of direct relevance to the present problem.

Upper and lower panels of Figure 111 show functional forms of the probability densities $f(Z)$ and $f(\phi)$ as given in equations (327) and (328).

When no signal is present (i.e., $\mathcal{A} = 0$), $f(Z)$ reduces to the Rayleigh probability density:

$$f(Z) = \frac{Z}{\sigma^2} e^{-\frac{Z^2}{2\sigma^2}}, \quad Z > 0,$$

and $f(\phi)$ becomes uniform (specifically, $f(\phi) = 1/2\pi$), as expected.

5.2.2 Expectations of Amplitude and Phase

Moran (1976) derived expectations of the observed amplitude and phase, using probability densities in equations (327) and (328). The results are

$$\begin{aligned} \langle Z \rangle &= \int_0^{\infty} Z f(Z) dZ = \frac{1}{\sigma^2} e^{-\frac{\mathcal{A}^2}{2\sigma^2}} \int_0^{\infty} e^{-\frac{Z^2}{2\sigma^2}} Z^2 I_0\left(\frac{\mathcal{A}Z^2}{4\sigma^2}\right) dZ \\ &= \sqrt{\frac{\pi}{2}} \sigma e^{-\frac{\mathcal{A}^2}{4\sigma^2}} \left[\left(1 + \frac{\mathcal{A}^2}{2\sigma^2}\right) I_0\left(\frac{\mathcal{A}^2}{4\sigma^2}\right) + \frac{\mathcal{A}^2}{2\sigma^2} I_1\left(\frac{\mathcal{A}^2}{4\sigma^2}\right) \right], \end{aligned} \quad (331)$$

$$\langle \phi \rangle = \int_{-\pi}^{\pi} \phi f(\phi) d\phi = 0. \quad (332)$$

One can verify equation (331) by using the recurrence formulae:

$$\frac{dI_0(x)}{dx} = I_1(x), \quad \text{and} \quad 2 \frac{dI_n(x)}{dx} = I_{n-1}(x) + I_{n+1}(x),$$

and also the integration formula:

$$\int_0^{\infty} e^{-a^2 x^2} I_{\nu}(bx) dx = \frac{\sqrt{\pi}}{2a} e^{\frac{b^2}{8a^2}} I_{\frac{1}{2}\nu}\left(\frac{b^2}{8a^2}\right),$$

for $\Re\nu > -1$ and $\Re a^2 > 0$, of the modified Bessel functions (Abramowitz and Stegun, 1964), and performing integration by parts twice.

On the other hand, equation (332) is rather trivial, since $f(\phi)$ in equation (328) is an even function of ϕ . Of course, this $\langle\phi\rangle = 0$ result is a consequence of our choice of the real axis along the signal \mathbf{A} . In a general case when the signal has non-zero fringe phase Φ , we must obtain

$$\langle\phi\rangle = \Phi. \quad (333)$$

We can easily calculate expectations of even-order power functions of the observed amplitude Z , by directly using the Gaussian probability density in equation (324). In particular, $\langle Z^2 \rangle$ is given by

$$\langle Z^2 \rangle = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (Z_x^2 + Z_y^2) f(Z_x, Z_y) dZ_x dZ_y = \mathcal{A}^2 + 2\sigma^2. \quad (334)$$

Using equation (334), we can calculate the standard deviation σ_Z of the observed amplitude:

$$\sigma_Z = \sqrt{\langle (Z - \langle Z \rangle)^2 \rangle} = \sqrt{\langle Z^2 \rangle - \langle Z \rangle^2} = \sqrt{\mathcal{A}^2 + 2\sigma^2 - \langle Z \rangle^2}, \quad (335)$$

where the expectation of the observed amplitude $\langle Z \rangle$ is given by equation (331).

5.2.3 In Limiting Case When S/N is Large

From Figure 111, we see that the probability densities of both the observed amplitude Z and phase ϕ take nearly Gaussian forms in the limit of large signal-to-noise ratio: $S/N = \mathcal{A}/\sigma \gg 1$.

In fact, using the first term of the asymptotic expansion of the modified Bessel function:

$$I_n(x) \cong \frac{e^x}{\sqrt{2\pi x}} \left[1 - \frac{m-1}{8x} + \frac{(m-1)(m-9)}{2(8x)^2} - \dots \right], \quad \text{for } x \gg n, \quad (336)$$

where n is any non-negative integer and $m = 4n^2$ (Abramowitz and Stegun, 1964), we obtain an approximate form of equation (327):

$$f(Z) \cong \frac{1}{\sqrt{2\pi}\sigma} \sqrt{\frac{Z}{\mathcal{A}}} e^{-\frac{(Z-\mathcal{A})^2}{2\sigma^2}}, \quad \text{when } \frac{\mathcal{A}}{\sigma} \gg 1, \quad (337)$$

for the probability density of the observed fringe amplitude Z . Of course, this approximate formula is not exactly Gaussian, in view of the extra factor

$\sqrt{Z/\mathcal{A}}$. However, when $\mathcal{A}/\sigma \gg 1$, Z is nearly constant: $Z \cong \mathcal{A}$, in the relatively narrow range of Z where $f(Z)$ is markedly non-zero, as Figure 111 shows. Thus, the probability density in equation (337) is nearly Gaussian in this sense.

Also, in view of the asymptotic behavior of the error function:

$$\operatorname{erf}(x) \cong 1, \quad \text{when } x \rightarrow \infty,$$

(Abramowitz and Stegun, 1964), as well as the fact that $f(\phi)$ is markedly non-zero only in the relatively narrow range of ϕ around zero: $|\phi| \ll 1$, as Figure 111 shows, where we can approximate the $\cos \phi$ function by

$$\cos \phi \cong 1 - \frac{1}{2} \phi^2,$$

we can reduce equation (328) to

$$f(\phi) \cong \frac{\mathcal{A}}{2\pi\sigma} e^{-\frac{\mathcal{A}^2 \phi^2}{2\sigma^2}}, \quad \text{when } \frac{\mathcal{A}}{\sigma} \gg 1. \quad (338)$$

This is the approximate formula at large signal-to-noise ratio for the probability density of the observed fringe phase ϕ . Note that the approximate equation (338) has the exact Gaussian form.

Applying equation (336) for the asymptotic expansion of the modified Bessel function to equations (331) and (335), we obtain approximate formulae for the expectation of the observed amplitude:

$$\langle Z \rangle \cong \mathcal{A} \left(1 + \frac{\sigma^2}{2\mathcal{A}^2} \right) = \mathcal{A} \left(1 + \frac{1}{2(S/N)^2} \right), \quad \text{for } \frac{\mathcal{A}}{\sigma} \gg 1, \quad (339)$$

and for the standard deviation σ_Z of the observed amplitude:

$$\frac{\sigma_Z}{\mathcal{A}} \cong \frac{\sigma}{\mathcal{A}} \left(1 - \frac{\sigma^2}{4\mathcal{A}^2} \right) = \frac{\sigma}{\mathcal{A}} \left(1 - \frac{1}{4(S/N)^2} \right), \quad \text{for } \frac{\mathcal{A}}{\sigma} \gg 1. \quad (340)$$

Figure 112 shows the expectation and the standard deviation of the observed amplitude Z normalized by the signal amplitude \mathcal{A} as functions of the signal-to-noise ratio $S/N = \mathcal{A}/\sigma$. Solid and broken lines compare results calculated by the exact formulae of equations (331) and (335) (solid) and by the approximate formulae of equations (339) and (340) (broken). We can see from the Figure that the approximate formulae are good enough as far as the signal-to-noise ratio is larger than the fringe detection threshold: $S/N \geq 6 \sim 7$.

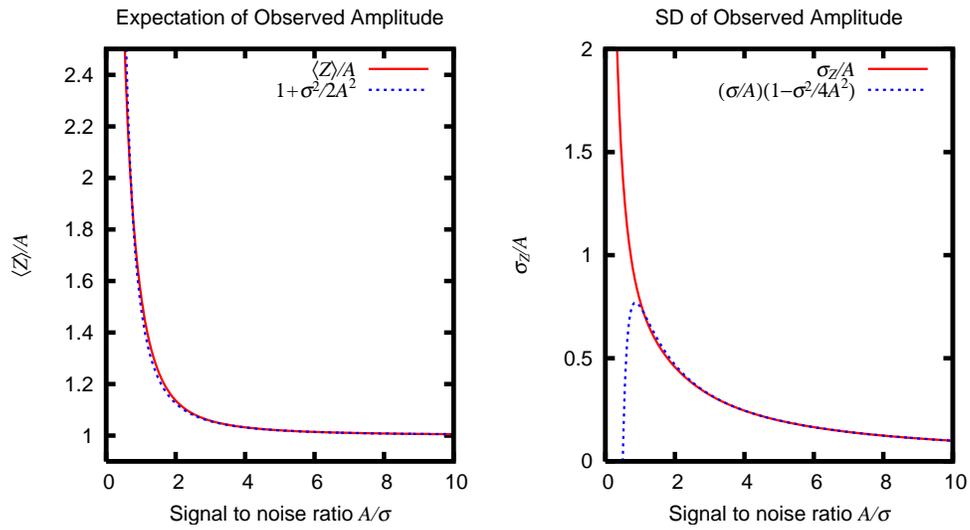


Figure 112: Expectation (left) and standard deviation (right) of observed amplitude Z both normalized by the signal amplitude \mathcal{A} versus signal to noise ratio \mathcal{A}/σ . Solid lines are calculated by exact formulae of equations (331) and (335), while broken lines are calculated by approximate formulae of equations (339) and (340).

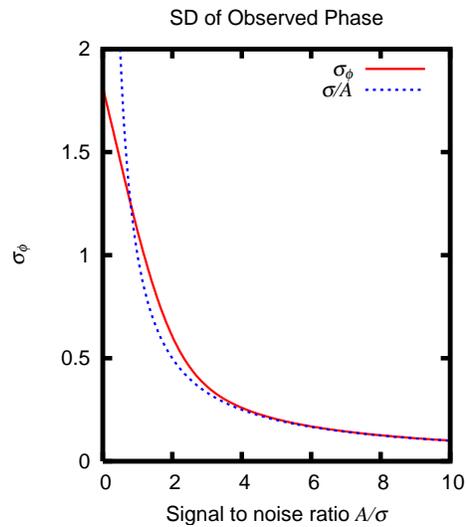


Figure 113: Standard deviation of observed phase ϕ versus signal to noise ratio \mathcal{A}/σ . Solid line is calculated by exact formula of equation (342), while broken line is calculated by approximate formula of equation (341).

Equation (338) for the approximate probability density of the observed phase at $S/N \gg 1$ gives the exact zero result for the expectation of the observed phase: $\langle \phi \rangle = 0$, and an approximate equation for the standard deviation of the observed phase:

$$\sigma_\phi \cong \frac{\sigma}{\mathcal{A}} = \frac{1}{S/N} \text{ (radian)}, \quad \text{for } \frac{\mathcal{A}}{\sigma} \gg 1. \quad (341)$$

We can compare this approximate result with the exact one:

$$\sigma_\phi = \sqrt{\left(\int_{-\pi}^{\pi} \phi^2 f(\phi) d\phi \right)}, \quad (342)$$

where $f(\phi)$ is the probability density of the observed phase given by equation (328). Figure 113 shows result of the comparison. We see again that the approximate formula of equation (341) is good enough when the signal-to-noise ratio is larger than the fringe detection threshold: $S/N \geq 6 \sim 7$. Note that $\sigma_\phi = \pi/\sqrt{3}$, when there is no signal ($\mathcal{A}/\sigma = 0$) and the probability density is uniform.

5.2.4 Accuracy of Fringe Amplitude Observable

The relative accuracy (or, more precisely, relative error due to the thermal noise) of the observed fringe amplitude with known signal-to-noise ratio $S/N = \mathcal{A}/\sigma$ is estimated by σ_Z/\mathcal{A} given by equation (335), or by equation (340) when $S/N \gg 1$.

Note that the estimation of the observed fringe amplitude is not strictly unbiased in the sense that the statistical expectation of the observed amplitude is always larger than the true signal amplitude due to the noise contribution, as evident from equation (339) or Figure 112. We must be careful about this effect when the signal-to-noise ratio is not sufficiently high: $S/N < 6 \sim 7$.

Naturally, the very long baseline of VLBI does not offer any advantage in the amplitude estimation. So, one must be satisfied by the usual accuracy of radio astronomy in general:

$$\frac{\sigma_Z}{\mathcal{A}} \cong \frac{1}{S/N}, \quad \text{when } S/N \gg 1.$$

The expected form of the fringe amplitude given in equation (317) does not contain information on the geometric delay τ_{g_0} . Therefore, the fringe

amplitude observable is not directly used in the geodetic VLBI. On the other hand, the visibility amplitude, which is obtained from the fringe amplitude after calibration of the bandpass characteristics, is the most important observable for the radio source imaging in the astrophysical VLBI.

5.2.5 Accuracy of Fringe Phase Observable

We can use the standard deviation of the observed fringe phase $\sigma_\phi \cong 1/(S/N)$ given in equation (341) as a measure of the accuracy (or thermal noise error) of the fringe phase observable.

We can easily understand this result if we look at an illustration in Figure 114, where we consider a general case when the signal has non-zero phase Φ . The Figure clearly explains why the expectation and the standard deviation of the observed fringe phase must be

$$\langle \phi \rangle = \Phi,$$

and

$$\sigma_\phi \cong \frac{1}{S/N} \quad (\text{radian}),$$

respectively, as we derived in equations (333) and (341).

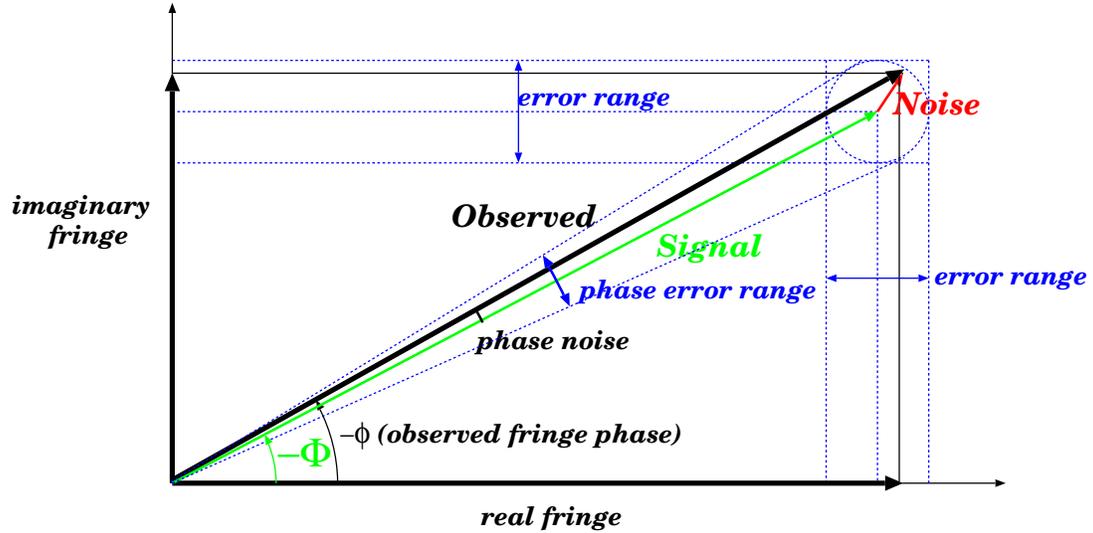


Figure 114: When real and imaginary parts of the complex fringe, which are obtained through the fringe fitting, have an equal noise standard deviation σ , and signal has amplitude \mathcal{A} , the standard deviation σ_ϕ of the observed phase is nearly equal to $\sigma/\mathcal{A} = 1/(S/N)$, as far as the signal to noise ratio S/N is large.

What does this fringe phase accuracy mean?

1. We can estimate the fringe phase Φ of the signal from an astronomical radio source with the $\sigma_\phi \cong 1/(S/N)$ radian accuracy. If we observe the source at a bandcenter frequency $\nu_0 = 22$ GHz with a baseline $D = 6000$ km, the minimum fringe spacing $\Delta\theta_F = c/(\nu_0 D)$ is around 0.48 milliarcsecond(mas). Therefore, if $S/N = 50$, then the $1/(S/N)$ phase accuracy corresponds to an angular distance:

$$\frac{\Delta\theta_F}{2\pi(S/N)} \cong 1.5 \text{ microarcsecond}(\mu\text{as})!$$

2. If we were allowed to uniquely determine the observed fringe phase ϕ , then we could in principle determine the “**phase delay**” τ_ϕ :

$$\tau_\phi = \frac{\phi}{\omega_0}, \quad (343)$$

with $\sigma_{\tau_\phi} \cong \sigma_\phi/\omega_0 \cong 0.14$ psec accuracy at 22 GHz, where $\omega_0 = 2\pi\nu_0$ is the RF bandcenter angular frequency.

3. If so, and if we could well calibrate the atmospheric, clock and rest of phase terms in equation (318) other than that due to the geometric delay $\omega_0 \tau_{g_0}$, then we could use the phase delay observable to estimate the radio source position or the Earth’s orientation with the μas level accuracy as well as the station position with $c\sigma_{\tau_\phi} \cong 0.04$ mm level accuracy in the geodetic VLBI.

Also, we could use the accurate fringe phase information for high-resolution radio source imaging in the astrophysical VLBI.

4. So, it is too regretful that we are not able to resolve the cycle ambiguity figuring in equation (318), as we discussed in Subsection 3.4.2, and therefore cannot directly use the fringe phase observable.
5. We must work hard to find ways to use, at least partially, the fringe phase observable. We will address this topic in the next Section.

5.2.6 Group Delay and Fringe Frequency Observables Free from Cycle Ambiguity

We already discussed in Subsection 3.4.2 the reason why we are not allowed to directly use the fringe phase observable. This was because we cannot uniquely know the fringe phase value due to the $2\pi n$ cycle ambiguity. The

unpredictable and time-varying atmospheric and clock effects are so large that they usually make it impossible to resolve the $2\pi n$ cycle ambiguity. Even if the correlation processing yields a phase value, for example $75.^{\circ}03$, say, the actual value could be $-284.^{\circ}97$, $435.^{\circ}03$, or some other $75.^{\circ}03 \pm n \times 360^{\circ}$.

On the other hand, the group delay and fringe frequency are derivatives of the phase spectrum of the cross-correlation of received voltages with respect to frequency and time, respectively. Therefore, the cycle ambiguity term, which is essentially constant both in frequency and time, does not give any contribution to these observables. In fact, the expected forms of the group delay and fringe frequency as given in equations (316) and (315):

$$\begin{aligned}\tau_G &= \tau_{g_0} + \tau_C + \tau_A - \frac{\partial\Phi_v}{\partial\omega} - \frac{\partial\Phi_b}{\partial\omega}, \\ F_r &= \omega_0(\dot{\tau}_{g_0} + \dot{\tau}_C + \dot{\tau}_A) + \dot{\theta}_1 - \dot{\theta}_2 - \dot{\Phi}_v - \dot{\Phi}_b,\end{aligned}$$

do not contain any ambiguity.

Therefore, we can directly use them in order to estimate the geometric delay τ_{g_0} for high precision VLBI geodesy and astrometry. Currently, they are the primary observables in the geodetic VLBI.

Of course, they are affected by the atmospheric and clock effects τ_A , τ_C and $\dot{\tau}_A$, $\dot{\tau}_C$ in equations (316) and (315) as severely as the fringe phase observable is. Unlike the fringe phase, however, the group delay and fringe frequency observables are uniquely determined without any ambiguity. In such a case, we can develop methods of analysis for estimating and removing, at least partially, the atmospheric and clock effects based on these observables themselves, which are obtained for many radio sources observed at different sky directions within hours or so. Such methods are widely used in analyses of current geodetic VLBI observations.

On the other hand, the group delay and fringe frequency observables are not directly relevant to the source imaging in the astrophysical VLBI except in cases where position information of the observed source is important.

5.2.7 Accuracy Estimation for Group Delay and Fringe Frequency

Thus, it is important for the VLBI geodesy and astrometry to know accuracies of the group delay and fringe frequency observables. This could be also important for VLBI astrophysicists as well, since they must be able to properly evaluate results of the fringe fitting often required in the imaging process.

However, the accuracy estimation of these observables appears a little complicated.

Of course, the estimation could be straightforward if we had sufficiently good data of the observed phase spectrum of the cross-correlation of the received voltages $\phi(\omega, t)$ as a function of frequency ω and time t . Then we could estimate the group delay $\partial\phi/\partial\omega$ and the fringe frequency $\partial\phi/\partial t$ together with their accuracies in terms of the standard least squares (or linear regression) analysis, i.e. by just fitting straight lines to [phase vs. frequency] and [phase vs. time] plots. For this purpose, it is natural to use the phase spectrum of the complex cross-correlation which is obtained from outputs of a multi-lag or multi-frequency-channel complex correlator at each AP (the accumulation period) as we saw in Subsections 4.3.4 and 4.3.14.

Unfortunately, however, such phase spectrum for a moderately weak source is usually quite noisy and easily scatter beyond 360 degrees. This is because hardware correlators tend to have large number of lags, or equivalently many frequency channels, such as 512 or 1024, and use short AP's around 1 second or shorter. In such a case, the signal in a component of the cross-power spectrum, which is contained within the small intervals of time and frequency, is usually too weak compared with the thermal noise. Consequently, the direct analysis of the phase spectrum is just meaningless, unless we had a magical power to exactly trace the $2\pi n$ jumps in the highly noisy phase values.

Instead, we usually obtain estimates of the residual group delay $\Delta\tau_G$ and the residual fringe frequency ΔF_r by the “peak-search” method in the fringe fitting, as we described in Subsection 4.3. In fact, we searched for optimal corrections $\delta\tau_G$ and δF_r which make the absolute value of the search function maximum and regarded them as the best estimates of $\Delta\tau_G$ and ΔF_r . The peak search method uses complex quantities without decomposing them into amplitudes and phases. Since the complex quantities are essentially continuous at any $2\pi n$ phase jumps, the method does not suffer from the phase jump problem and is able to reliably provide all necessary observables with high enough signal-to-noise ratio after integration both in time and frequency. Though this method looks very clever and quite reasonable, it does not seem easy to judge how accurate the obtained estimates are.

Thus it is important for us to estimate accuracies of the group delay and fringe frequency observables as obtained through the peak search method.

5.2.8 Equivalence of Peak Search Method and Least Squares Estimation

Actually, estimation of the group delay and the fringe frequency with the peak search method is equivalent to the least squares estimation of phase slopes in time and frequency, as far as signal-to-noise ratio of the final result

(integrated fringe) is sufficiently high. In the followings, we will show the equivalence, partly based on a discussion given in Appendix of PhD Thesis by Mamoru Sekido (Sekido, 2001).

Search Function with Discrete Cross-Power Spectrum

In order to see the equivalence, it is rather convenient to use the discrete Fourier transformation at finite number of sample points, instead of the continuous one which we have used for conceptual familiarity.

Let us consider outputs of a multi-lag hardware correlator, i.e. the complex cross-correlations given at n equally spaced lags and N equally spaced times (accumulation periods). We denoted the complex cross-correlation at l -th lag and k -th time as $\mathcal{R}_k(\Delta\tau_l)$ in Subsection 4.2.13. Again we assume the case when the number of lags n is an even number. The additional delay introduced by the n -lag correlator $\Delta\tau_l$ with $l = 1, 2, \dots, n/2, \dots, n$ is chosen to be zero at the central lag number $l = n/2$, i.e., $\Delta\tau_l = (l - n/2)t_s$, as before, where t_s is the sampling interval. We assume the Nyquist sampling with the sampling interval $t_s = 1/(2B)$, where B is the frequency bandwidth of our observation. We describe the additional delay $\Delta\tau_l$ also as $\Delta\tau_m = m t_s$, introducing the new lag number $m = l - n/2$, as we did in Subsection 4.2.9.

The discrete Fourier transform $S_k(\omega_j)$ of the complex cross-correlation $\mathcal{R}_k(\Delta\tau_l)$ given at n discrete points (lags) is given by

$$S_k(\omega_j) = \sum_{l=1}^n \mathcal{R}_k(\Delta\tau_l) e^{-i\omega_j \Delta\tau_l t_s},$$

or, equivalently,

$$S_k(\omega_j) = \sum_{m=-\frac{n}{2}+1}^{\frac{n}{2}} \mathcal{R}_k(\Delta\tau_m) e^{-i\omega_j \Delta\tau_m t_s} = \sum_{m=-\frac{n}{2}+1}^{\frac{n}{2}} \mathcal{R}_k(m t_s) e^{-i\omega_j m t_s t_s}, \quad (344)$$

at n discrete points of angular frequency:

$$\omega_j = \frac{2\pi}{n t_s} j = 2\pi \frac{2B}{n} j, \quad \text{where } j = -\frac{n}{2} + 1, \dots, 0, 1, \dots, \frac{n}{2}. \quad (345)$$

It is evident that equation (344) is the discrete analog of the continuous cross-power spectrum:

$$S_k(\omega) = \int_{-\infty}^{\infty} \mathcal{R}_k(\tau) e^{-i\omega\tau} d\tau,$$

corresponding to the continuous complex cross-correlation $\mathcal{R}_k(\tau)$. Thus equation (344) defines the discrete cross-power spectrum $S_k(\omega_j)$ at discrete points of frequency ω_j given in equation (345), corresponding to the discrete complex cross-correlation $\mathcal{R}_k(\Delta\tau_m)$.

Following Papoulis (1984), we introduced in Subsection 1.1.8 a discrete Fourier transform $S_D(\omega)$ of a function $R[m]$ given at infinite number of discrete points $-\infty < m < \infty$:

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

where \mathcal{T} is an arbitrary parameter (see equations (26) and (27)).

We can easily derive equation (344) from this discrete Fourier transform, if we take a case where $\mathcal{T} = t_s$ and

$$R[m] = \begin{cases} \mathcal{R}_k(m t_s) & \text{if } -\frac{n}{2} + 1 \leq m \leq \frac{n}{2}, \\ 0 & \text{otherwise,} \end{cases}$$

and if we multiply to the RHS a constant factor t_s . This last operation makes the LHS to have a dimension of spectral density (ex. power per unit frequency bandwidth). Thus we obtain a spectrum of the discrete complex cross-correlation:

$$S_k(\omega) = \sum_{m=-\frac{n}{2}+1}^{\frac{n}{2}} \mathcal{R}_k(m t_s) e^{-i\omega m t_s} t_s.$$

Since the discrete complex cross-correlation $\mathcal{R}_k(m t_s)$ is *band-limited* within a delay range from $(-n/2 + 1) t_s$ to $(n/2) t_s$, the sampling theorem tells us that *an optimum sampling interval in the frequency domain* should be the Nyquist interval $1/(n t_s)$ in the physical frequency, or $2\pi/(n t_s)$ in the angular frequency. Moreover, $S_k(\omega)$ above is essentially a periodic function of period $2\pi/t_s = 2\pi 2B$. Therefore, the spectrum $S_k(\omega)$ is uniquely determined as far as it is given at finite number of discrete sample points of the angular frequency ω_j given in equation (345). Thus we obtain equation (344).

Now, the search function, which we defined earlier by equation (302):

$$\mathcal{F}(\delta F_r, \delta\tau_G) = \frac{1}{2\pi N} \sum_{k=1}^N \int_0^{2\pi B} S_k(\omega) e^{i[\delta F_r(t_k - t_0) + \delta\tau_G(\omega - \omega_I)]} d\omega,$$

for the continuous case, is given for the discrete case by

$$\mathcal{F}(\delta F_r, \delta \tau_G) = \frac{2B}{Nn} \sum_{k=1}^N \sum_{j=1}^{n/2} S_k(\omega_j) e^{i[\delta F_r(t_k-t_0) + \delta \tau_G(\omega_j - \omega_I)]}, \quad (346)$$

in terms of the discrete cross-power spectrum $S_k(\omega_j)$, where we replaced the infinitesimal frequency increment $d\omega/2\pi$ by the finite frequency interval:

$$\frac{d\omega}{2\pi} \Rightarrow \Delta\nu = \frac{1}{nt_s} = \frac{2B}{n}. \quad (347)$$

This interval determines frequency resolution of the discrete cross-power spectrum $S_k(\omega_j)$.

Amplitude and Phase of Observed Cross-Power Spectrum

Our cross-power spectrum $S_k(\omega_j)$ obtained from observed data always contains the signal $\langle S_k(\omega_j) \rangle$ and the noise N_{jk} :

$$S_k(\omega_j) = \langle S_k(\omega_j) \rangle + N_{jk}. \quad (348)$$

Let us express such a spectrum through the observed amplitude $Z_k(\omega_j)$ and the observed residual phase $\Delta\phi_{jk}$ defined by

$$S_k(\omega_j) = Z_k(\omega_j) e^{-i\Delta\phi_{jk}}. \quad (349)$$

Then the search function is reduced to

$$\mathcal{F}(\delta F_r, \delta \tau_G) = \frac{2B}{Nn} \sum_{k=1}^N \sum_{j=1}^{n/2} Z_k(\omega_j) e^{-i[\Delta\phi_{jk} - \delta F_r(t_k-t_0) - \delta \tau_G(\omega_j - \omega_I)]}. \quad (350)$$

If we decompose the observed residual phase $\Delta\phi_{jk}$ into its mean value $\Delta\phi$:

$$\Delta\phi = \frac{2}{Nn} \sum_{k=1}^N \sum_{j=1}^{n/2} \Delta\phi_{jk}, \quad (351)$$

and deviation from the mean $\Delta\psi_{jk}$, i.e.

$$\Delta\phi_{jk} = \Delta\phi + \Delta\psi_{jk}, \quad (352)$$

where $\Delta\psi_{jk}$ satisfies

$$\frac{2}{Nn} \sum_{k=1}^N \sum_{j=1}^{n/2} \Delta\psi_{jk} = 0, \quad (353)$$

we can express the search function in the form:

$$\mathcal{F}(\delta F_r, \delta \tau_G) = \frac{2B}{Nn} e^{-i\Delta\phi} \sum_{k=1}^N \sum_{j=1}^{n/2} Z_k(\omega_j) e^{-i[\Delta\psi_{jk} - \delta F_r(t_k - t_0) - \delta \tau_G(\omega_j - \omega_I)]}. \quad (354)$$

Cross-Power Spectrum with High Signal-to-Noise Ratio

Now let us consider a case when the signal-to-noise ratio r_{sn} of the cross-power spectrum $S_k(\omega_j)$ is sufficiently large: $r_{sn} \gg 1$.

This may sound a little strange. Of course, the large signal-to-noise ratio is unrealistic if our cross-power spectrum is the one which is directly obtained from outputs of a hardware correlator with large number of lags (or frequency channels) and short interval t_{AP} of an AP, as we discussed in the previous Subsection. We certainly use such noisy cross-power spectrum derived from the correlator outputs in the initial stage of the peak search, when we calculate the search function at all mesh points in a wide search window, as we saw in Subsections 4.3.13 and 4.3.14. Let us denote here such large initial numbers of lags and AP's as n' and N' , respectively, which are determined by the hardware correlator.

However, in the final stage, such as the “mountaineering” stage, we search for the optimal correction values for the residual group delay and the residual fringe frequency in a very narrow area of the search window around the correct correlation peak. In other words, we already know in this stage good correction values δF_{r*} and $\delta \tau_{G*}$ which are very close to the actual residuals of the fringe frequency ΔF_r and group delay $\Delta \tau_G$, respectively.

Then, suppose in mind that we improve the delay model of our correlator using these correction values, and rerun the hardware correlator with the improved delay model. Since the delay model is sufficiently accurate in this case, we would need only a small number of lags $n \ll n'$ to reliably find the correlation peak and we could use a long accumulation period $t_{eAP} \gg t_{AP}$ without significantly losing the coherence. The long accumulation period corresponds to a small number of AP's during our observing time, i.e. $N \ll N'$. This would give us a new narrow search window of the size $n t_s \times 1/t_{eAP}$, which is just suited for our search in the final stage, and is much smaller than the initial window size $n' t_s \times 1/t_{AP}$.

Now we could derive a new cross-power spectrum $S_k(\omega_j)$ given at $n \times N$ points from “outputs of the correlator with the improved delay model”. The new spectrum must have much higher signal-to-noise ratio r_{sn} compared with the initial one given at $n' \times N'$ points, since the noise in the new spectrum

must be greatly suppressed in the course of the new hardware integration over much wider frequency interval $2B/n$ and longer time interval t_{eAP} than the initial $2B/n'$ and t_{AP} . We could expect that $r_{sn} \gg 1$, when we are certain that the final integrated fringe is actually detected with high enough signal-to-noise ratio.

Since the hardware and software processings are just two steps of the single correlation processing, final peak-search results, i.e. estimated values of the observables and signal-to-noise ratio of the detected fringe, should not significantly depend on length of the accumulation period or number of lags used in the hardware processing, as far as the delay model is sufficiently accurate. Therefore, we can estimate accuracies of the final search results on the basis of the hypothetical new cross-power spectrum $S_k(\omega_j)$ having large enough signal-to-noise ratio $r_{sn} \gg 1$.

Taylor Expansion of Search Function

In the case when $r_{sn} \gg 1$, the observed amplitude must be close to the signal amplitude: $Z_k(\omega_j) \cong |\langle S_k(\omega_j) \rangle|$, as equations (339) and (340) imply. When we observe a continuum spectrum source with no significant time variability, the signal amplitude $\mathcal{S} \equiv |\langle S_k(\omega_j) \rangle|$ must be constant both in time and frequency. Then, approximating $Z_k(\omega_j)$ with \mathcal{S} , i.e. $Z_k(\omega_j) \cong \mathcal{S}$, we reduce equation (354) to

$$\mathcal{F}(\delta F_r, \delta \tau_G) = B \mathcal{S} e^{-i\Delta\phi} \frac{2}{N n} \sum_{k=1}^N \sum_{j=1}^{n/2} e^{-i[\Delta\psi_{jk} - \delta F_r(t_k - t_0) - \delta \tau_G(\omega_j - \omega_I)]}. \quad (355)$$

Also, the phase noise in $\Delta\psi_{jk}$ of equations (354) and (355) must be small, if $r_{sn} \gg 1$, since its standard deviation $\sigma_{\Delta\psi}$ must satisfy $\sigma_{\Delta\psi} \cong 1/r_{sn} \ll 1$, as we saw in equation (341). The signal part of $\Delta\psi_{jk}$, on the other hand, is given by

$$\langle \Delta\psi_{jk} \rangle = \Delta F_r(t_k - t_0) + \Delta \tau_G(\omega_j - \omega_I),$$

where ΔF_r and $\Delta \tau_G$ are the actual residual fringe frequency and the residual group delay, respectively, as we saw in (276). Therefore, for corrections δF_r and $\delta \tau_G$, which are sufficiently close to the actual ΔF_r and $\Delta \tau_G$, the argument of the exponential function in the search function of equation (355) should be small:

$$|\Delta\psi_{jk} - \delta F_r(t_k - t_0) - \delta \tau_G(\omega_j - \omega_I)| \ll 1. \quad (356)$$

Then, if we expand the exponential term into the Taylor series with respect to the small argument given in equation (356), the search function of

equation (355) will be

$$\begin{aligned} & \mathcal{F}(\delta F_r, \delta \tau_G) \\ & \cong B S e^{-i\Delta\phi} \frac{2}{N n} \sum_{k=1}^N \sum_{j=1}^{n/2} \left\{ 1 - i [\Delta\psi_{jk} - \delta F_r(t_k - t_0) - \delta \tau_G(\omega_j - \omega_I)] \right. \\ & \quad \left. - \frac{1}{2} [\Delta\psi_{jk} - \delta F_r(t_k - t_0) - \delta \tau_G(\omega_j - \omega_I)]^2 + \dots \right\}. \end{aligned} \quad (357)$$

The first order term in the above expansion (the second term in the RHS) is equal to zero, because

$$\frac{2}{N n} \sum_{k=1}^N \sum_{j=1}^{n/2} \Delta\psi_{jk} = 0, \quad (\text{equation (353)}),$$

$$\frac{1}{N} \sum_{k=1}^N t_k = t_0, \quad (t_0 \text{ is the central time of the integration}),$$

and

$$\frac{2}{n} \sum_{j=1}^{n/2} \omega_j = \omega_I, \quad (\omega_I \text{ is the central frequency of the observing band}).$$

Therefore, we can now approximate the search function by

$$\begin{aligned} & \mathcal{F}(\delta F_r, \delta \tau_G) \\ & \cong B S e^{-i\Delta\phi} \frac{2}{N n} \sum_{k=1}^N \sum_{j=1}^{n/2} \left\{ 1 - \frac{1}{2} [\Delta\psi_{jk} - \delta F_r(t_k - t_0) - \delta \tau_G(\omega_j - \omega_I)]^2 \right\}. \end{aligned} \quad (358)$$

Equivalence to Least Squares Estimation

Equation (358) implies that the “peak search” condition:

$$| \mathcal{F}(\delta F_r, \delta \tau_G) | = \max,$$

is equivalent to the condition:

$$\frac{2}{N n} \sum_{k=1}^N \sum_{j=1}^{n/2} [\Delta\psi_{jk} - \delta F_r(t_k - t_0) - \delta \tau_G(\omega_j - \omega_I)]^2 = \min. \quad (359)$$

We can rewrite this condition in the form

$$\frac{2}{Nn} \sum_{k=1}^N \sum_{j=1}^{n/2} [\Delta\phi_{jk} - \Delta\phi - \delta F_r (t_k - t_0) - \delta\tau_G (\omega_j - \omega_I)]^2 = \min, \quad (360)$$

since we defined in equation (352) that

$$\Delta\psi_{jk} = \Delta\phi_{jk} - \Delta\phi.$$

Equation (360) is nothing but the condition for the least squares estimation of three parameters $\Delta\phi$ (mean residual phase), δF_r (residual fringe frequency) and $\delta\tau_G$ (residual group delay) from the $N \times n/2$ “observed residual phase values” $\Delta\phi_{jk}$ in the familiar linear regression scheme (see, for example, Rao (1973), for general discussions of the method of least squares). Therefore, using the standard formulae of the linear regression, we obtain the best estimates for the three parameters:

$$\begin{aligned} \Delta\phi &= \frac{2}{Nn} \sum_{k=1}^N \sum_{j=1}^{n/2} \Delta\phi_{jk}, \\ \delta F_r &= \frac{\sum_{k=1}^N (t_k - t_0) \left(\frac{2}{n} \sum_{j=0}^{n/2-1} \Delta\phi_{jk} \right)}{\sum_{k=1}^N (t_k - t_0)^2}, \\ \delta\tau_G &= \frac{\sum_{j=0}^{n/2-1} (\omega_j - \omega_I) \left(\frac{1}{N} \sum_{k=1}^N \Delta\phi_{jk} \right)}{\sum_{j=0}^{n/2-1} (\omega_j - \omega_I)^2}. \end{aligned} \quad (361)$$

Thus we confirmed the equivalence of the peak search method and the least squares estimation.

5.2.9 Dispersions of Fringe Frequency and Group Delay Observables

If we assume, for simplicity, a common dispersion $\sigma_{\Delta\psi}^2$ for all $N \times n/2$ observed residual phase values $\Delta\phi_{jk}$, the least squares method in the linear regression

problem gives us dispersions of the estimated parameters. Specifically, we obtain dispersion of the detected fringe phase observable:

$$\sigma_{\Delta\phi}^2 = \frac{2}{Nn} \sigma_{\Delta\psi}^2, \quad (362)$$

dispersion of the fringe frequency observable:

$$\sigma_{\delta F_r}^2 = \frac{\frac{2}{n} \sigma_{\Delta\psi}^2}{N \sum_{k=1} (t_k - t_0)^2}, \quad (363)$$

and dispersion of the group delay observable:

$$\sigma_{\delta\tau_G}^2 = \frac{\frac{1}{N} \sigma_{\Delta\psi}^2}{\sum_{j=0}^{n/2-1} (\omega_j - \omega_I)^2}. \quad (364)$$

Substituting equation (362) into equations (363) and (364), we obtain

$$\sigma_{\delta F_r}^2 = \frac{N \sigma_{\Delta\phi}^2}{N \sum_{k=1} (t_k - t_0)^2}, \quad (365)$$

for the dispersion of the fringe frequency observable, and

$$\sigma_{\delta\tau_G}^2 = \frac{\frac{n}{2} \sigma_{\Delta\phi}^2}{\sum_{j=0}^{n/2-1} (\omega_j - \omega_I)^2}. \quad (366)$$

for the dispersion of the group delay observable, respectively.

We can further reduce equations (365) and (366) to simpler forms if we approximate the summations in these equations by continuous integrations.

In general, for a series of x_i ($i = 1, 2, \dots, M$), which is defined by

$$x_1 = -\frac{X}{2}, \dots, x_i = -\frac{X}{2} + \frac{(i-1)X}{M-1}, \dots, x_M = \frac{X}{2}, \quad (367)$$

and satisfies

$$\sum_{i=1}^M x_i = 0,$$

we have an approximate formula:

$$\frac{1}{M} \sum_{i=1}^M x_i^2 \cong \frac{1}{X} \int_{-\frac{X}{2}}^{\frac{X}{2}} x^2 dx = \frac{X^2}{12}. \quad (368)$$

Applying this approximate formula to summations in denominators of equations (365) and (366), and performing integrations through the observing time $\tau_a = N t_{AP}$ and the observing bandwidth $\Delta\omega = 2\pi B$, respectively, we obtain

$$\sigma_{\delta F_r}^2 \cong \frac{12 \sigma_{\Delta\phi}^2}{\tau_a^2}, \quad (369)$$

for the dispersion of the fringe frequency observable, and

$$\sigma_{\delta\tau_G}^2 \cong \frac{12 \sigma_{\Delta\phi}^2}{\Delta\omega^2}, \quad (370)$$

for the dispersion of the group delay observable (see, for example, Thompson, Moran and Swenson, 2001).

Furthermore, equation (341) shows that $\sigma_{\Delta\phi} = 1/(S/N)$, when S/N is large, where S/N is the signal-to-noise ratio of the detected fringe at the correlation peak (Moran, 1976). Applying this to equations (369) and (370), we obtain estimates of the standard deviations of the fringe frequency and the group delay observables for an observation with time duration τ_a and bandwidth $\Delta\omega$:

$$\sigma_{\delta F_r} \cong \frac{\sqrt{12}}{\tau_a} \sigma_{\Delta\phi} = \frac{\sqrt{12}}{\tau_a} \frac{1}{S/N}, \quad (371)$$

and

$$\sigma_{\delta\tau_G} \cong \frac{\sqrt{12}}{\Delta\omega} \sigma_{\Delta\phi} = \frac{\sqrt{12}}{\Delta\omega} \frac{1}{S/N}, \quad (372)$$

as measures of accuracies of these two observables.

5.2.10 Accuracy of Fringe Frequency Observable

Figure 115 illustrates the linear regression model which yielded equation (371) for estimation of the accuracy (standard deviation of the thermal noise error) of the fringe frequency observable.

The signal-to-noise ratio of the detected fringe is given by equation (129):

$$S/N = \eta_c \sqrt{\frac{T_{A_1} T_{A_2}}{T_{S_1} T_{S_2}}} \sqrt{2 B \tau_a},$$

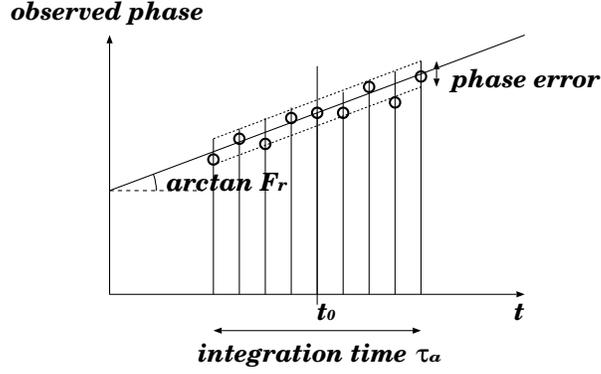


Figure 115: Linear regression model of the estimation of the fringe frequency as the phase slope in time.

where T_A is the antenna temperature, T_S is the system noise temperature, B is the bandwidth, τ_a is the integration time, and η_c is the coherence factor.

Substituting this formula to equation (371), we see that

$$\sigma_{\delta F_r} \propto \tau_a^{-3/2}.$$

However, the integration time τ_a is limited by the coherence time which is determined mainly by the atmospheric condition of our VLBI observation. Therefore, we are not allowed to improve the fringe frequency accuracy $\sigma_{\delta F_r}$ by increasing τ_a beyond the limit.

As an example, let us consider a case when $\tau_a = 400$ sec, and $S/N = 20$. Then, equation (371) gives

$$\sigma_{\delta F_r} \cong 4.3 \times 10^{-4} \text{ rad / sec.}$$

For the *physical* fringe frequency $f_r = F_r/(2\pi)$ and the delay rate $\dot{\tau}_\phi = F_r/\omega_0$ at band-center angular frequency ω_0 , which we introduced in Subsection 4.3.1, standard deviations are given by

$$\sigma_{\delta f_r} \cong \frac{\sqrt{12}}{2\pi\tau_a} \frac{1}{S/N},$$

and

$$\sigma_{\delta \dot{\tau}_\phi} \cong \frac{\sqrt{12}}{\omega_0 \tau_a} \frac{1}{S/N}.$$

Therefore, for the case of $\tau_a = 400$ sec and $S/N = 20$, we obtain

$$\sigma_{\delta f_r} \cong 6.9 \times 10^{-5} \text{ Hz,}$$

and

$$\sigma_{\delta\dot{\tau}_\phi} \cong 0.0086 \text{ psec / sec},$$

if the band-center frequency is $\nu_0 = \omega_0/(2\pi) = 8 \text{ GHz}$.

Let us try to convert this accuracy to a positional (or angular) accuracy σ_θ of the radio source position, say, derived from the fringe frequency observable. For a ground-based VLBI observing a point-like radio source towards direction \mathbf{s} with a baseline \mathbf{D} at RF frequency ω_0 , the fringe frequency could be roughly estimated by

$$F_r \cong \omega_0 \dot{\tau}_g = \omega_0 \frac{(\boldsymbol{\omega} \times \mathbf{D}) \cdot \mathbf{s}}{c}, \quad (373)$$

if the diurnal change of the geometric delay was an only main factor determining the fringe frequency, where τ_g is the geometric delay, c is the light velocity, and $\boldsymbol{\omega}$ is the Earth's angular velocity vector ($|\boldsymbol{\omega}| \simeq 7.3 \times 10^{-5} \text{ rad/sec}$). Therefore, we could roughly estimate the positional accuracy by

$$\sigma_\theta \sim \sigma_s \sim \frac{\sigma_{\delta F_r}}{2\pi(D/\lambda) |\boldsymbol{\omega}|},$$

where $\lambda = 2\pi c/\omega_0$ is the wavelength. For $D = 6000 \text{ km}$ and $\lambda = 3.7 \text{ cm}$ (8 GHz) we would obtain $\sigma_\theta \sim 1.2 \text{ mas}$ if the fringe frequency accuracy is $\sigma_{\delta F_r} \cong 4.3 \times 10^{-4} \text{ rad / sec}$. Although this estimated accuracy (thermal noise error) is fairly high, we cannot actually obtain positional accuracy of this level from the fringe frequency observable. Such limitation comes from the disturbing effects of the atmosphere as we will see below.

5.2.11 Atmospheric Limitation to Usefulness of Fringe Frequency

The irregular propagation delay due to the wet component of the troposphere severely limits actual positional accuracy obtained from the fringe frequency observable in the ordinary VLBI observation.

As we saw in equation (315), the expected form of the fringe frequency is given by

$$F_r = \omega_0(\dot{\tau}_{g_0} + \dot{\tau}_C + \dot{\tau}_A) + \dot{\theta}_1 - \dot{\theta}_2 - \dot{\Phi}_v - \dot{\Phi}_b,$$

where ω_0 is the RF (or sky) center frequency, τ_{g_0} is the geometric delay, τ_C is the clock offset, τ_A is the atmospheric delay, θ_1 , and θ_2 are the LO initial phases, Φ_v is the visibility phase, and Φ_b is the phase of the bandwidth pattern.

Using the fringe frequency observable F_r , we wish to derive useful information on radio source positions or baseline vectors, which are contained in the geometric delay term $\dot{\tau}_{g_0}$. All other terms in the above equation appear

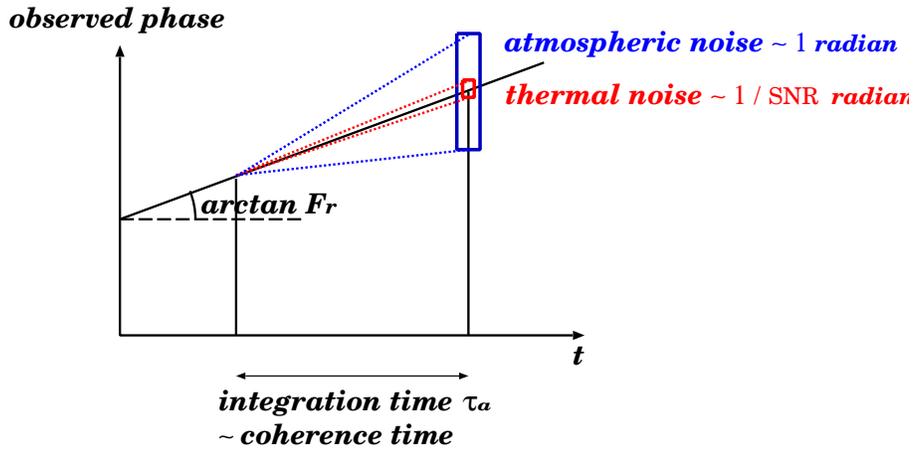


Figure 116: Accumulated atmospheric phase fluctuation during the coherence time is typically ~ 1 rad, while the thermal phase noise after the integration for the coherence time is $\simeq 1/S/N$ rad. This is the reason why the fringe frequency observable, the phase slope in time, is severely contaminated by the short timescale atmospheric phase fluctuations.

as noises for this purpose. In the case of the fringe frequency observable, the atmospheric phase variations with short timescales of the order of the coherence time, i.e. several minutes or shorter, cause the most severe effect.

In fact, accumulated phase noise due to the atmosphere is typically 1 radian during the coherence time (this is just the definition of the coherence time), while the thermal phase noise after integration for the coherence time is $\simeq 1/S/N$ radian (Figure 116). The signal-to-noise ratio of the detected fringe is usually $S/N \sim 10$ or larger, as far as the detection is reliable.

Therefore, the atmospheric phase noise causes much larger scatter of the phase slope in time, i.e. of the fringe frequency value, than the thermal phase noise, during a coherence time. This means that individual fringe frequency values, which are estimated during integration times comparable with the coherence time, scatter each other much more than the thermal noise level due to the atmosphere. This is the reason why the actual positional accuracy obtained from the fringe frequency observable is usually more than one order of magnitude worse than what is expected from the thermal noise estimation.

5.2.12 Accuracy of Group Delay Observable

Equation (372) shows that the standard deviation of the group delay observable due to the thermal noise is given by

$$\sigma_{\delta\tau_G} \cong \frac{\sqrt{12}}{\Delta\omega} \sigma_{\Delta\phi} = \frac{\sqrt{12}}{2\pi B} \frac{1}{S/N}.$$

Figure 117 illustrates the linear regression model for the group delay estimation.

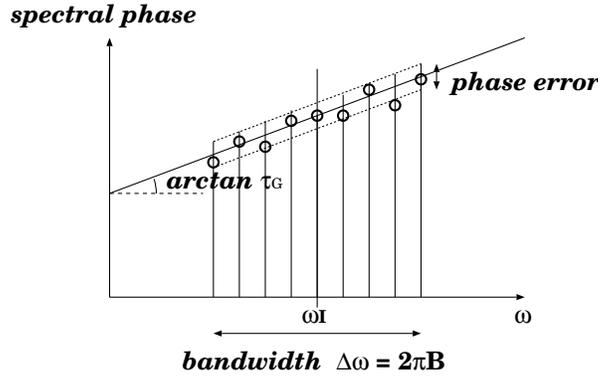


Figure 117: Linear regression model of the estimation of the group delay as the phase slope in frequency.

In view of equation (129) for the signal-to-noise ratio:

$$S/N = \eta_c \sqrt{\frac{T_{A_1} T_{A_2}}{T_{S_1} T_{S_2}}} \sqrt{2 B \tau_a},$$

the group delay accuracy is inversely proportional to 1.5-th power of the bandwidth B : $\sigma_{\delta\tau_G} \propto B^{-3/2}$. Unlike the integration time in the fringe frequency case, there is no natural limitation, other than technological, in the bandwidth B . Therefore, we can really improve the accuracy in proportion to $\propto B^{-3/2}$ by increasing the bandwidth B .

As an example, suppose that we observe a relatively strong source, which gives $S/N = 10$ after integration for a certain time τ_a with a narrow bandwidth $B = 2$ MHz. Then, suppose that we increase the bandwidth to 64 MHz, and 256 MHz, say, and observe the same source with the same integration time τ_a . In these cases, we expect group delay accuracies shown in the following table.

B	S/N	$\sigma_{\delta\tau_G}$	$c\sigma_{\delta\tau_G}$	$\sigma_{\theta(D=6000\text{km})}$
MHz		nsec	cm	mas
2	10	28	830	280
64	57	0.15	4.6	1.6
256	113	0.02	0.6	0.2

Table 11: Accuracies of group delay observables with various bandwidths when the same source is observed with the same integration time.

Here we roughly estimated the positional accuracy σ_{θ} by a simple formula:

$$\sigma_{\theta} \sim \frac{c\sigma_{\delta\tau_G}}{D},$$

assuming a baseline with length $D = 6000$ km. We can clearly see that the bandwidth is very important for the group delay observation.

5.2.13 Atmospheric Effect in Group Delay Observable

As we saw in equation (316), the expected form of the group delay is given by

$$\tau_G = \tau_{g_0} + \tau_C + \tau_A - \frac{\partial\Phi_v}{\partial\omega} - \frac{\partial\Phi_b}{\partial\omega},$$

where τ_{g_0} is the geometric delay, τ_C is the clock offset, τ_A is the atmospheric delay, Φ_v is the visibility phase, and Φ_b is the phase of the bandwidth pattern.

Of course, what we would actually like to obtain from the group delay observable is again the geometric delay τ_{g_0} , and all other terms in the above equation appear as noises for that purpose. It is clear from the above equation that the atmospheric and clock effects severely affect the group delay observation.

However, unlike in the fringe frequency case, the atmospheric phase variations with short timescales, of the order of the coherence time, do not affect the group delay observable very much. In fact, the 1 radian level accumulated phase noise during the coherence time means

$$\omega_0\Delta\tau_A \sim 1 \quad \text{radian},$$

where ω_0 is the central RF (or sky) frequency of the observing band, and $\Delta\tau_A$ is the accumulated atmospheric delay during the coherence time. This

corresponds to the accumulated excess path length $\Delta\mathcal{L}_A$:

$$\Delta\mathcal{L}_A = c \Delta\tau_A \sim \frac{c}{\omega_0} = \frac{\lambda}{2\pi},$$

where $\lambda = 2\pi c / \omega_0$ is the wavelength at the band center. Thus, if we observe at 8 GHz ($\lambda = 3.7$ cm), the accumulated excess path length during the coherence time is about 0.6 cm. This value is barely comparable with the thermal noise effect when the thermal noise is very small, i.e. in the very favorable case such as the one with $S/N > 100$ and $B = 256$ MHz in Table 11. If we observe at higher frequency, or with integration time shorter than the coherence time, or S/N is smaller, or the bandwidth B is narrower, than the above case, the short timescale atmospheric effect is smaller than the thermal noise.

Therefore, in individual group delay values obtained through integrations for the coherence time or shorter, the thermal noise normally predominates the atmospheric fluctuation. This means that the individual group delay values usually scatter with respect to each other in the range of the thermal noise level only. In this sense, the individual group delay values are well determined within the thermal noise level, unlike the fringe frequency observable.

Of course, this does not altogether mean that the atmospheric effects are not important in the group delay observables. In timescales of the order of hours or days, the irregular propagation delay due to the wet troposphere fluctuates at a few to several tens cm level, which is not small at all compared with the thermal noise, and may seriously degrade the positional accuracy, if not corrected.

Such a behavior of the atmosphere is expected from the theory of the Kolmogorov turbulence, which is known to be a good model for the atmospheric phase fluctuations. In fact, magnitudes of fluctuations with larger temporal and spatial scales are larger, than those with smaller scales, in the Kolmogorov turbulence.

Although the atmospheric effect is large, the time variation of the effect is rather slow, and one can clearly trace such variation by watching the individual group delay values, which are well determined at the thermal noise level. Therefore, it is possible to estimate and then remove, to some extent, the atmospheric delay, using results of large numbers of group delay observations made at various sky directions during 24 hours or so. Such estimation and removal of the atmospheric delay effect are routinely performed in analyses of current geodetic VLBI observations.

5.2.14 Phase Delay and Group Delay

We mentioned earlier that the thermal noise in the phase delay is much smaller than the one in the group delay, though, unfortunately, it is usually difficult to directly use the phase delay as an observable.

Figure 118 graphically illustrates why the phase delay accuracy is much higher than the group delay accuracy.

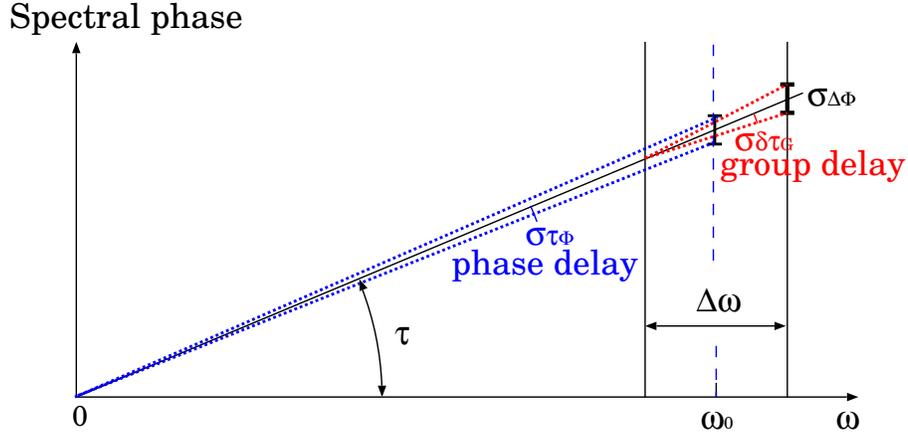


Figure 118: Thermal noise in the phase delay is usually much smaller than that in the group delay roughly by a ratio between the bandwidth and the center frequency: $\Delta\omega / \omega_0$.

In the phase delay, the phase slope in frequency, i.e. the delay τ , is determined in the whole range of frequency from 0 to the RF center frequency ω_0 . Thus, as we saw in equation (343), the phase delay τ_ϕ is derived by

$$\tau_\phi = \frac{\phi}{\omega_0},$$

from the fringe phase ϕ . On the other hand, in the group delay, the phase slope is determined within the observing bandwidth $\Delta\omega$ only.

Therefore, if we denote the accuracy of the detected fringe phase by $\sigma_{\Delta\phi}$, accuracies of the phase delay and the group delay will be roughly

$$\sigma_{\tau_\phi} \sim \frac{\sigma_{\Delta\phi}}{\omega_0},$$

and

$$\sigma_{\delta\tau_G} \sim \frac{\sigma_{\Delta\phi}}{\Delta\omega}.$$

More strict treatment with the linear regression model, which takes into account that the accuracy of the detected fringe phase results from the integration through the bandwidth $\Delta\omega$, added the extra factor of $\sqrt{12}$ to the above formula of the group delay accuracy, as we saw in equation (372).

But, apart from the factors, the ratio of the accuracies (thermal noise errors) of the phase delay and the group delay is roughly determined by the ratio of the bandwidth and the RF center frequency:

$$\frac{\sigma_{\tau_\phi}}{\sigma_{\delta\tau_G}} \sim \frac{\Delta\omega}{\omega_0}. \quad (374)$$

Thus the ratio is of the order of 0.01, if the RF center frequency is 22 GHz and the bandwidth is 256 MHz.

Since it is usually difficult to directly use the VLBI phase, we have to use the group delay, as the second accurate, but still good enough, observable in the current international geodetic and astrometric VLBI.

5.3 Bandwidth Synthesis Technique

As we saw in previous subsections, accuracy of the group delay observable improves with increasing bandwidth. However, the wider bandwidth requires the higher sampling, recording, and processing rates, which are limited by levels of relevant technologies of the date. The bandwidth synthesis technique was invented for achieving higher group–delay accuracy by effectively increasing the bandwidth, but without increasing the sampling, recording, and processing rates too much.

For this purpose, the multi–IF– (or multi–baseband–)channel design and the phase calibration (P-cal) system, as schematically shown in Figures 44 and 46, were developed in the MK–3 system.

5.3.1 For Higher Accuracy of Group–Delay Estimation

Basic concept of the bandwidth synthesis is the following.

First, let us imagine that we made a VLBI observation of a continuum spectrum source with a pretty wide bandwidth B_w , and we could somehow cross–correlate the data and obtained a spectrum $S_a(\omega)$ of the complex cross–correlation. In order to concentrate our attention to the group–delay estimation problem, we assume, for a while, that we already know a sufficiently correct value of the fringe frequency and time–average the spectrum for an integration time using the correct value. We regard that $S_a(\omega)$ is such a “time–averaged” spectrum. Signal component of the spectrum $S_a(\omega)$ is

localized in the positive frequency range and is equal to twice of the positive frequency part of the cross-power spectrum of received signals, as we discussed in Subsections 4.3.4.

We consider that the spectrum $S_a(\omega)$ covers a range of RF frequency from ω_1 to $\omega_1 + 2\pi B_w$, as schematically shown by dotted lines in Figure 119. Then, we could estimate the residual group delay by searching for the correction value $\delta\tau_G$, which maximizes amplitude of the “time-averaged search function”:

$$\mathcal{F}_a(\delta\tau_G) = \frac{1}{2\pi} \int_{\omega_1}^{\omega_1 + 2\pi B_w} S_a(\omega) e^{i\delta\tau_G(\omega - \omega_c)} d\omega,$$

where ω_c is the central frequency of the wide band, as we did in Subsections 4.3.9 and 4.3.14. Unlike in equation (302), we omit here the time averaging with respect to AP (accumulation period) numbers, since the spectrum $S_a(\omega)$ is assumed to be already time-averaged. As we saw in Subsections 4.3.9 and 5.2.8, such estimation is equivalent to determining inclination of a straight line fitted to the phase spectrum of the complex cross-correlation, or to the phase of the cross-power spectrum of received voltages, as illustrated by the dotted line in the bottom panel of Figure 119.

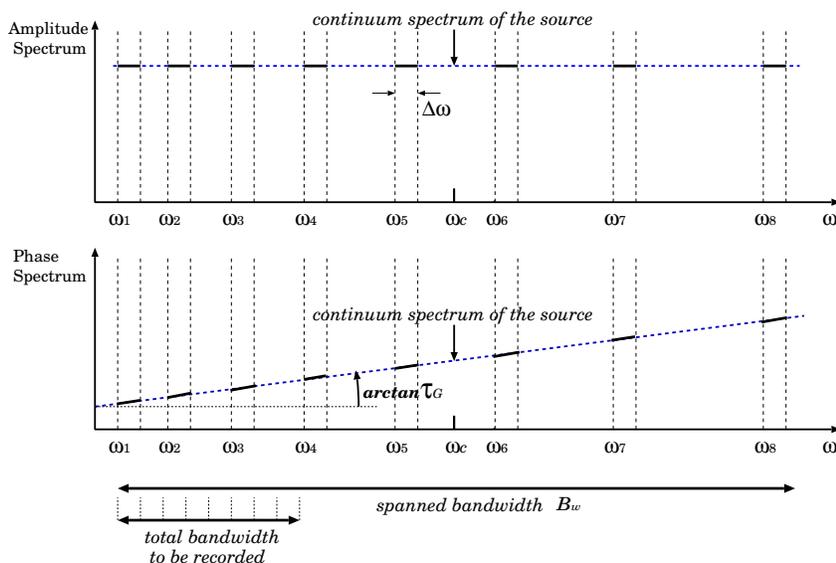


Figure 119: “Broken-comb” like spectrum (solid line) obtained by cutting a continuum spectrum (dotted line) of the complex cross-correlation with a number of rectangular filters.

Next, let us imagine that we cut out L narrow bands from the continuum spectrum of the source at lower-edge frequencies $\omega_1, \omega_2, \dots, \omega_L$ by L identical rectangular bandpass filters with a common bandwidth B , as shown in Figure 119. Then, instead of the search function $\mathcal{F}_a(\delta_G)$ for the continuum spectrum, we have a new search function $D_R(\delta\tau_G)$:

$$D_R(\delta\tau_G) = \frac{1}{2\pi} \sum_{l=1}^L \int_{\omega_l}^{\omega_l + \Delta\omega} S_a(\omega) e^{i\delta\tau_G(\omega - \omega_c)} d\omega, \quad (375)$$

where $\Delta\omega = 2\pi B$ is the angular bandwidth of a single rectangular filter, and ω_l is the lower-edge frequency of l -th band. Thus, l -th band, or l -th “IF channel”, covers an RF frequency range of $\omega_l \leq \omega < \omega_l + \Delta\omega$. We call this the “IF channel” because in actual VLBI systems the narrow band is usually cut out from an broad IF band by a video converter (or baseband converter BBC) which we mentioned in Subsection 2.1.5. In the bandwidth synthesis practice, a function composed of multi-IF-channel data such as $D_R(\delta\tau_G)$ in equation (375) is generally called the “**delay resolution function**”.

The correction value $\delta\tau_G$, which makes the amplitude of the delay resolution function $D_R(\delta\tau_G)$ maximum, must give us a good estimate of the residual group delay of the source, since the $D_R(\delta\tau_G)$ function is nothing but a search function of a hypothetical radio source, existing at the same celestial position as the actual source, but having a strange “broken-comb” like spectrum. This is equivalent to accurately estimating the phase inclination using segments of the phase spectrum spread over the wide range of frequency, or the “spanned bandwidth”, B_w , though the bandwidth B of each segment is narrow, as Figure 119 illustrates.

Now, for each IF channel with a lower-edge frequency ω_l , let us introduce a videoband frequency $\omega' = \omega - \omega_l$. Then each channel spans a common video-frequency range $0 \leq \omega' < \Delta\omega$ with a center frequency $\omega_I = \Delta\omega / 2 = \pi B$. Changing argument of integrations ω to ω' for all IF channels in equation (375), we obtain a new form of the delay resolution function $D_R(\delta\tau_G)$:

$$D_R(\delta\tau_G) = \frac{1}{2\pi} e^{-i\delta\tau_G(\omega_c - \omega_I)} \sum_{l=1}^L e^{i\delta\tau_G \omega_l} \int_0^{\Delta\omega} S_a(\omega' + \omega_l) e^{i\delta\tau_G(\omega' - \omega_I)} d\omega'. \quad (376)$$

An important point here is that we can calculate the delay resolution function as given in equation (376) using spectra of complex cross-correlations of **videoband signals** which are given by video converters with LO (local oscillator) frequencies $\omega_1, \dots, \omega_l, \dots, \omega_L$ in a MK-3 or similar multi-IF-channel VLBI system.

In fact, any frequency conversion shifts a band-limited RF spectrum to an IF band keeping its spectral shape almost unchanged, as we discussed in Chapter 3. The only exceptions are initial phases of LO's used in analog frequency conversions which are put in the phase of an IF cross-power spectrum of an interferometer.

Specifically, if we use down-converters 1 and 2 with a common LO frequency ω_{LO} and respective initial phases ϕ_{LO_1} and ϕ_{LO_2} for the frequency conversion in a two-element interferometer, then a band-limited cross-power spectrum of IF voltages $S_{v_{I_1}v_{I_2}}(\omega')$ is related to an equally band-limited original RF cross-power spectrum $S_{v_{R_1}v_{R_2}}(\omega)$, where ω and $\omega' = \omega - \omega_{LO}$ are RF and IF frequencies, respectively, by an equation

$$S_{v_{I_1}v_{I_2}}(\omega') = S_{v_{R_1}v_{R_2}}(\omega' + \omega_{LO}) e^{-i(\phi_{LO_1} - \phi_{LO_2})},$$

in the upper-sideband (USB) reception case, as we saw in Chapter 3.

In our case, different analog video converters add different LO initial phases to phases of L videoband cross-power spectra. But, we are allowed to remove the LO initial-phase differences in phases of the videoband spectra, such as $\phi_{LO_1} - \phi_{LO_2}$ above, by using outputs of the P-cal detector units of our correlator, as we discussed in Subsections 2.1.8 and 4.2.4. Therefore, we readily obtain a P-cal-corrected videoband spectrum for l -th IF channel which is just equal to the frequency shifted RF spectrum $S_a(\omega' + \omega_l)$ figuring in equation (376).

Thus, we can calculate the delay resolution function for the “broken-comb” like spectrum by using a multi-IF-channel VLBI system. In fact, video converters in such a system cut off desired bands and down-convert them to videobands. The videoband signals in each IF channel are individually formatted, recorded, and correlated. Then we obtain videoband spectra for all IF channels which, after the P-cal correction, enable us to construct the delay resolution function $D_R(\delta\tau_G)$ in equation (376).

Therefore, although we can accurately estimate the group delay using the advantage of the wide spanned bandwidth B_w , yet we do not need to sample the data with the very high Nyquist rate $2B_w$. Instead, we can sample each IF-channel data with much lower rate of $2B$ only. Moreover, we are allowed to record not the whole spanned bandwidth B_w , but only $L \times B$, as total recorded bandwidth (Figure 119). Correlation processing can be done with a moderate rate of $2B$ or so by a multi-IF-channel correlator.

This is the essence of the bandwidth synthesis technique.

5.3.2 Model Delay Resolution Function

The ‘time-averaged’ cross-power spectrum of signal parts of received voltages of a continuum spectrum source is approximately given by

$$S_a(\omega) = \begin{cases} \mathcal{S}(\omega) e^{-i[\Delta\Phi(\omega_c) + \Delta\tau_G(\omega - \omega_c)]} & \text{for } \omega \geq 0, \\ 0 & \text{for } \omega < 0, \end{cases} \quad (377)$$

as we saw in equation (277), where $\mathcal{S}(\omega)$ is the amplitude spectrum, $\Delta\Phi(\omega_c)$ is the residual phase spectrum at the central frequency ω_c of the band, and $\Delta\tau_G$ is the residual group delay. Substituting this into equation (376), we obtain

$$D_R(\delta\tau_G) = \frac{1}{2\pi} e^{-i[\Delta\Phi(\omega_c) - (\Delta\tau_G - \delta\tau_G)(\omega_c - \omega_I)]} \\ \times \sum_{l=1}^L \left[e^{-i(\Delta\tau_G - \delta\tau_G)\omega_l} \int_0^{\Delta\omega} \mathcal{S}(\omega' + \omega_l) e^{-i(\Delta\tau_G - \delta\tau_G)(\omega' - \omega_I)} d\omega' \right]. \quad (378)$$

If we model the actual spectrum by a flat spectrum:

$$\mathcal{S}(\omega) = \mathcal{S} = \text{const},$$

then the integral in equation (378) yields a sinc function, and we obtain

$$D_R(\delta\tau_G) = e^{-i[\Delta\Phi(\omega_c) - (\Delta\tau_G - \delta\tau_G)(\omega_c - \omega_I)]} \\ \times \mathcal{S} B \frac{\sin[\pi B(\Delta\tau_G - \delta\tau_G)]}{\pi B(\Delta\tau_G - \delta\tau_G)} \sum_{l=1}^L e^{-i(\Delta\tau_G - \delta\tau_G)\omega_l}. \quad (379)$$

Equation (379) shows that a good model for the amplitude of the delay resolution function is given by

$$|D_R(\delta\tau_G)| = \mathcal{S} B \left| \frac{\sin[\pi B(\Delta\tau_G - \delta\tau_G)]}{\pi B(\Delta\tau_G - \delta\tau_G)} \right| \left| \sum_{l=1}^L e^{-i(\Delta\tau_G - \delta\tau_G)\omega_l} \right|, \quad (380)$$

as far as the real spectrum is well approximated by the flat spectrum.

5.3.3 Examples of Model Delay Resolution Functions

Figure 120 shows examples of amplitudes of model delay resolution functions calculated with equation (380) for two sets of frequency distributions ω_l with number of IF channels $L = 8$.

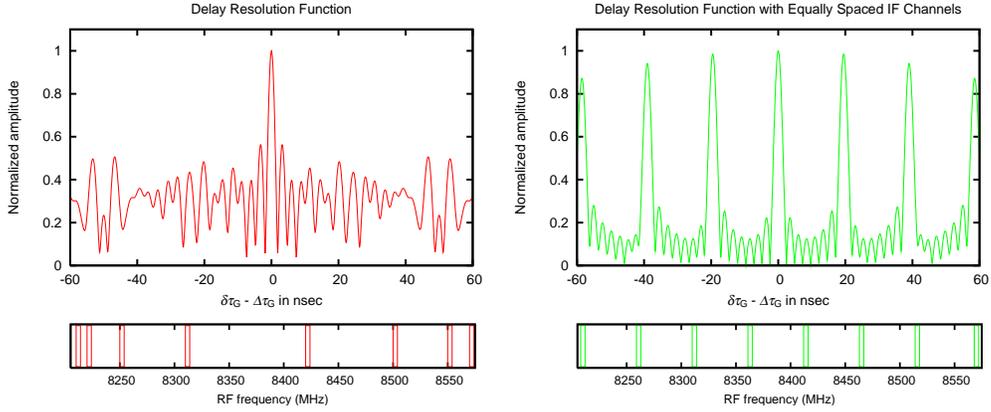


Figure 120: Amplitudes of model delay resolution functions (upper panels) with frequency distributions of IF channels shown in lower panels.

Left panel shows a case of frequency distribution ω_l shown in Table 6 of Subsection 2.1.5 for 8 IF channels in X-band with $B = 4$ MHz bandwidth each, which is actually used in some of geodetic VLBI observations. On the other hand, right panel shows a case when 8 IF channels of 4 MHz bandwidth each are distributed with almost equal intervals. In both cases, the spanned bandwidth is $B_w \cong 370$ MHz.

Both cases show sharp peaks appearing within a broad envelope of the sinc function in equation (380). This envelope is determined by the bandwidth pattern of a single filter with bandwidth B .

In either case, the central peak at $\delta\tau_G = \Delta\tau_G$ is the highest one and thus could give an accurate estimate of the residual group delay $\Delta\tau_G$. Of course, it is evident from Figure 120 that this peak is much sharper than the broad envelope corresponding to the single filter of bandwidth B . Moreover, the central peak is much sharper than a bandwidth pattern of a single filter of bandwidth $8B$ which is formed when we combine all 8 IF filters to the single filter. In fact, Figure 120 shows that full width of half maximum of the central peak in terms of the delay (which is $\sim 2/B_w$) is several nsec, and is much narrower than $2/(8B) = 62.5$ nsec expected from the bandwidth pattern of the single filter with bandwidth $8B$.

Thus, it is clear that the bandwidth synthesis is really useful for getting the sharp central peak, and therefore the higher group-delay accuracy, without too much increasing the recording bandwidth, which is only $8B = 32$ MHz in the above cases.

However, comparison of the left and right panels of Figure 120 shows that it is very important to properly choose an appropriate frequency distribution

of IF channels in order to avoid high side peaks and possible misidentification of the true peak.

In the left panel, the central $\delta\tau_G = \Delta\tau_G$ peak is well distinguished from other minor peaks. In the right panel, however, high side peaks, which are comparable with the central peak, line up with equal intervals. One of these side peaks may happen to appear as the highest due to the noise, when the signal-to-noise ratio is not high enough.

This is the reason why a little complicated frequency distribution of IF channels, such as shown in Table 6, is adopted in actual geodetic VLBI observations. In general, the narrowest interval between IF channels is important for avoiding the misidentification of the true peak, while the spanned bandwidth determines the group-delay accuracy.

Table 12 and Figure 121 exhibit another example of frequency distribution which is used in geodetic VLBI observations of the VERA array with 16 IF

Channel number	Frequency [MHz]	Bandwidth [MHz]	Bits/Sample
1	2244.00	16	2
2	2276.00	16	2
3	2308.00	16	2
4	2340.00	16	2
5	8208.00	16	2
6	8240.00	16	2
7	8272.00	16	2
8	8304.00	16	2
9	8336.00	16	2
10	8368.00	16	2
11	8400.00	16	2
12	8432.00	16	2
13	8464.00	16	2
14	8496.00	16	2
15	8528.00	16	2
16	8560.00	16	2

Table 12: Frequency distribution of 16×16 MHz IF channels in the VERA/KVN geodetic observation mode.

channels of 16 MHz bandwidth each. The same frequency distribution can be used also in geodetic VLBI observations with the KVN array.

Figure 121 shows the amplitude of the delay resolution function formed by the 12 X-band IF channels of the VERA/KVN geodetic observation mode.

Due to the current limitation of the digital baseband converter (BBC) technology, an equal-interval frequency distribution is adopted for the IF channels, as shown in Table 12. Nevertheless, Figure 121 clearly shows that the 16 MHz bandwidth of each IF channel is wide enough, compared with the channel separation, in order to make the sinc function envelope fairly

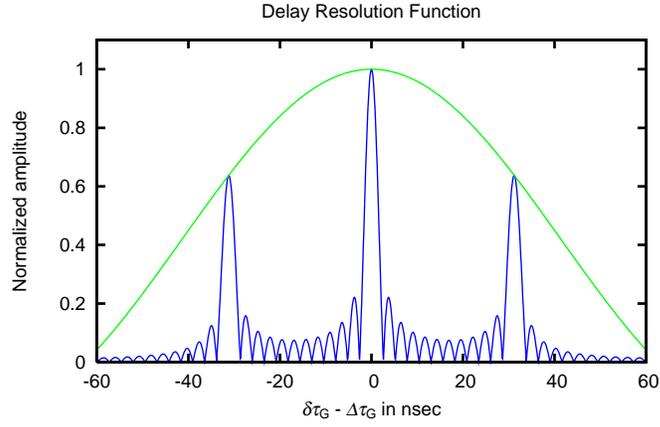


Figure 121: Amplitude of the delay resolution function formed by 12 X-band IF channels of the VERA/KVN geodetic observation mode, shown in Table 12. Also shown is the sinc function envelope determined by the bandwidth of each IF channel (16 MHz, in this case).

narrow and allow us to easily distinguish the central peak from other side peaks.

5.3.4 Ambiguity in Bandwidth Synthesis

Being a frequency derivative of phase of the cross-power spectrum of received voltages, the group delay observable does not have any cycle ambiguity unlike the fringe phase observable, as we saw in subsection 5.2.6. However, estimation of the group delay in terms of the bandwidth synthesis technique does suffer from a kind of ambiguity (e.g., Takahashi et al., 2000).

Let us consider the amplitude of sum of exponential terms in equation (380) as a function $g(\delta\tau_G)$ of the trial correction $\delta\tau_G$:

$$g(\delta\tau_G) = \left| \sum_{l=1}^L e^{-i(\Delta\tau_G - \delta\tau_G)\omega_l} \right| = \left| \sum_{l=1}^L e^{-2\pi i \nu_l (\Delta\tau_G - \delta\tau_G)} \right|, \quad (381)$$

where ν_l is the physical frequency at the lower band-edge of l -th IF channel $\nu_l = \omega_l / (2\pi)$. Then, we have

$$\begin{aligned} g(\delta\tau_G) &= \left| e^{-2\pi i \nu_1 (\Delta\tau_G - \delta\tau_G)} + e^{-2\pi i \nu_2 (\Delta\tau_G - \delta\tau_G)} + \dots + e^{-2\pi i \nu_L (\Delta\tau_G - \delta\tau_G)} \right| \\ &= \left| e^{-2\pi i \nu_1 (\Delta\tau_G - \delta\tau_G)} [1 + e^{-2\pi i (\nu_2 - \nu_1) (\Delta\tau_G - \delta\tau_G)} + \dots + e^{-2\pi i (\nu_L - \nu_1) (\Delta\tau_G - \delta\tau_G)}] \right| \\ &= \left| 1 + e^{-2\pi i (\nu_2 - \nu_1) (\Delta\tau_G - \delta\tau_G)} + \dots + e^{-2\pi i (\nu_L - \nu_1) (\Delta\tau_G - \delta\tau_G)} \right|. \end{aligned} \quad (382)$$

Equation (382) shows that $g(\delta\tau_G)$ is a periodic function of $\delta\tau_G$ with a period $T_a = 1 / \Delta\nu_{GCF}$, where $\Delta\nu_{GCF}$ is the greatest common factor of frequency differences $\nu_2 - \nu_1$, $\nu_3 - \nu_1$, \dots , and $\nu_L - \nu_1$ (we assume here that the frequency differences in Hz are integers). In fact,

$$\begin{aligned}
& g(\delta\tau_G + n T_a) \\
&= |1 + e^{-2\pi i (\nu_2 - \nu_1) (\Delta\tau_G - \delta\tau_G - n T_a)} + \dots + e^{-2\pi i (\nu_L - \nu_1) (\Delta\tau_G - \delta\tau_G - n T_a)}| \\
&= |1 + e^{-2\pi i (\nu_2 - \nu_1) (\Delta\tau_G - \delta\tau_G)} + \dots + e^{-2\pi i (\nu_L - \nu_1) (\Delta\tau_G - \delta\tau_G)}| \\
&= g(\delta\tau_G),
\end{aligned} \tag{383}$$

for any integer n , since

$$(\nu_m - \nu_1) T_a = \frac{\nu_m - \nu_1}{\Delta\nu_{GCF}} = \text{integer}, \quad (\text{where } m = 2, 3, \dots, L),$$

for any m by definition of the greatest common factor. Therefore, the peak at $\delta\tau_G = \Delta\tau_G$ is not the only highest peak in the function $g(\delta\tau_G)$. Identical peaks are obtained also at $\delta\tau_G = \Delta\tau_G \pm T_a$, $\delta\tau_G = \Delta\tau_G \pm 2 T_a$, and so on.

Of course, amplitude of the actual delay resolution function $|D_R(\delta\tau_G)|$ given in equation (380) has the sinc function envelope:

$$\left| \frac{\sin[\pi B(\Delta\tau_G - \delta\tau_G)]}{\pi B(\Delta\tau_G - \delta\tau_G)} \right|,$$

multiplied to the $g(\delta\tau_G)$ function. Therefore, the peak at $\delta\tau_G = \Delta\tau_G$ of $|D_R(\delta\tau_G)|$ is higher than those at $\delta\tau_G = \Delta\tau_G \pm T_a$, $\delta\tau_G = \Delta\tau_G \pm 2 T_a$, \dots , since the sinc function envelope takes its maximum at $\delta\tau_G = \Delta\tau_G$.

Nevertheless, when the bandwidth B of a single IF channel is not wide enough, specifically $B < 1/T_a$, then the sinc function envelope decreases slowly with increasing $|\delta\tau_G - \Delta\tau_G|$. As a result, peaks at $\delta\tau_G = \Delta\tau_G \pm T_a$, $\delta\tau_G = \Delta\tau_G \pm 2 T_a$, \dots may be still high enough compared with the true highest peak at $\delta\tau_G = \Delta\tau_G$, so that they could be misidentified with the true one under circumstances that the noise contribution is not negligibly small. This is the ambiguity problem in the bandwidth synthesis.

As an example, let us again consider the frequency distribution:

ν_1	ν_2	ν_3	ν_4	ν_5	ν_6	ν_7	ν_8
8209.99	8219.99	8249.99	8309.99	8419.99	8499.99	8549.99	8569.99

for 8 channels in X-band with 4 MHz bandwidth listed in Table 6 of Subsection 2.1.5. Unit of frequency here is MHz. In this case, the frequency differences are

$\nu_2 - \nu_1$	$\nu_3 - \nu_1$	$\nu_4 - \nu_1$	$\nu_5 - \nu_1$	$\nu_6 - \nu_1$	$\nu_7 - \nu_1$	$\nu_8 - \nu_1$
10	40	100	210	290	340	360

in MHz. The greatest common factor of the above frequency differences is 10 MHz. Therefore, the period of the $g(\delta\tau_G)$ function, or the “delay ambiguity”, T_a is equal to $T_a = 1 / (10 \text{ MHz}) = 100 \text{ nsec}$.

Figure 122 shows the amplitude of the model delay resolution function for the above frequency distribution which is the same with the left panel in Figure 120 but given in much wider delay range from -325 to 325 nsec . We can

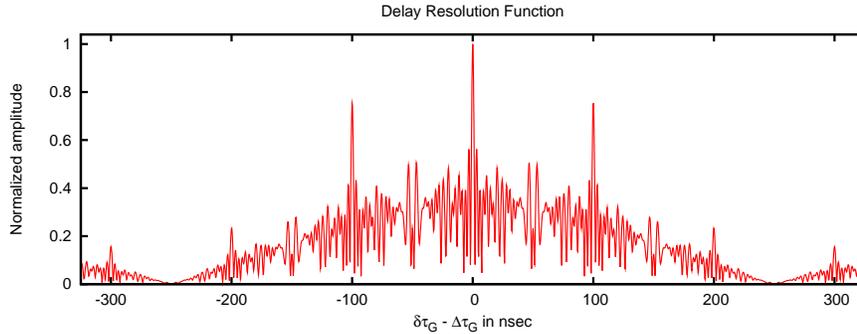


Figure 122: Amplitude of the model delay resolution function formed by 8 X-band IF channels listed in Table 6 of Subsection 2.1.5.

clearly see “ambiguity peaks” at ± 100 , ± 200 , and $\pm 300 \text{ nsec}$ corresponding to the 100 nsec delay ambiguity T_a . Heights of the $\pm 100 \text{ nsec}$ peaks reach nearly 80 % of the true central peak. Therefore, chances of misidentification of these peaks with the true peak are not very small unless the signal-to-noise ratio is high enough.

Actually, the ambiguity peaks are often detected in data processings with the bandwidth synthesis technique. In such a case, estimated group delay shows a large offset from theoretically predicted delay, which is close to an integer multiple of the delay ambiguity T_a . When this happens, one should correct the estimated group delay value by subtracting the integer multiple of the delay ambiguity T_a , so that corrected group delay values are well aligned around the theoretically predicted values.

As we already mentioned in Subsection 5.3.3, we may also misidentify side peaks of the delay resolution function other than the ambiguity peaks when frequency distribution of IF channels is not adequately selected.

Therefore, we must be always careful about the misidentification problems in actual processings of our data with the bandwidth synthesis technique.

5.3.5 Actual Bandwidth Synthesis

In actual data processings, we usually perform the bandwidth synthesis in the following way (e.g., Takahashi et al., 2000).

1. We obtain complex cross-correlations and their spectra for all L IF channels from outputs of a hardware correlator. We denote a videoband spectrum of a complex cross-correlation of l -th IF channel at k -th accumulation period (AP) as $S_k^l(\omega')$, where ω' is the videoband frequency and superscript l indicates the IF-channel number. Note that this is an “instantaneous” spectrum at each AP, unlike the “time-averaged spectrum” $S_a(\omega)$ which we introduced as an auxiliary concept in Subsection 5.3.1. Note also that we keep describing the spectrum as a continuous function of frequency ω' for simplicity, although in actual processings we deal with discrete spectra obtained from outputs of a digital correlator.
2. We first construct search functions used in the peak search as described in Subsection 4.3.14 for all IF channels. For l -th IF channel, the search function $\mathcal{F}^l(\delta F_r, \delta \tau_G)$ is given by equation (302):

$$\mathcal{F}^l(\delta F_r, \delta \tau_G) = \frac{1}{2\pi N} \sum_{k=1}^N \int_0^{2\pi B} S_k^l(\omega') e^{i[\delta F_r(t_k - t_0) + \delta \tau_G(\omega' - \omega_I)]} d\omega', \quad (384)$$

where δF_r and $\delta \tau_G$ are trial corrections for the residual fringe frequency and the residual group delay, respectively, N is the number of AP's in an integration, t_k is a central time of k -th AP, and t_0 and ω_I are reference time and reference frequency which are usually selected at a central time of an integration and at a central frequency of a videoband, respectively, as before. Next we average amplitudes of search functions of all L IF channels to make a new combined search function:

$$\mathcal{F}_C(\delta F_r, \delta \tau_G) = \frac{1}{L} \sum_{l=1}^L |\mathcal{F}^l(\delta F_r, \delta \tau_G)|.$$

Then we obtain optimum estimates for the residual fringe frequency ΔF_r and the residual group delay $\Delta \tau_G$ by searching for corrections δF_{rC} and $\delta \tau_{GC}$ which maximize the combined search function $\mathcal{F}_C(\delta F_r, \delta \tau_G)$. This search process performed within a narrow bandwidth B of baseband channels is called the “**coarse search**” in the bandwidth-synthesis practice.

3. We correct the videoband spectrum $S_k^l(\omega')$ using the estimated residual fringe frequency and residual group delay values obtained in the coarse search. Then we get a “corrected spectrum” $S_{C_k}^l(\omega')$ for l -th IF channel:

$$S_{C_k}^l(\omega') = S_k^l(\omega') e^{i[\delta F_{rC}(t_k - t_0) + \delta \tau_{GC}(\omega' - \omega_I)]}. \quad (385)$$

Of course, this correction is not complete due to errors of the estimated residuals δF_{rC} and $\delta \tau_{GC}$, and therefore the corrected spectrum $S_{C_k}^l(\omega')$ still contains remaining residuals of the fringe frequency $\Delta' F_r = \Delta F_r - \delta F_{rC}$ and group delay $\Delta' \tau_G = \Delta \tau_G - \delta \tau_{GC}$. But the remaining residuals are now reduced to levels smaller than the standard deviations given in equations (371) and (372) which are evaluated for a single channel of bandwidth B .

4. We correct for the LO initial-phase effects due to analog video converters by applying phase correction values (the “P-cal corrections”) which we get from outputs of the P-cal detector unit in our correlator. Let us denote the P-cal correction for l -th IF channel as $\delta \phi_l$. Then we obtain a “P-cal corrected spectrum” $S_{P_k}^l(\omega')$ by an equation:

$$S_{P_k}^l(\omega') = S_{C_k}^l(\omega') e^{i\delta \phi_l}. \quad (386)$$

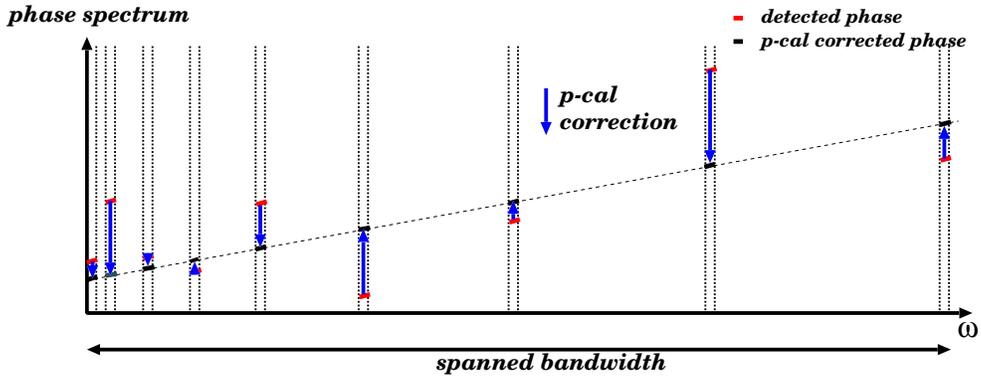


Figure 123: Removal of instrumental phases in IF channels by means of P-cal corrections allows us to determine the phase slope with frequency, i.e. the group delay.

Figure 123 illustrates importance of the P-cal correction in the bandwidth synthesis with analog video converters.

5. Now the P-cal corrected videoband spectrum $S_{P_k^l}(\omega')$ of l -th IF channel must be equal to the frequency shifted RF spectrum as we discussed in Subsection 5.3.1. Therefore, we are now in position to construct the delay resolution function following equation (376). However, we no longer use the “time-averaged” spectrum $S_a(\omega)$ as we did in Subsection 5.3.1. Instead, we introduce a new delay resolution function $\mathcal{D}_R(\delta F_r, \delta \tau_G)$:

$$\mathcal{D}_R(\delta F_r, \delta \tau_G) = \frac{1}{2\pi N} e^{-i\delta\tau_G(\omega_c - \omega_I)} \times \sum_{l=1}^L e^{i\delta\tau_G \omega_l} \sum_{k=1}^N \int_0^{2\pi B} S_{P_k^l}(\omega') e^{i[\delta F_r(t_k - t_0) + \delta\tau_G(\omega' - \omega_I)]} d\omega', \quad (387)$$

which directly use the P-cal corrected videoband spectrum $S_{P_k^l}(\omega')$ at k -th AP. Time averaging (averaging over all AP's) is included in this new delay resolution function.

6. We conduct a two-dimensional search for correction values δF_{r_f} and $\delta \tau_{G_f}$ which maximize amplitude of the new delay resolution function $\mathcal{D}_R(\delta F_r, \delta \tau_G)$ in equation (387). δF_{r_f} gives a good estimate of the residual fringe frequency $\Delta' F_r$ suited to the data in all IF channels, while $\delta \tau_{G_f}$ gives an accurate estimate of the residual group delay $\Delta' \tau_G$ determined by using the advantage of the wide spanned bandwidth B_w . This second search process is called the “**fine search**” in the bandwidth-synthesis practice.

Actual peak search is conducted in much the same way as in the single band case, i.e., $|\mathcal{D}_R(\delta F_r, \delta \tau_G)|$ is first calculated on grid points of a mesh in the $\delta F_r - \delta \tau_G$ plane, and then, starting from the grid point which gives the maximum amplitude, the true peak is searched for by the ‘mountaineering method’.

The search window for the fine search could be much narrower than that in the coarse search since we applied corrections δF_{r_c} and $\delta \tau_{G_c}$ obtained in the coarse search to the videoband spectra. It is desirable to choose the delay-window size to be larger than twice of the delay ambiguity T_a in order to avoid misidentification of an ambiguity peak with the true maximum peak. Also, it is desirable to choose the grid interval in the delay window to be smaller than $1/B_w$ in order not to overlook the very sharp peak.

Figures 124 and 125, brought from the NICT webpage, show the amplitude (denoted as “COR. AMP” in the Figures) of the delay resolution

function and its close-up view, respectively, which was obtained in 1995 in an actual VLBI observation of a continuum source for the KSP Project. The Figures show results of the two-dimensional fine search for the residual fringe frequency and the residual group delay based on correlator outputs of the observed data. 7 IF channels of X-band, distributed from 7700 to 8600 MHz, are used in this bandwidth synthesis observation. The central peak is very sharp owing to the 900 MHz wide spanned bandwidth adopted in the KSP observing system.

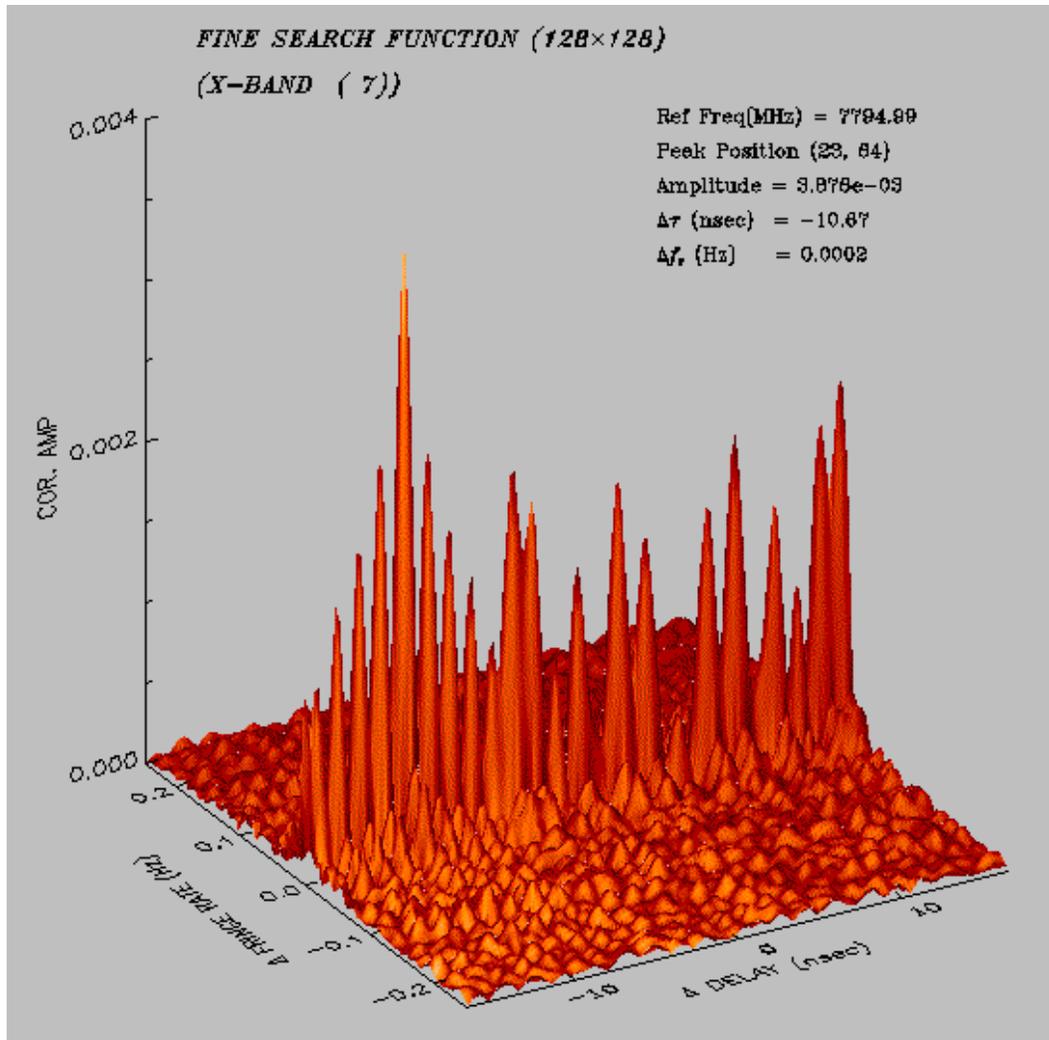


Figure 124: Amplitude of the delay resolution function calculated in the two-dimensional fine search of the bandwidth synthesis on the basis of observed data obtained in the KSP Project. Figure courtesy of the NICT.

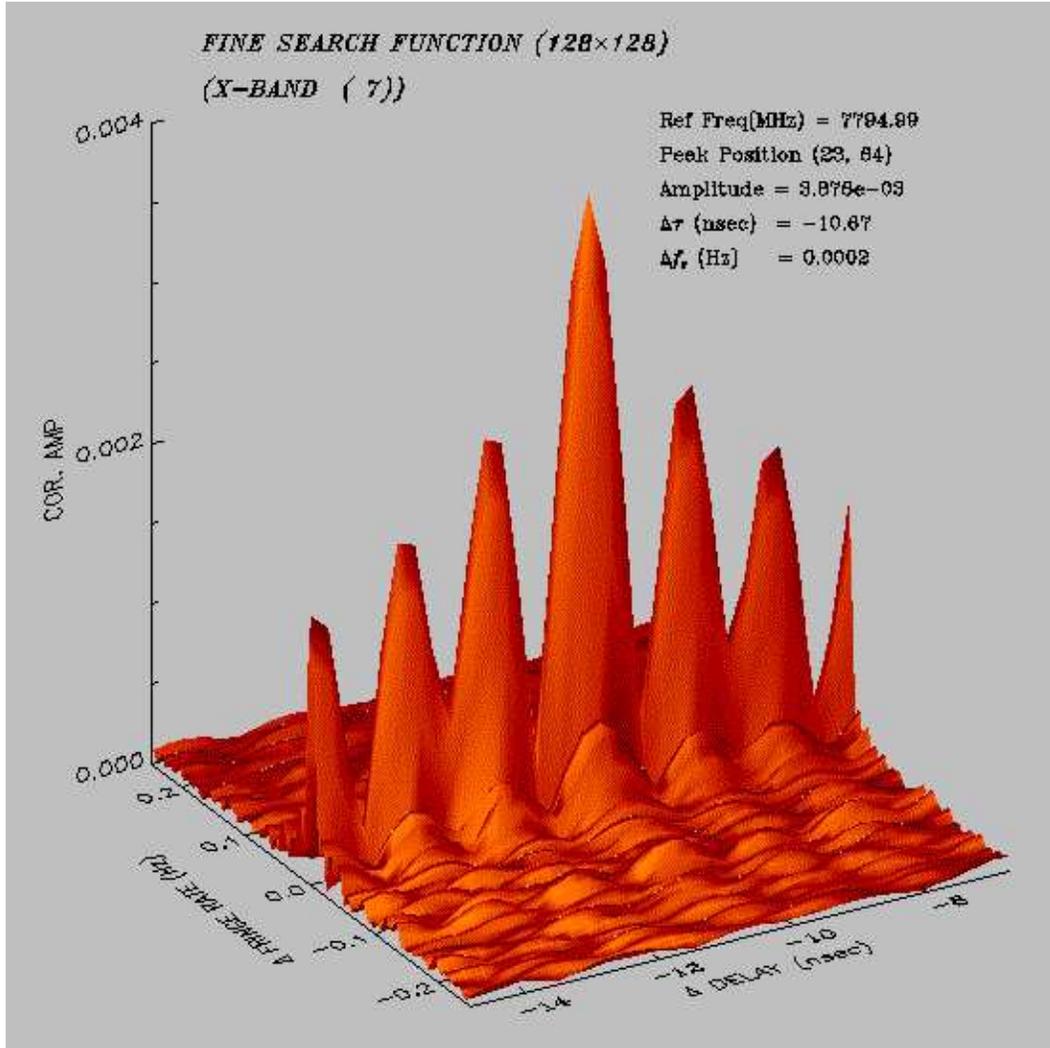


Figure 125: Close-up view of the amplitude of the delay resolution function given in Figure 124. Figure courtesy of the NICT.

5.3.6 Group-Delay Accuracy in Bandwidth Synthesis

We can estimate the accuracy of the group delay observable determined with the bandwidth synthesis technique, using the least squares scheme for the linear regression problem (Sekido, 2001), such as shown in Figure 126. We

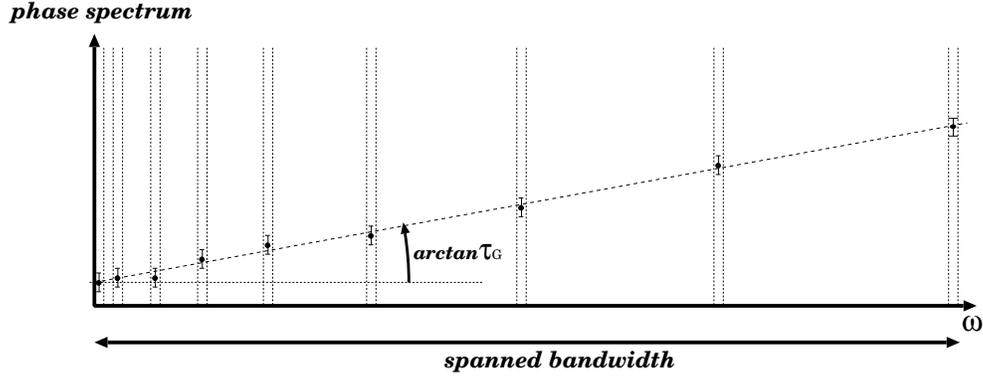


Figure 126: The group-delay accuracy in the bandwidth synthesis is estimated by solving the linear regression problem for fitting a straight line to the phase values in IF channels.

could easily justify this procedure on the basis of the discussions in Subsection 5.2.8, since the peak search algorithm in the bandwidth synthesis is essentially the same with the one for the single band discussed there.

Solving the linear regression problem, we obtain an estimation of the thermal noise standard deviation of the group delay determined with the bandwidth synthesis technique:

$$\sigma_{\delta\tau_G} = \frac{1}{2\pi \Delta\nu_{rms} S/N}, \quad (388)$$

with

$$\Delta\nu_{rms} = \sqrt{\frac{\sum_{l=1}^L (\nu_{b_l} - \nu_c)^2}{L}}, \quad (389)$$

where $\Delta\nu_{rms}$ is called the “root mean square bandwidth”, ν_{b_l} is a center frequency of l -th out of L IF channels ($\nu_{b_l} = (\omega_l + \omega_I) / (2\pi)$), and the “central frequency of the wide band” ν_c is chosen to fulfill a condition:

$$\sum_{l=1}^L (\nu_{b_l} - \nu_c) = 0. \quad (390)$$

The signal-to-noise ratio S/N here is determined by the total recorded bandwidth $L \times B$, where B is the bandwidth of each IF channel.

If $S/N = 16$ and $\Delta\nu_{rms} = 100$ MHz, then $\sigma_{\delta\tau_G} \cong 0.1$ nsec, or $c\sigma_{\delta\tau_G} \cong 3$ cm. Thus, the MK-3 system realized the 0.1 nsec, or cm-level, accuracy, even when the total recorded bandwidth was only $8 \times 2 = 16$ MHz (X band) and $6 \times 2 = 12$ MHz (S band), owing to the bandwidth synthesis technique.

With today's technology, we could realize 16×16 MHz system with the root mean square bandwidth of $\Delta\nu_{rms} \cong 800$ MHz. Then, the standard deviation of the thermal noise in the group delay observable would be below 0.01 nsec, or 3 mm.

6 Atmosphere-Related Difficulties and Their Solutions in VLBI

Apparently, VLBI researches in astrophysics are currently developing towards following directions:

- higher frequency (for short mm-wave VLBI),
- higher angular resolution (to unveil very centers of AGNs),
- higher astrometric accuracy ($\sim 10 \mu\text{as}$ for kpc-scale measurements),
- higher sensitivity (from mJy to μJy),
- higher time resolution (to elucidate subdaily variations of AGNs, microquasars, and other active objects).

In geodesy, large efforts are made for

- rapid production of highly accurate geodetic data (~ 1 mm, initial results within 24 hours after observation) for better understanding of the planet Earth, prediction of natural hazards, and deep-space navigation.

Naturally, these trends require further advancements in VLBI technologies and solution of existing difficulties.

Rather rapid progress in VLBI technologies, which we have seen so far, such as dedicated global arrays including spaceborne antennas, large aperture mm-wave antennas, low noise receiving systems, ultra-wideband recording systems, high-speed samplers, digital filters, fiber-linked e-VLBIs, and PC-based large scale correlators, is likely to eventually meet the above trends.

So, hopefully, we will encounter no essential technological difficulty in near future, as far as instrumental sides of VLBI systems are concerned.

On the other hand, the atmosphere-related difficulties remain quite serious and affect all the above directions of future development of VLBI.

While the atmospheric absorption effects are mostly unavoidable unless we go to high mountains or space orbits, there are ways to compensate, at least partially, the refraction effects.

We will discuss in this section the refraction-related difficulties and their possible solutions. This is the field of research which is still under the trial-and-error stage. So, we will examine basic problems and ideas in this field without being involved in details.

Discussions in this section are partly based on a review report (Sasao, 2003) in the Proceedings of the IVS Symposium “New Technologies in VLBI” held at Kyeongju in November, 2002.

6.1 Difficulties Associated with mm-Wave VLBI

Difficulties in VLBI can be best perceived in its high-frequency frontier, the mm-wave VLBI, since all instrumental and environmental problems become increasingly severe as we go to higher frequency.

In 1988, James Moran concisely discussed “VLBI at Millimeter Wavelengths” in his lecture “Introduction to VLBI” given in a conference held at Bologna (Moran, 1989). Perhaps, this was done to warn people who might naively wish to obtain high angular resolution in terms of the mm-wave VLBI without seriously thinking difficulties associated with it. Moran himself regarded the mm-VLBI as “an important new frontier in VLBI research”, and undoubtedly believed that the difficulties he mentioned will be eventually overcome by future advancements in technology.

Moran took the standard formula for the signal-to-noise ratio on a single baseline:

$$S/N = \frac{S_\nu A_e \sqrt{2B\tau_c}}{kT_s},$$

where S_ν is the flux density of the observed source, T_s is the geometric mean of the system noise temperatures of the two telescopes, A_e is the geometric mean of their effective apertures, B is the receiving bandwidth, τ_c is the coherent integration time, and k is Boltzmann’s constant.

Then, Moran discussed the wavelength λ -dependence of the sensitivity in the following way.

The system noise temperature is roughly

$$T_s \propto \lambda^{-1},$$

since the best T_s at that time was 30 K at 8 GHz, and 300 K at 80 GHz.

The antenna aperture is

$$A_e \propto \lambda^2,$$

since the ratio [diameter] / [surface tolerance] is roughly constant, in view of structural stiffness and cost.

The bandwidth is

$$B \propto \lambda^0,$$

since it is recorder-limited, irrespective of the observing wavelength.

The flux densities of the radio sources generally vary as

$$S_\nu \propto \lambda^{0.5},$$

though compact sources tend to have flat spectra at high frequency.

The coherent integration time is

$$t_c \propto \lambda,$$

since the RMS phase fluctuations are proportional to λ^{-1} for both the frequency standard and the atmosphere.

Combining all dependences above, Moran gave a terrifying formula:

$$S/N \propto \lambda^4,$$

which predicts a 10000 times drop of the signal-to-noise ratio as we go from the 10 GHz to the 100 GHz band!

Nevertheless, we see remarkable progress in mm-wave VLBI observations in recent years. This includes the first successful intercontinental VLBI observation of AGNs at 147 GHz, the highest frequency VLBI detection of SiO maser lines at 129 GHz, and the extensive survey of a large number of continuum sources at 86 GHz.

Advancements in mm-wave radio astronomy during the past decades were really impressive.

Now the best system noise temperature T_s at 147 GHz in actual VLBI observations is ~ 200 K, while 30 K at 15 GHz is still more or less a good figure. Therefore,

$$T_s \propto \lambda^{-0.65},$$

rather than λ^{-1} .

The largest improvement occurred in the antenna aperture A_e . Large aperture mm-wave telescopes, such as IRAM 30 m in Pico Veleta, Nobeyama 45 m, and inner 60 m of Effelsberg 100 m, became involved in VLBI observations. Moreover, recent inclusion of mm-wave interferometer arrays as VLBI

elements further increased collecting areas of mm–wave VLBI. Now, the “[diameter] / [surface tolerance] \sim const” law is no longer valid, and practically we do not see any essential difference between aperture sizes of telescopes used in recent cm–wave and mm–wave VLBI. Therefore, roughly speaking,

$$A_e \propto \lambda^0.$$

The bandwidth is still recorder–limited, and therefore

$$B \propto \lambda^0.$$

But we have to notice that the recordable bandwidth itself increased more than 10 times since the end of 1980’s, greatly enhancing the VLBI sensitivity.

VLBI surveys of large numbers of extragalactic radio sources at various frequencies revealed that spectra of compact sources are really flatter than $\lambda^{0.5}$. Although we still do not have enough material to derive a reliable average spectral index, it seems safe to assume that the flux density S_ν varies with wavelength not faster than

$$S_\nu \propto \lambda^{0.35}.$$

The atmospheric phase fluctuations still limit the coherent integration time τ_c of mm–wave VLBI strictly according to the equation:

$$\tau_c \propto \lambda.$$

In view of recent successful applications of phase–referencing techniques in cm–wave VLBI, which enabled hours of integration for detecting weak sources, mm–wave VLBI here lags most severely behind cm–wave VLBI.

Combining above dependences, we now have

$$S/N \propto \lambda^{1.5},$$

which is much better than what we had decades ago: $S/N \propto \lambda^4$. Thus, it is understandable that we now have tolerable sensitivity to make brilliant successes in short mm–wave VLBI observations, even including intercontinental ones, under favorite atmospheric conditions.

References

- Abramowitz, M., and Irene A. Stegun, I. A., 1964, Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, *Dover* (New York).
- Allan, D.W., 1966, Statistics of Atomic Frequency Standards, *Proc. IEEE*, **54**, 221–230.
- Hachisuka, K., 2001, Toward Determination of Outer Galactic Rotation Curve — High Precision Astrometry with Phase-Referencing VLBI, PhD Thesis, Graduate University for Advanced Studies, Japan.
- 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.
- Moran, J.M., 1976, Very Long Baseline Interferometric Observations and Data Reduction, in *Methods of Experimental Physics*, **12C**, ed. M.L. Meeks, Academic Press (New York), 228–260.
- Moran, J.M., 1989, Introduction to VLBI, in *Very Long Baseline Interferometry Techniques and Applications*, eds. M. Felli and R.E. Spencer, Kluwer Academic Publishers (Dordrecht), 27–45.
- 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.
- Rao, C.R., 1973, Linear Statistical Inference and Its Applications, Second Edition, *John Wiley & Sons* (New York).
- Rogers, A.E.E., and Moran, J.M., 1981, Coherence Limits for Very-Long-Baseline Interferometry, *IEEE Trans. Instrum. Meas.*, **IM-30**, 283–286.
- Sasao, T., 2003, Technologies for mm and Sub-mm VLBI: Multi-frequency Feed as a Tool for mm-Wave VLBI, in *New Technologies in VLBI: Proceedings of the International VLBI Service for Geodesy and Astrometry*, ed. Y.C. Minh, ASP Conference Series Vol. 306, 53–74.
- Sekido, M., 2001, Pulsar Astrometry by VLBI, PhD Thesis, Grad-

- uate University for Advanced Studies, Japan.
- 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 Middleton, D., 1966, The Spectrum of Clipped Noise, *Proc. IEEE*, **54**, 2–19.